OBSERVATIONAL CONSTRAINTS ON THE STRUCTURE AND EVOLUTION

OF QUASARS

by

Brandon C. Kelly

———————————————

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF ASTRONOMY

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2 0 0 8

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Brandon C. Kelly entitled "Observational Constraints on the Structure and Evolution of Quasars" and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

_____ Date: 11 April 2008
Jill Bechtold

_____ Date: 11 April 2008
Xiaohui Fan

_____ Date: 11 April 2008
Romeel Davé

_____ Date: 11 April 2008
Christopher Walker

_____ Date: 11 April 2008
Gregory Rudnick

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

_____ Date: 11 April 2008
Dissertation Director: Jill Bechtold

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Brandon C. Kelly

Acknowledgments

DEDICATION

TABLE OF CONTENTS

TABLE OF CONTENTS — *Continued*

TABLE OF CONTENTS — *Continued*

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

   I use X-ray and optical data to investigate the structure of quasars, and its de-
pendence on luminosity, redshift, black hole mass, and Eddington ratio. In order
to facilitate my work, I develop new statistical methods of accounting for mea-
surement error, non-detections, and survey selection functions. The main results
of this thesis follow. (1) The statistical uncertainty in the broad line mass esti-
mates can lead to significant artificial broadening of the observed distribution of
black hole mass. (2) The $z = 0.2$ broad line quasar black hole mass function falls
off approximately as a power law with slope ~ 2 for $M_{BH} \gtrsim 10^8 M_{\odot}$. (3) Radio-
quiet quasars become more X-ray quiet as their optical/UV luminosity, black hole
mass, or Eddington ratio increase, and more X-ray loud at higher redshift. These
correlations imply that quasars emit a larger fraction of their bolometric luminos-
ity through the accretion disk component, as compared to the corona component,
as black hole mass and Eddington ratio increase. (4) The X-ray spectral slopes of
radio-quiet quasars display a non-monotonic trend with Eddington ratio, where
the X-ray continuum softens with increasing Eddington ratio until $L/L_{Edd} \sim 0.3$,
and then begins to harden. This observed non-monotonic trend may be caused
by a change in the structure of the disk/corona system at $L/L_{Edd} \sim 0.3$, possibly
due to increased radiation pressure. (5) The characteristic time scales of quasar
optical flux variations increase with increasing M_{BH}, and are consistent with disk
orbital or thermal time scales. In addition the amplitude of short time scale vari-
ability decreases with increasing M_{BH}. I interpret quasar optical light curves as
being driven by thermal fluctuations, which in turn are driven by some other un-
derlying stochastic process with characteristic time scale long compared to the
disk thermal time scale. The stochastic model I use is able to explain both short

and long time scale optical fluctuations.

CHAPTER 1

INTRODUCTION

Shortly after the discovery that quasars[1] lie at extragalactic distances (Schmidt, 1963; Sandage, 1965), it was proposed that they are powered by accretion onto a supermassive black hole (SMBH, Salpeter, 1964; Lynden-Bell, 1969; Rees, 1984). A succesful model for the nature of the quasar accretion process involves the UV/optical continuum arising from a geometrically thin, optically thick cold accretion disk (e.g., Shakura & Syunyaev, 1973), and the X-ray continuum arising from a hot, optically thin corona that Compton upscatters the disk UV photons (e.g., Haardt & Maraschi, 1991). In radio-loud active galactic nuclei (AGN) the radio continuum arises from synchrotron emission from a jet (e.g., see Krolik, 1999a, and references therein), and recently there is evidence that even radio-quiet objects possess weak jets or outflows (Leipski et al., 2006).

While it appears that the basic ingredients of the quasar engine are currently understood, a number of uncertainties remain. First, a considerable amount of our interpretation and understanding of AGN accretion disks, and consequently their optical/UV emission, is based on the so-called α-prescription (Shakura & Syunyaev, 1973). Within the standard α model, the viscosity, thought to be the source of the thermal disk emission and outward transfer of angular momentum in the disk, is parameterized as being proportional to the total disk pressure. However, when the disk is dominated by radiation pressure, as thought to be the case in the regions emitting the optical and UV flux in quasars (Krolik, 1999a), an α-disk is both thermally and viscously unstable (Shakura & Sunyaev, 1976; Light-

---

[1]Throughout this thesis I will refer to quasars and Active Galactic Nuclei interchangeably; no luminosity difference between the two objects is assumed

man & Eardley, 1974). In particular, for a radiation pressure dominated disk, the thermal instability is expected to grow exponentially on a time scale similar to the thermal time scale. For AGN, this is $t_{th} \sim 5$ yrs, assuming typical parameters. It seems odd that nature would be forced to make inherently unstable accretion disks in AGN, and indeed there is no evidence for instabilities in the optical light curves of AGN with time span $\sim t_{th}$ (Collier & Peterson, 2001, see also Chapter 8). This fact is also played out in 3-dimensional magneto-hydrodynamic simulations of radiation pressure-dominated AGN accretion disks, where no thermal instability is seen (e.g., Turner, 2004). Moreover, the most promising physical mechanism behind the viscous torque is the magneto-rotational instability (MRI, Balbus & Hawley, 1991, 1998), and recent numerical and analytical work has suggested that the α prescription may be a poor representation for MRI-driven viscosity (Pessah et al., 2008).

In addition, the standard disk model of Shakura & Syunyaev (1973) is unable to explain the X-ray emission from AGN, as the accretion disk temperature is too cold to produce the large amount of X-ray photons. Currently, the accepted model involves Compton upscattering of UV disk photons by a hot, optically thin corona, possibly heated by bouyant magnetic fields escaping the disk (Miller & Stone, 2000). One of the current important uncertainties in the source of the X-ray emission is its geometry, and possibilities include an accretion disk that evaporates into a hot inner flow (e.g., Shapiro et al., 1976; Zdziarski et al., 1999), a hot ionized 'skin' that sandwiches the cold disk (the 'slab' geometry, e.g., Bisnovatyi-Kogan & Blinnikov, 1977; Liang & Price, 1977; Nayakshin, 2000), a combination of a hot inner flow and a corona that sandwiches the disk (e.g., Poutanen et al., 1997; Sobolewska et al., 2004a), or a patchy corona, consisting of a number of hot spots above the accretion disk (e.g., Galeev et al., 1979; Malzac et al., 2001; Sobolewska

et al., 2004b). In addition, radiation pressure from the UV photons may drive an outflow from the disk into the corona (Proga, 2005), coupling the two processes. In highly accreting objects, like quasars ($0.01 \lesssim L_{bol}/L_{Edd} \lesssim 1$, e.g., Woo & Urry, 2002; Vestergaard, 2004; McLure & Dunlop, 2004; Kollmeier et al., 2006), the X-ray plasma geometry is expected to be in the slab geometry, but the evidence for this is by no means conclusive.

Another salient feature of AGN is variability. Aperiodic variability across all wavebands is ubiquitous in AGN, with the most rapid variations occuring in the X-rays (for a review, see Ulrich et al., 1997). However, the physical mechanism that drives the variability is unclear. Recent results from reverberation mapping have shown that the broad emission lines respond to variations in the continuum emission after some time lag (e.g., Peterson et al., 2004), implying that the optical/UV continuum variations are dominated by processes intrinsic to the accretion disk. In addition, a successful model for quasar X-ray variability describes the X-ray variations as being the result of perturbations in the accretion rate that occur outside of the X-ray emitting region and then travel inwards, modulating the X-ray emission (e.g., Lyubarskii, 1997; Mayer & Pringle, 2006; Janiuk & Czerny, 2007). If the optical/UV variations are intrinsic to the accretion disk, then thermal fluctuations appear to be a natural choice for driving the optical/UV variations, as the optical/UV emission is thought to be thermal emission from the accretion disk. The fact that quasars become bluer as they brighten is consistent with a thermal origin (e.g., Giveon et al., 1999; Trèvese et al., 2001; Geha et al., 2003). However, as noted above, the standard α disks are unstable to thermal fluctuations when they are dominated by radiation pressure. As such, the variations in quasar optical flux are a powerful probe of accretion disk physics.

The correlation between SMBH mass and both host galaxy luminosity (e.g.,

Kormendy & Richstone, 1995; Magorrian et al., 1998; McLure & Dunlop, 2001; Marconi & Hunt, 2003) and stellar velocity dispersion (M_{BH}–σ relationship, e.g., Gebhardt et al., 2000a; Merritt & Ferrarese, 2001; Tremaine et al., 2002), together with the fact that quasars have been observed to reside in early-type galaxies (McLure et al., 1999; Kukula et al., 2001; McLeod & McLeod, 2001; Nolan et al., 2001; Percival et al., 2001; Dunlop et al., 2003), implies that the evolution of spheroidal galaxies and quasars is intricately tied together (e.g., Silk & Rees, 1998; Haehnelt & Kauffmann, 2000; Merritt & Poon, 2004; Di Matteo et al., 2005; Hopkins et al., 2006a). Recently, there have been attempts to link the evolution of SMBHs to analytic and semi-analytic models of structure formation (e.g., Kauffmann & Haehnelt, 2000; Hatziminaoglou et al., 2003; Bromley et al., 2004), where black holes grow by accreting gas funneled toward the center during a galaxy merger until feedback energy from the SMBH expels gas and shuts off the accretion process (e.g., Silk & Rees, 1998; Fabian, 1999; Wyithe & Loeb, 2003; Begelman & Nath, 2005). This 'self-regulated' growth of black holes has recently been successfully applied in smoothed particle hydrodynamics simulations (Di Matteo et al., 2005; Springel et al., 2005). Within this framework, the AGN or quasar phase occurs during the episode of significant accretion that follows the galaxy merger, persisting until feedback from the black hole 'blows' the gas away (e.g., Hopkins et al., 2006b). Alternatively, mergers alone do not appear to be enough to reproduce the number of faint AGN (e.g., Marulli et al., 2007), and accretion of ambient gas (e.g., Ciotti & Ostriker, 2001; Hopkins & Hernquist, 2006), may fuel these fainter, lower M_{BH} AGN at lower z, resulting in a different growth mechanism for these SMBHs.

Quasar feedback has also been invoked in areas outside of black hole growth (for example, see the discussion by Elvis, 2006). Quasar feedback may quench

star-formation in mergers of gas-rich galaxies by reheating the cold gas and forming winds (e.g., Silk & Rees, 1998; Di Matteo et al., 2005). As noted by Elvis (2006), powerful winds generated by AGN feedback may contribute to the enrichment of the intergalactic medium (IGM) seen in both the Lyman-α forest (e.g., Pettini, 2004) and the 'X-ray forest' (e.g., Nicastro et al., 2005). Recent observational work has found evidence for both the quenching of star formation (e.g., Bundy et al., 2006) and the existence of high velocity ($\sim 1000 \text{ km s}^{-1}$) winds in post-starburst galaxies (Tremonti et al., 2007). In addition, quasar winds may provide a source of dust (Elvis et al., 2002a). This is especially important at high redshift, where dust is expected to be important for enhanced star formation (Hirashita & Ferrara, 2002). Indeed, large amounts of dust are seen in high-z quasars (e.g., Omont et al., 2001; Jiang et al., 2006b; Wang et al., 2008), and how dust can form this early is a significant problem (Edmunds & Eales, 1998). Formation of dust through a quasar wind provides a possible explanation. Finally, mechanical feedback via a quasar jet has been invoked to simultaneously explain the lack of massive elliptical galaxies and the halting of cooling flows (e.g., Croton et al., 2006).

Feedback from the central quasar can be injected through both radiative processes and mechanical processes, such as a jet or wind. Figure 1.1 illustrates the effects of quasar feedback. Considering the potential importance of AGN feedback in regulating the growth and evolution of the SMBH and host galaxy, as well as the evolution and enrichment of the surrounding IGM, it is of significant interest to understand how the quasar SED depends on the structure of the accretion disk system, black hole mass, and accretion rate. Hydrodynamic calculations of accretion flows have shown that the efficiency of the quasar in driving an outflow depends on the fraction of energy emitted through he UV/disk component as compared to the X-ray/corona component (Proga, 2007). The disk component

produces luminosity in the UV, which is responsible for driving an outflow via radiation pressure on lines, whereas the corona component produces luminosity in the X-rays, which is responsible for driving an outflow via thermal expansion. Calculations by Proga (2007) have shown that radiation driving produces an outflow that carries more mass and energy than thermal driving. If the efficiency of black hole feedback depends on the quasar SED, any dependence of the quasar SED on M_{BH} and the accretion rate relative to Eddington, $\dot{m}$, has potentially important consequences for models of quasar feedback, black hole growth, and galaxy evolution. In particular, if quasars become more X-ray quiet at high $\dot{m}$ or M_{BH}, then it may become more difficult to grow the most massive SMBHs.

Unfortunately, as is often the case in astronomy, there is a large gap between what we can do theoretically, and what is possible observationally. Quasars are certainly no exception to this, and indeed, in many ways the situation is more difficult for these objects. It is often cynically noted by those outside of the field of AGN research that although we have known of these objects for over 50 years, we still don't really understand how they work. Considering that all of our knowledge of quasars is derived from the photons we detect, and that it is often difficult to obtain these photons, especially for the most interesting bands (i.e., the ultraviolet and X-ray), it is not surprising that it has been difficult for observations to shed light on the physics governing these complex objects. In spite of this, progress has been made in understanding the quasar engine, and intensive studies of some individual bright objects at low-z have been possible. Unfortunately, intensive studies are significantly more difficult for the fainter, higher redshift AGN with larger M_{BH}. Instead, progress has often been made through the analysis of the distribution of quasar luminosities, or through correlations involving various summaries of quasar emission, such as the ratio of flux in two different
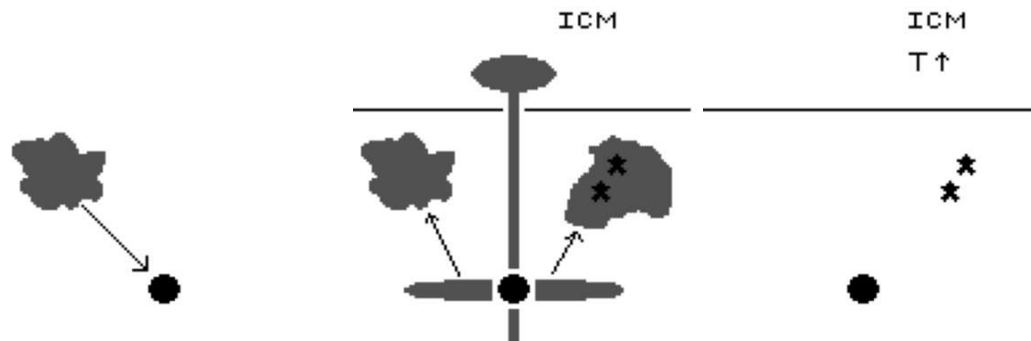
Figure 1.1 Diagram illustrating quasar feedback. First, matter is funneled toward the black hole and accreted by it (left). The accreted material then generates radiative feedback via the conversion of gravitational potential energy to radiation, and mechanical feedback via the formation of an outflow, such as a jet or wind (middle). The radiative feedback injects energy into the surrounding ISM, which heats and expels the accreting gas, halting the black hole's growth (right). In addition, radiative feedback heats the ambient ISM, quenching star formation, and mechanical feedback from a jet increases the temperature of the ICM, halting cooling flows and the accretion of gas onto massive elliptical galaxies.

bands.

Numerous correlations involving quasar properties have been found in previous work, and here I only summarize the most salient of those relevant to this thesis. Further information on previous work can be found in the introductions to each chapter. First, it has been noted by numerous authors that radio-quiet quasars become more X-ray quiet as their optical/UV luminosity increases (e.g., Avni & Tananbaum, 1986; Wilkes et al., 1994; Bechtold et al., 2003; Strateva et al., 2005; Just et al., 2007, and references therein). In addition, a number of previous studies have found evidence that quasars become less variable as their luminosity increases (e.g., Hook et al., 1994; Giveon et al., 1999; Vanden Berk et al., 2004; de Vries et al., 2005, and references therein), and more variable at increasing rest frame wavelength (e.g., Cutri et al., 1985; di Clemente et al., 1996; Helfand et al., 2001; Vanden Berk et al., 2004). While correlations among luminosity measures are important, ideally one would like to investigate how quasar emission depends on more fundamental parameters, such as black hole mass and Eddington ratio.

Perhaps one of the most important advances in observational quasar research over the past decade has been the development of reverberation mapping of quasar broad emission lines. Reverberation mapping (Blandford & McKee, 1982; Peterson, 1993) is often used to estimate SMBH mass, M_{BH}, in Type 1 AGN (e.g., Wandel et al., 1999; Peterson et al., 2004). The basic idea behind this technique is measure the time lag between variations in the emission line flux and in the continuum flux. This time lag is then used to estimate the distance between the gas emitting the broad emission lines, and the central continuum source (i.e., the accretion disk). Under the assumption that the broad line region (BLR) is in Keplerian motion, the time lag is combined with the line width to give an estimate

of SMBH mass. This, in principal, makes it immediately applicable for both faint and distant quasars. Although there are many potential systematic uncertainties in the technique (Krolik, 2001), there has been good agreement between SMBH masses inferred from reverberation mapping and those inferred from the stellar velocity dispersion (Gebhardt et al., 2000b; Ferrarese et al., 2001; Nelson et al., 2004; Onken et al., 2004).

Unfortunately, reverberation mapping requires long-term intensive monitoring, which is not practical for large samples. Fortunately, a correlation has been found that relates the BLR size and monochromatic continuum luminosity (the R–L relationship, Kaspi et al., 2000; Bentz et al., 2006), making it possible to estimate virial masses from single-epoch spectra (e.g., Wandel et al., 1999; Vestergaard, 2002; McLure & Jarvis, 2002; Vestergaard & Peterson, 2006). This therefore opens up the possibility of not only investigating how quasar emission features depend on fundamental parameters, such as M_{BH}, but also of mapping the growth of SMBHs in the quasar phase. However, broad line estimates of M_{BH} based on single-epoch spectra can have significant statistical uncertainties (~ 0.4 dex, e.g., Vestergaard & Peterson, 2006), a fact which must be kept in mind when attempting to derive information from these estimates.

Motivated by the potential in recent advances in estimating M_{BH} for helping to further our understanding of various outstanding issues in quasar research, I have attempted in this thesis to place some observational constraints on the structure and physics of the quasar accretion disk, the distribution and growth of SMBHs in AGN, and how quasar structure is important within the larger context of AGN feedback and galaxy evolution. Specifically, I have attempted to answer, at least partially, the following questions:

- **How does the statistical uncertainty in estimates of M_{BH} derived from the**

broad emission lines affect conclusions based on these estimates? Are mass estimates derived from the Hβ and C IV emission lines consistent? In Chapter 2 I describe a formalism that describes how to calculate virial estimates of M_{BH} from single-epoch spectra of broad line AGN, with an emphasis on the role that uncertainty in M_{BH} plays in drawing conclusions from these estimates. I also show that mass estimates derived from the Hβ and C IV emission lines are consistent.

- **How does one correct for the significant measurement errors that can bias the analysis of correlations, especially in the presence of non-detections?** In Chapter 3 I describe the effects that measurement error can have on observed correlations, and the biases that can result from improperly correcting for measurement error. I also derive a Bayesian approach to linear regression that is able to correct for measurement error in both the dependent and independent variable, even when some of the values of the dependent variable are only upper or lower limits.

- **How does one estimate the distribution of luminosities for a population of astronomical objects in the presence of a selection function? Similarly, how does one derive the distribution of M_{BH} from a set of estimates of M_{BH} derived from the broad emission lines, when one's sample is selected based on flux densities?** In Chapter 4 I describe a flexible Bayesian approach to estimating luminosity functions, and in Chapter 5 I extend this approach to also estimate the broad line quasar black hole mass function. The approach is able to correct for a selection function and the large intrinsic uncertainties in the broad line estimates of M_{BH}. I conclude these Chapters by applying the technique to sources from the Bright Quasar Survey (BQS,

Schmidt & Green, 1983) in order to estimate the local mass function of broad line AGN.

- **How does the quasar optical–UV–X-ray SED correlate with luminosity? Does the quasar SED evolve? How does the quasar SED depend on M_{BH} and $\dot{m}$, and what does this imply about the disk/corona system and quasar feedback?** In Chapter 6 I use new *Chandra* observations of faint high-z radio-quiet quasars, in combination with a $z < 4$ comparison sample drawn from the Sloan Digital Sky Survey (SDSS), to investigate the dependence of X-ray loudness on UV luminosity and redshift. In addition, I also investigate the dependence of the spectral slope of the X-ray continuum, Γ_X, on luminosity and redshift. In Chapter 7 I investigate the dependence of X-ray loudness and Γ_X on M_{BH} and Eddington ratio, and discuss the implications for quasar feedback and the structure of the disk/corona system. I also argue that the observed evolution in the optical–UV–X-ray SED is likely driven by the SED's dependence on M_{BH}.

- **What is the nature of quasar optical flux variations? Is there any evidence for a characteristic time scale of the optical variations, and do the properties of quasar variability depend on M_{BH} or Eddington ratio? What does this tell us about the physical properties of the accretion disk?** In Chapter 8 I model the optical light curves of quasars as a stochastic process, parameterized with a characteristic time scale and amplitude of short time scale variability. I compare the characteristic time scale and magnitude of AGN optical variations with luminosity, redshift, M_{BH}, and Eddington ratio. I interpret the quasar optical variations as being the result of thermal fluctuations in the accretion disk, which are in turn driven by a different

stochastic process with characteristic time scale much longer than the thermal time scale. I also argue that this stochastic model is able to explain both short and long time scale optical flux variations.

Although I summarize the conclusions for each chapter at the end of that chapter, in Chapter 9 I also summarize some of most important points from this thesis. In addition, in Chapter 9 I describe directions for my current and planned future work, which builds on my work described in this thesis.

CHAPTER 2

VIRIAL MASSES OF BLACK HOLES FROM SINGLE EPOCH SPECTRA OF AGN

2.1 CHAPTER ABSTRACT

We describe the general problem of estimating black hole masses of AGN by calculating the conditional probability distribution of M_{BH} given some set of observables. Special attention is given to the case where one uses the AGN continuum luminosity and emission line widths to estimate M_{BH}, and we outline how to set up the conditional probability distribution of M_{BH} given the observed luminosity, line width, and redshift. We show how to combine the broad line estimates of M_{BH} with information from an intrinsic correlation between M_{BH} and L, and from the intrinsic distribution of M_{BH}, in a manner that improves the estimates of M_{BH}. Simulation was used to assess how the distribution of M_{BH} inferred from the broad line mass estimates differs from the intrinsic distribution, and we find that this can lead to an inferred distribution that is too broad. We use these results and a sample of 25 sources that have recent reverberation mapping estimates of AGN black hole masses to investigate the effectiveness of using the C IV emission line to estimate M_{BH} and to indirectly probe the C IV region size–luminosity (R–L) relationship. A linear regression of $\log L_\lambda(1549\text{\AA})$ on $\log M_{BH}$ found that $L_{1549} \propto M_{BH}^{1.17\pm0.22}$. A linear regression also found that $M_{BH} \propto L_{1549}^{0.41\pm0.07} FWHM_{CIV}^2$, implying a C IV R–L relationship of the form $R_{CIV} \propto L_{1549}^{0.41\pm0.07}$. Including the C IV line $FWHM$ resulted in a reduction of a factor of $\sim 1/3$ in the error in the estimates of M_{BH} over simply using the continuum luminosity, statistically justifying its use. We estimated M_{BH} from both C IV and Hβ for a sample of 100 sources, including new spectra of 29 quasars. We find that the two emission lines give consistent estimates if one assumes $R \propto L_{UV}^{1/2}$

for both lines.

2.2   CHAPTER INTRODUCTION

It is widely accepted that the extraordinary activity associated with quasars involves accretion onto a supermassive black hole (SMBH). Furthermore, the evidence that almost all massive galaxies host SMBHs has become quite convincing. It has been found that SMBH mass is correlated with the host galaxy's bulge luminosity (e.g., Kormendy & Richstone, 1995; Magorrian et al., 1998; McLure & Dunlop, 2001; Marconi & Hunt, 2003) as well as the stellar velocity dispersion (e.g., Gebhardt et al., 2000a; Merritt & Ferrarese, 2001; McLure & Dunlop, 2002; Tremaine et al., 2002). Because luminous quasars have been observed to reside in massive early-type galaxies (McLure et al., 1999; Kukula et al., 2001; McLeod & McLeod, 2001; Nolan et al., 2001; Percival et al., 2001; Dunlop et al., 2003), this implies that the evolution of spheroidal galaxies and quasars is intricately tied together (e.g., Silk & Rees, 1998; Haehnelt & Kauffmann, 2000; Adams et al., 2001; Merritt & Poon, 2004; Di Matteo et al., 2005). Therefore, understanding the cosmic evolution of SMBHs is an important task of modern astronomy.

Reverberation mapping (Blandford & McKee, 1982; Peterson, 1993) is often used to estimate SMBH mass, M_{BH}, in Type 1 active galaxies (Wandel et al., 1999; Kaspi et al., 2000; Peterson et al., 2004). One of the principal advantages of this method is that it does not require high spatial resolution, but rather relies on the time lag between the continuum and emission line variability. Under the assumption that the broad line region (BLR) is in Keplerian motion, the time lag is combined with the line width to give an estimate of SMBH mass. This, in principal, makes it immediately applicable for both faint and distant quasars. Although there are many potential systematic uncertainties in the technique (Krolik, 2001),

there has been good agreement between SMBH masses inferred from reverberation mapping and those inferred from the stellar velocity dispersion (Gebhardt et al., 2000b; Ferrarese et al., 2001; Nelson et al., 2004; Onken et al., 2004).

Unfortunately, reverberation mapping requires long-term intensive monitoring, which is not practical for large samples. In addition, the long time scales for variability in bright high redshift sources make reverberation mapping for these sources unfeasable. Fortunately, a correlation has been found that relates the BLR size and monochromatic continuum luminosity (the R–L relationship, Kaspi et al., 2000, 2005), making it possible to estimate virial masses from single-epoch spectra using the Hβ width (Wandel et al., 1999). In addition, Wu et al. (2004) find a relationship between BLR size and the luminosity of Hβ, but Woo & Urry (2002) do not find any correlation between SMBH mass and bolometric luminosity. Hβ is redshifted into the near-infrared at $z \sim 0.9$, making it difficult to observe from the ground for large samples. Vestergaard (2002) and McLure & Jarvis (2002) have argued for the use of C IV and Mg II, respectively, to estimate virial masses from single-epoch spectra, allowing quasar SMBH masses to be estimated at high z from the ground. Many studies have exploited these results and estimated SMBH masses for large samples of quasars (e.g., Bechtold et al., 2003; Corbett et al., 2003; Warner et al., 2003; Vestergaard, 2004; McLure & Dunlop, 2004; McLure & Jarvis, 2004). Dietrich & Hamann (2004) have found that the estimates based on Hβ and C IV agree well for high z quasars, whereas Mg II-based estimates are typically a factor of ~ 5 times lower. In constrast to this, Shemmer et al. (2004) and Baskin & Laor (2005) argue that C IV does not give as accurate of an estimate of SMBH mass as Hβ.

Previous methods that utilize single-epoch spectra have relied on empirical linear relationships that estimate the BLR size for a given source luminosity, and

then use this estimate of R in the virial relationship. Although this will give good estimates for R, this may give less efficient estimates for SMBH mass. The reason for this is that the luminosity may also be correlated with M_{BH} in a manner independent of the R–L relationship, such as through the accretion process. By only using the luminosity as a proxy for R when estimating M_{BH}, one is ignoring the additional information of M_{BH} that is contained within L. In other words, the standard broad line estimates are based on the distribution of velocities at a given luminosity and M_{BH}, as folded through the R–L correlation and under the assumption of Keplerian motion. However, the distribution of black hole masses at a given luminosity and line width is not simply obtained by inverting the distribution of velocities at a given luminosity and black hole mass, unless L and M_{BH} are statistically independent.

Recent studies have found evidence for a correlation between luminosity and black hole mass (e.g., Corbett et al., 2003; Netzer, 2003; Peterson et al., 2004), suggesting that black hole mass and luminosity are not statistically independent. However, in contrast to other studies, Woo & Urry (2002) have argued that there is no significant correlation between bolometric luminosity and M_{BH}. If a correlation between M_{BH} and L exists, then we can combine the M_{BH}–L correlation with the broad line mass estimates to obtain, on average, more accurate estimates of M_{BH} for a given luminosity and line width.

In this work, we outline a formalism that allows one to estimate the probability distribution of an AGN's black hole mass, given some set of observables. We focus on the special case of estimating M_{BH} given some monochromatic luminosity, L_λ, and the width of an emission line. We also search for other parameters of the C IV line that may contribute additional information of M_{BH}, thus decreasing the uncertainty in the estimated M_{BH}. Although it is possible to include other

quasar properties, such as radio loudness, X-ray loudness, or variability (e.g., Xie et al., 2005; O'Neill et al., 2005; Pessah et al., 2006), for simplicity we have chosen to only include parameters that may be measured from a single spectrum containing C IV. This allows M_{BH} to be estimated using only one spectrum, and it is thus not necessary to compile observations from several different spectral regions.

We use the formalism developed here to justify using the C IV line width in estimating M_{BH}, and attempt to indirectly infer the C IV R–L relationship. We have chosen the C IV line because it is readily observable from the ground over a wide range in redshift ($1.5 \lesssim z \lesssim 4.5$), is less effected by blends with iron and other lines, has shown to give consistent mass estimates with Hβ (Vestergaard, 2002), is commonly employed to estimate SMBH mass, and archival UV spectra are available for most of the sources with reverberation-based masses (Peterson et al., 2004). Despite its common usage, the R–L relationship for C IV is mostly unexplored, as there are only seven data points with reliable C IV reverberation mapping data (Peterson et al., 2005). Often it is assumed that C IV BLR size has the same dependence on luminosity as Hβ (e.g., Vestergaard, 2002; Netzer, 2003; Vestergaard & Peterson, 2006). Another possibility is to assume $R \propto L^{1/2}$, as predicted from simple photoionization theory or if the BLR size is set by the dust sublimation radius (Netzer & Laor, 1993); this was done by Wandel et al. (1999) and Shields et al. (2003) for the Hβ line. Peterson et al. (2005) performed a linear regression using five AGN with a total of seven data points and find $R \propto L_{UV}^{0.61 \pm 0.05}$, similar to that for the Balmer lines[1]. Unfortunately, this result is almost entirely dependent on the inclusion of NGC 4395, the least luminous known Seyfert 1 galaxy, since the data points for the other AGN are clustered

---

[1]This value appears in an erratum to this paper

around $\lambda L_\lambda(1350\mathring{A}) \approx 10^{43.75}$ ergs s^{-1}. Vestergaard & Peterson (2006) found that using the C IV line to calculate black hole masses assuming $R \propto L_{UV}^{0.53}$ gave results consistent with masses obtained by reverberation mapping.

The layout of the paper is as follows. In § 2.3 we describe the general problem of estimating SMBH mass from single-epoch spectra, and in § 2.4 we describe our two samples. The first sample is a set of 25 quasars with black holes from reverberation mapping and archival UV spectra, and the second sample is a set of 100 quasars for which we have spectra containing both the Hβ and C IV emission lines. In § 2.5 we describe the method we employ to estimate the emission line parameters. In § 2.6 we test if any other UV or C IV parameters contribute useful information about M_{BH}. In § 2.7.1 we use the sample with reverberation mapping data to investigate the M_{BH}–L relationship, and in § 2.7.2 we test if including the C IV line $FWHM$ is preferred by this sample and investigate the nature of the C IV R–L relationship. In § 2.8 we use our larger sample to compare estimates of M_{BH} obtained from both single-epoch Hβ and C IV. We summarize our results in § 2.9.

In this work we adopt the WMAP best-fit cosmological parameters ($h = 0.71, \Omega_m = 0.27, \Omega_\Lambda = 0.73$, Spergel et al., 2003). We will use the common statistical notation where a point estimate of a quantity is denoted by placing a ˆabove it, e.g., $\hat{M}_{BH}$ would be an estimate of M_{BH}. We will also commonly refer to the bias of an estimate, where the bias of $\hat{M}_{BH}$ is $Bias = E(\hat{M}_{BH}) - M_{BH}$. Here, $E(\hat{M}_{BH})$ is the expectation value of $\hat{M}_{BH}$. An unbiased estimate of M_{BH} is one such that $E(\hat{M}_{BH}) = M_{BH}$.

2.3 ESTIMATING BLACK HOLE VIRIAL MASSES

If one assumes that the BLR gas is in Keplerian motion, then the mass of the central black hole, M_{BH}, may be estimated from the virial theorem :

$$M_{BH} = f \frac{RV^2}{G}.$$ (2.1)

Here R is the distance between the BLR that is emitting a particular line and the central continuum source, V is the velocity dispersion of the line-emitting gas, and G is the gravitational constant. The velocity dispersion is inferred from the width of the line, quantified using either the $FWHM$ or line dispersion, σ_*. The factor f is a scaling factor that converts the virial product, RV^2/G, to a mass. Typically, f has been set to the value appropriate for an isotropic velocity field, where $f = 3/4$ if one uses the broad line $FWHM$ to estimate V (e.g., Netzer, 1990a); however, there may be systematic effects that can significantly effect the value of f (Krolik, 2001). Onken et al. (2004) used the correlation between M_{BH} and stellar velocity dispersion to estimate an average scale factor of $\langle f \rangle = 5.5$ when the line dispersion is used to estimate V. For simplicity, in the rest of this work we will assume $f = 1$, so that what is really being estimated is the virial product. After estimating the virial product, we can convert it to a mass using any adopted value of f.

2.3.1 Estimating M_{BH} from Reverberation Mapping

For the case of reverberation mapping, estimating M_{BH} is straightforward. The BLR size, R, can be estimated as $c\tau$, where c is the speed of light and τ is either the peak or the centroid of the line–continuum cross-correlation function. The velocity dispersion of the line-emitting gas, V, is estimated from the width of the broad emission line as measured in the variable part of the spectrum. Measuring V from the variable (RMS) spectrum ensures that one is probing the line emission

that is actually varying, i.e., the BLR gas that is at the distance $c\tau$. One then uses $c\tau$ and V in Equation (2.1) to estimate M_{BH}. See Peterson et al. (2004) for recent reverberation mapping results.

2.3.2   Estimating M_{BH} from Single-Epoch Spectra

Estimating M_{BH} from a single-epoch spectrum (SES) is a somewhat different problem for several reasons. In this section we illustrate why the SES case is different, and provide a general formalism for calculating the conditional probability distribution of M_{BH}, given a set of observables.

In the single-epoch case, one cannot directly observed the BLR size, but instead employs a correlation between R and continuum luminosity. However, there is considerable scatter in the R–L relationship, which is propogated through when using L instead of R to estimate M_{BH}. In this case, the conditional probability distribution of M_{BH}, given the broad line estimate, is broad, typically a factor of a few (Vestergaard, 2002; Kaspi et al., 2005). This scatter can be reduced by incorporating information about M_{BH} from other observables. In the more general case, one may have several parameters that contain information about M_{BH}, such as emission line width, variability, luminosity, etc. For example, Merloni et al. (2003) have found evidence that black hole mass is correlated with X-ray and radio luminosity. In this case, one could combine the information from the line width, UV luminosity, X-ray luminosity, and radio luminosity to obtain an estimate of M_{BH} that is more accurate than would have be obtained solely from some subset of these parameters.

To be more specific, suppose that one has a set of observables, denoted by X. Then, given the observables, X, the conditional probability of M_{BH} given X is

given by Bayes' Formula

$$p(M_{BH}|X) = \frac{p(X|M_{BH})p(M_{BH})}{p(X)}, \tag{2.2}$$

where $p(M_{BH})$ is the intrinsic probability distribution of M_{BH} for a sample, and $p(X)$ is the distribution of the observables; $p(X)$ is just a normalizing constant and may be ignored.

In this work we are concerned with the distribution of M_{BH} given L, the emission line width, and redshift. For this case, Equation (2.2) becomes

$$p(m|l, v, z) \propto p(v|l, m, z)p(l|m, z)p(m|z). \tag{2.3}$$

Here, we are using the notation, $m \equiv \log M_{BH}$, $l \equiv \log \lambda L_\lambda$, and $v \equiv \log V$. The first term, $p(v|l, m, z)$, is the distribution of line widths at a given l, m, and z. The second term, $p(l|m, z)$, is the distribution of luminosities at a given m and z. The last term, $p(m|z)$, is the distribution of m at a given redshift. As we will show later in this section, $p(v|l, m, z)$ is obtained by plugging the R–L relationship into the Virial theorem (Eq.[2.1]). This is the standard method of estimating M_{BH} from the broad lines, but it implicitly assumes $p(m|l, v, z) \propto p(v|l, m, z)$, and therefore that $p(m|z)$ and $p(l|m, z)$ are uniform. Taking $p(m|z)$ and $p(l|m, z)$ to be uniform results in a broader distribution of M_{BH}, given L, V, and z, and thus a less efficient, albeit still unbiased, estimate of M_{BH}. However, by incorporating information on both the distribution of luminosities at a given black hole mass and redshift, and the distribution of M_{BH} at a given redshift, one can obtain a better estimate of M_{BH}.

To estimate a functional form for $p(l|m)$ and $p(v|l, m, z)$, suppose we observe some quasar with SMBH mass M_{BH} at a redshift z, where m is drawn from some probability density $p(m|z)$, $m|z \sim p(m|z)$. We assume that the accretion process for this source generates a luminosity, $L_\lambda \propto M_{BH}^{\alpha_m(z)}$, by

$$l|m, z = \alpha_0 + \alpha_m(z)m + \epsilon_l(z, m). \tag{2.4}$$

Here, α_0 is some constant of proportionality, and $\epsilon_l(z, m)$ is the random error term representing the scatter about this relationship. We will refer to Equation (2.4) as the M_{BH}–L relationship. The stochastic term, $\epsilon_l(z, m)$, encompasses variations at a given M_{BH} in the accretion rate ($\dot{M}$), radiative efficiency (ϵ), source inclination, bolometric correction (C_{bol}), etc.; since we do not observe these quantities we model them as being random. For now, we allow the logarithmic slope, α_m, to depend on z, and the random error to depend on z and M_{BH}.

The parameter α_m can be predicted from accretion physics. The radiated bolometric luminosity from an accretion flow can be written as

$$L = (1.26 \times 10^{31}) \epsilon \dot{m} \frac{M_{BH}}{M_\odot} \text{ W}, \tag{2.5}$$

where $\dot{m} = \dot{M}/\dot{M}_{edd}$ is the accretion rate normalized to Eddington. If one assumes $L_\lambda = C_{bol}L$, then from Equation (2.5) it follows that for this case $\alpha_m = 1$ and $\epsilon_l = \log \epsilon + \log \dot{m} + \log C_{bol}$. A more careful analysis of the thin disk case suggests $\lambda L_\lambda \propto (M_{BH}\dot{M})^{\alpha_m}, \alpha_m = 2/3$, after employing some simplifying approximations (e.g., Collin et al., 2002).

Given this luminosity, the BLR distance R is assumed to be set by the luminosity according to the R–L relationship, $R \propto L^{\theta_l}$:

$$r|l, z = \theta_0 + \theta_l l + \epsilon_r(z). \tag{2.6}$$

Similar to above, $r \equiv \log R$ and $\epsilon_r(z)$ is the stochastic component. In Equation (2.6) we assume that given L, R is independent of M_{BH}. In addition, we have not assumed any redshift dependence for θ_l because it is likely that the form of the R–L relationship is independent of z (Vestergaard, 2004). However, in Equation (2.6) we have allowed for the possibility of a redshift dependence for ϵ_r as the the intrinsic scatter in the R–L relationship may depend on z.

Simple photoionization theory predicts that $R \propto L_{ion}^{1/2}$, where L_{ion} is the luminosity of the ionizing continuum (e.g., Wandel et al., 1999; Kaspi et al., 2000). From the definition of the ionization parameter, U, it follows that

$$2r = \log L_{ion} - \log(4\pi c) - \log U - \log n_e - \log \bar{E}, \qquad (2.7)$$

where $\bar{E}$ is the average energy of an ionizing photon and n_e is the BLR gas density. If we make the simplifying assumptions that r is set by Equation (2.7), $L_\lambda \propto L_{ion}$, and the means of the distributions of U, n_e, and $\bar{E}$ are independent of L_λ, then comparison with Equation (2.6) shows that $\theta_l = 1/2$ and $\epsilon_r = -\log U - \log n_e - \log \bar{E}$. If this is not the case, but rather the means of the distributions of U, n_e, and $\bar{E}$ have a power-law dependence on L_{ion}, then Equation (2.7) is still valid, but in general $\theta_l \neq 1/2$. Either way, in this model the scatter about the R–L relationship is partly the result of variations in U, n_e, and $\bar{E}$. Other sources of scatter may include variations in the conversion between L_λ and L_{ion}, source inclination, and the non-instantaneous response of the BLR to continuum variations.

If R is set by the dust sublimation radius, then we also expect $\theta_l = 1/2$ (Netzer & Laor, 1993), but in this case $R \propto L^{1/2}$. A relationship of the form $R \propto L_\lambda^{1/2}$ is consistent with the results of Peterson et al. (2004) for the Balmer lines if one uses the UV continuum luminosity.

Finally, from Equation (2.1) the observed SES line width depends on R and M_{BH} as

$$v|r, m = v_0 - \frac{1}{2}r + \frac{1}{2}m + \epsilon_v, \qquad (2.8)$$

where, $v_0 = \log(\sqrt{G}/f_{SES})$, and ϵ_v is the stochastic term. The term f_{SES} converts the SES line width measurement into a velocity dispersion, and ϵ_v describes random deviations of single-epoch v from that for the RMS spectrum. Vestergaard (2002) has found that the single-epoch Hβ $FWHM$ and the RMS Hβ $FWHM$

are consistent with a scatter of $\sim 20\%$, if one does not subtract the SES Hβ narrow component. When the $FWHM$ is used, $f_{SES} \approx 1/2$. Through the stochastic terms ϵ_l, ϵ_r, and ϵ_v, Equations (2.4), (2.6), and (2.8) define the conditional probability densities that describe how L and V depend on M_{BH}, and how R depends on L.

It is useful to examine the special case of Gaussian error terms, ϵ_l, ϵ_r, and ϵ_v, and Gaussian $p(m|z)$. In this case, the distribution of m at a given l and v is also normal. If we assume that ϵ_l, ϵ_r, and ϵ_v are uncorrelated, have zero mean, and variances σ_l^2, σ_r^2, and σ_v^2, respectively, then the optimal estimate of m is the mean of $p(m|l, v, z)$, μ. If we make the further assumption that ϵ_l, ϵ_r, and ϵ_v are independent of m and z, and that α_m does not depend on z, then $p(m|l, v, z)$ takes a particularly simple form. The mean of $p(m|l, v, z)$ may be calculated from the properties of the normal distribution (e.g., Gelman et al., 2004) as

$$\mu = \frac{\hat{m}_{BL}/\sigma_{BL}^2 + \hat{m}_{ML}/\sigma_{ML}^2 + \bar{m}(z)/\sigma^2(z)}{1/\sigma_{BL}^2 + 1/\sigma_{ML}^2 + 1/\sigma^2(z)}. \tag{2.9}$$

Here, $\hat{m}_{BL} = m_0^{BL} + \theta_l l + 2v$ is the standard broad line mass estimate, found by plugging Equation (2.6) into Equation (2.8), $m_0^{BL} = \theta_0 - 2v_0$, $\sigma_{BL}^2 = \sigma_r^2 + 4\sigma_v^2$ is the variance in $\hat{m}_{BL}$, $\hat{m}_{ML} = (l - \alpha_0)/\alpha_m$ is the estimate of m based on the M_{BH}–L relationship, $\sigma_{ML}^2 = \sigma_l^2/\alpha_m^2$ is the variance in $\hat{m}_{ML}$, $\bar{m}(z)$ is the mean m at a given z, and $\sigma^2(z)$ is the variance in m at a given z.

The variance in μ as an estimate for m, σ_μ^2, is

$$\sigma_\mu^2 = \left(\frac{1}{\sigma_{BL}^2} + \frac{1}{\sigma_{ML}^2} + \frac{1}{\sigma^2(z)} \right)^{-1}. \tag{2.10}$$

As is apparent from Equation (2.10), the variance in μ as an estimate for m is always less than that for the standard broad line estimate, $\hat{m}_{BL}$. In fact, in the limit $(\sigma_{ML}^2, \sigma^2(z)) \to \infty$, $p(m|l)$ and $p(m|z)$ supply no information on the black hole mass and μ converges to the broad line estimate. However, by combining the

broad line mass estimate with the information on M_{BH} that is contained within the luminosity and redshift we can obtained a better estimate of m.

In reality, the distribution of $\dot{m}$ is unlikely to be Gaussian, and since $\dot{m}$ is a component of ϵ_l, this would violate the assumption that ϵ_l is Gaussian. Instead, the distribution of $\dot{m}$ is likely bimodal (Ho, 2002; Marchesini et al., 2004; Hopkins et al., 2006b; Cao & Xu, 2007) as a result of a transition from a radiatively inefficient flow to an efficient one at $\dot{m} \sim 0.01$ (e.g., Jester, 2005). However, because of the flux limits of modern surveys, most observed broad line quasars will have $\dot{m} \gtrsim 0.01$ (McLure & Dunlop, 2004; Vestergaard, 2004). This results in a unimodal, relatively smooth and symmetric distribution of $\dot{m}$ for observed quasars (Hopkins et al., 2006a). It may also be that BLRs do not form in sources with $\dot{m} \lesssim 10^{-3}$ (Nicastro et al., 2003; Czerny et al., 2004), and therefore the distribution of $\dot{m}$ for broad line AGN would have a lower limit at $\dot{m} \sim 0.001$. If true, then unimodality in the distribution of $\dot{m}$ for broad line AGN is ensured.

The stochastic scatter about the M_{BH}–L relationship, ϵ_l, is the sum of random deviations in $\dot{m}$, bolometric correction, radiative efficiency, etc. We assume that the distributions of most, if not all, of the constituent components of ϵ_l are not too different from a Gaussian distribution, i.e., unimodal, smooth, and fairly symmetric. Then, by the central limit theorem, the distribution of the sum of these components, ϵ_l, will tend toward a normal density. Therefore, without any evidence to the contrary, the normal density should provide an accurate approximation to the true form of $p(l|m)$. A similar argument may be used to justify the assumption of normality for ϵ_r and ϵ_v.

To illustrate the improvement that Equation (2.9) offers over the broad line mass estimate, we simulate values of m, l, and v. The simulations were performed as follows. First, we simulated values of m from a smoothly-connected

double power-law, with a mean of $\bar{m} \approx 7.87$ and a dispersion of $\sigma \approx 0.45$ dex. Values of l were then generated according to Equation (2.4), with $\alpha_0 = 37, \alpha_m = 1$. The gaussian scatter about this relationship had a dispersion of 0.7 dex. Then, we generated values of v as $v = 11 - l/4 + m/2 + \epsilon$, where ϵ was a gaussian random variable with dispersion 0.2 dex. This form for v assumes $R \propto L^{1/2}$, and corresponds to a intrinsic scatter in the broad line mass estimates of 0.4 dex. The parameters for simulating l and v were chosen to be similar to the results found in § 2.7. Finally, we calculated broad line mass estimates from the simulated luminosities and line widths, $\hat{m}_{BL} = -22 + l/2 + 2v$, and mass estimates μ according to Equation (2.9).

The results are shown in Figure 2.1. As can be seen from the distribution of residuals, the mass estimates that combine all available information on the black hole mass, μ, are more accurate on average than the broad line estimates, $\hat{m}_{BL}$. In addition, the distribution of μ provides a more accurate estimate of the true distribution of m than does $\hat{m}_{BL}$, with the distribution of m inferred from $\hat{m}_{BL}$ being too broad. It is interesting to note that both of these results are in spite of the fact that the intrinsic distribution of m is not Gaussian, which was assumed when deriving Equation (2.9). This suggests that if the intrinsic distribution of m for a sample is not too different from a normal density, Equation (2.9) will still give more efficient estimates than the broad line estimates. This is reasonable, considering that Equation (2.9) 'shrinks' the black hole mass estimates towards the sample mean by an amount inversely proportional to the intrinsic variance in m of a sample.

2.3.3 Cautions for Using Quantities Calculated from the SES Mass Estimates

The intrinsic uncertainty in M_{BH} inferred from L_λ and the line width may be thought of as the 'measurement' error in M_{BH}. This intrinsic uncertainty can
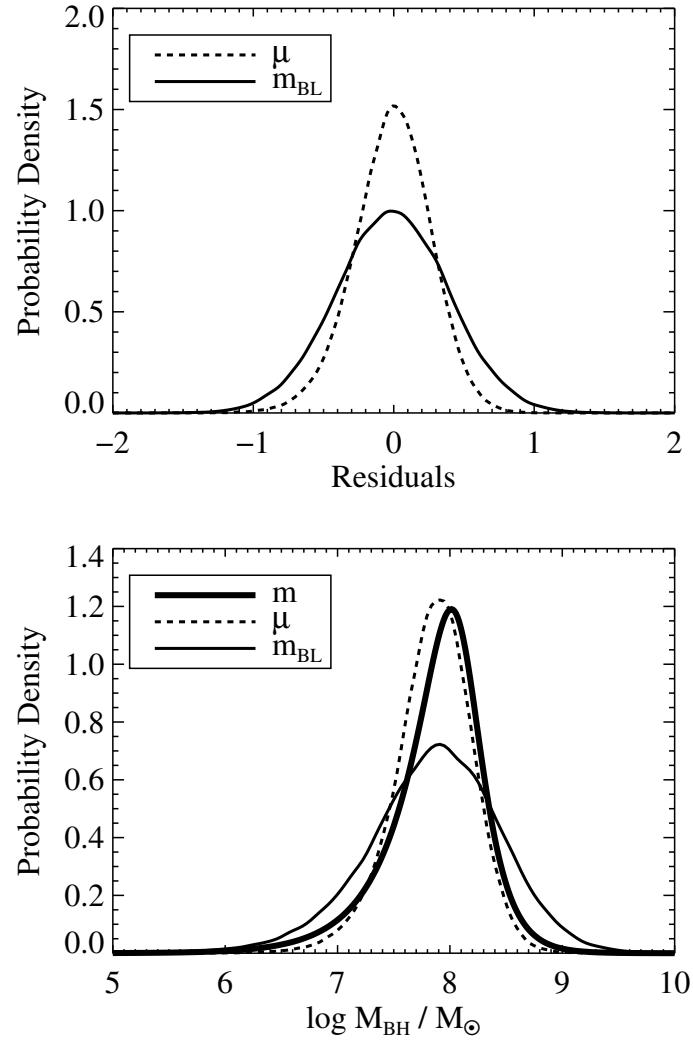
Figure 2.1 Results of the simulations described in § 2.3.2, illustrating the effective-ness of the mass estimate, μ, that combines the broad line mass estimates with the distribution of luminosities at a given black hole mass and the intrinsic distribu-tion of M_{BH}. The upper panel shows the distribution of the residuals when using μ (dashed line) and the broad line mass estimate, $\hat{m}_{BL}$ (solid line). The bottom panel compares the intrinsic distribution of m (thick solid line) with that inferred from the distribution of μ (dashed line) and $\hat{m}_{BL}$ (thin solid line).

cause problems when using the black hole mass estimates, $\hat{m}$, to calculate additional quantities. In particular, quantities based on the square of the estimated $\hat{m}$, such as correlation coefficients and linear regressions, can be significantly effected (e.g., see Akritas & Bershady, 1996; Fox, 1997).

Suppose we are interested in calculating the correlation between M_{BH} and some other parameter X. To estimate this correlation, we would obtain a sample of quasars with black hole masses estimated from some assumed form of $p(m|l, v, z)$, $\hat{m}$. Typically, these are the standard broad line estimates. We then calculate the correlation coefficient between X and $\hat{m}$. However, because we do not have the actual black hole masses for our sample, but instead obtained estimates from the continuum luminosities and widths of the broad lines, this is not the *true* correlation coefficient between X and m. Because our estimated black hole masses have been 'measured' with error, this broadens the observed distribution of black hole masses, and thus biases the observed correlation coefficient towards zero. Therefore, the correlation coefficient obtained from the estimated black hole masses will be, on average, *less* in magnitude than the correlation coefficient that would have been obtained using the actual black hole masses.

To prove this point, we note that the logarithmic black hole mass estimates are related to the actual black hole masses as $\hat{m} = m + \epsilon_m$, where ϵ_m is the random error between $\hat{m}$ and m. The linear correlation between the parameter of interest X and m is $\rho = Covar(X, m)/[Var(X)Var(m)]^{1/2}$, where $Covar$ and Var are the sample covariance and variance, respectively. Since we don't actually observe m, we can't calculate the true correlation coefficient, but instead we calculate the correlation between X and $\hat{m}$, $\hat{\rho}$. While the covariance between X and m is unaffected by using $\hat{m}$ instead of m, the sample variance of $\hat{m}$ is $Var(\hat{m}) = Var(m) + \sigma_{\hat{m}}^2$, where $\sigma_{\hat{m}}^2$ is the intrinsic uncertainty in $\hat{m}$ as an estimate

for m, $\sigma_{\hat{m}}^2 = E(\epsilon_m^2)$. The observed correlation is then given by

$$\hat{\rho} = \left[\frac{Var(m)}{Var(m) + \sigma_{\hat{m}}^2} . \right]^{1/2} \rho \tag{2.11}$$

Therefore, correlation coefficients calculated using the estimated black hole masses will be reduced in magnitude from the true correlation by a factor of $[1 + \sigma_{\hat{m}}^2/Var(m)]^{-1/2}$. For broad line mass estimates based on Hβ, Vestergaard & Peterson (2006) find $\sigma_{\hat{m}} = 0.43$ dex. In this work we find that $\sigma_{\hat{m}} \approx 0.40$ dex for C IV-based broad line estimates . For a sample with an intrinsic dispersion in m of 0.75 dex, these values of $\sigma_{\hat{m}}$ correspond to a decrease in the magnitude of any observed correlation by $\approx 12\%$. If the sample has an intrinsic dispersion of 0.4 dex, similar to the intrinsic uncertainties in the broad line estimates, then the magnitude of the observed correlation coefficient is reduced by $\approx 30\%$. These effects is not negligible, and can be more serious for linear regression (e.g., Fox, 1997). In light of these issues, care must be taken when calculating quantities from black hole mass estimates based on single-epoch spectra.

Another problem arises when one is using a flux limited sample to estimate the intrinsic distribution of the black hole mass, i.e., the active black hole mass function (Wang et al., 2006), based on broad line estimates. There is currently significant interest in this problem, as the active black hole mass function is an important tool in understanding SMBH formation and evolution. Unfortunately, the limiting flux of a survey causes incompleteness in black hole mass, the degree of which depends on the distribution of l at a given m and z, $p(l|m, z)$. In addition, because the broad line mass estimates are measured with error, the distribution of m inferred from the broad lines can be significantly broader than the intrinsic distribution (cf., Fig.2.1). Because of these issues, estimates of the active black hole mass function obtained from broad line estimates should be interpreted with caution.

In order to estimate the degree of incompleteness in m, it is necessary to re-express the survey selection function as a function of m. Following Gelman et al. (2004), we introduce an indicator variable, I, denoting whether a source is included in the survey, where $I = 1$ if a is source included and $I = 0$ if a source is missed. Then the selection function of the survey is the probability that a source is included in the survey for a given luminosity and redshift, $p(I = 1|l, z)$. The selection function for black hole mass is then

$$p(I = 1|m, z) = \int_{-\infty}^{\infty} p(I = 1|l, z)p(l|m, z)dl. \tag{2.12}$$

As can be seen, estimating the completeness in m for a survey depends on the form of $p(l|m, z)$. Therefore, it is important to understand $p(l|m, z)$, even if such an understanding does not result in significantly better estimates of m.

2.4 THE SAMPLE

We employ two samples in our analysis. The first sample consists of a set of 25 low-z sources that have reverberation mapping data from Peterson et al. (2004). We use this sample to investigate the M_{BH}–L relationship and the C IV R–L relationship, and to justify using the C IV line to estimate AGN black hole masses. The second sample is a set of 100 quasars for which we have seperate spectra containing the Hβ and C IV emission lines. We use this sample to compare the broad line estimates of M_{BH} obtained from the two lines.

Peterson et al. (2004) calculated virial products for 35 AGNs based on the reverberation mapping method. Of those 35 sources, we selected ones with archival UV spectra. We did not include those sources which Peterson et al. (2004) listed as having unreliable virial products (PG 0844+349, PG 1211+143, PG 1229+204, and NGC 4593). We also did not include NGC 3227 or NGC 4151, as the C IV line for these sources had significant absorption. In addition, we removed NGC

4051 from the analysis because this source is an outlier in the BLR R-L relationship (Vestergaard, 2002; Kaspi et al., 2005), and thus probably does not follow the linear relationship assumed in Equation (2.6). Our sample consists of 25 AGN: 14 sources with HST FOS spectra, 9 sources with IUE spectra, 1 source with HST GHRS spectra, and 1 source with HST STIS spectra. Of the 14 sources with FOS spectra, 13 were taken from Bechtold et al. (2002), and the other (NGC 5548) from Evans & Koratkar (2004). For archival sources with more than one observation, we took the source with the longest exposure time. All spectra are single-epoch, except for NGC 5548, which is averaged over approximately a month of HST observations. The sample is summarized in Table 2.1.

The luminosities were calculated from the predicted continuum flux at $1549\mathring{A}$, assuming a power law continuum (see $\S$ 2.5.1). We corrected luminosities for galactic absorption using the $E(B-V)$ values taken from Schlegel et al. (1998), as listed in the NASA/IPAC Extragalactic Database (NED), and the extinction curve of Cardelli et al. (1989), assuming a value of $A_V/E(B-V) = 3.1$. We did not do this for the Bechtold et al. (2002) sources as they have already been corrected for galactic absorption. The C IV line widths for the IUE sources were corrected by subtracting an assumed instrumental resolution of 1000 km s^{-1} in quadrature from the measured line widths; the resolution for the other instruments is negligible compared to the emission line widths, so no correction was performed.

We compiled UV and optical spectra for a sample of 100 sources for the purpose of comparing the C IV-based estimates derived here with the estimates based on Hβ and the empirical R–L relationship. Of these sources, 89 have $z < 0.8$, 6 have $z \sim 2.3$, and 2 have $z \sim 3.3$. The UV spectra for the $z < 0.8$ sources are FOS spectra from Bechtold et al. (2002), the optical spectra for 51 of these sources are from Marziani et al. (2003), and the optical spectra for 9 of these

Table 2.1. Object List

Object	RA (J2000)	DEC (J2000)	Redshift	$\log(M_{BH}/M_\odot)$[a]	Rad. Type[b]	Instrument
MRK 335	00 06 19.5	+20 12 10.5	0.025	6.41 ± 0.11	Quiet	FOS
PG 0026+129	00 29 13.7	+13 16 03.8	0.142	7.85 ± 0.10	Quiet	FOS
PG 0052+251	00 54 52.1	+25 25 39.3	0.155	7.82 ± 0.08	Quiet	FOS
FAIRALL9	01 23 45.7	-58 48 21.8	0.046	7.66 ± 0.09	Quiet	FOS
MRK 590	02 14 33.6	-00 46 00.1	0.027	6.93 ± 0.06	Quiet	IUE
3C 120	04 33 11.1	+05 21 15.6	0.033	7.00 ± 0.21[c]	Loud	IUE
ARK 120	05 16 11.4	-00 08 59.4	0.033	7.43 ± 0.05	Quiet	FOS
MRK 79	07 42 32.8	+49 48 34.7	0.0221	6.97 ± 0.11	Quiet	IUE
PG 0804+761	08 10 58.6	+76 02 42.0	0.1000	8.10 ± 0.05	Quiet	IUE
MRK 110	09 25 12.9	+52 17 10.5	0.0352	6.65 ± 0.10	Quiet	IUE
PG 0953+414	09 56 52.4	+41 15 23.0	0.2341	7.69 ± 0.09	Quiet	FOS
NGC 3516	11 06 47.5	+72 34 06.9	0.0088	6.88 ± 0.14	Quiet	FOS
NGC 3783	11 39 01.7	-37 44 18.9	0.009	6.73 ± 0.07	Quiet	FOS
3C 273.0	12 29 06.7	+02 03 09.0	0.1583	8.20 ± 0.09	Loud	FOS
PG 1307+085	13 09 47.0	+08 19 49.8	0.1550	7.90 ± 0.12	Quiet	FOS
MRK 279	13 53 03.4	+69 18 29.6	0.0304	6.80 ± 0.26	Quiet	STIS
NGC 5548	14 17 59.5	+25 08 12.4	0.0171	7.08 ± 0.01	Quiet	FOS
PG 1426+015	14 29 06.6	+01 17 06.5	0.0864	8.37 ± 0.12	Quiet	IUE
MRK 817	14 36 22.1	+58 47 39.4	0.0314	6.95 ± 0.06	Quiet	IUE
PG 1613+658	16 13 57.2	+65 43 09.6	0.1290	7.70 ± 0.20[c]	Quiet	IUE
PG 1617+175	16 20 11.3	+17 24 27.7	0.1124	8.03 ± 0.10	Quiet	IUE
3C 390.3	18 42 08.8	+79 46 17.0	0.0560	7.71 ± 0.09	Loud	FOS
MRK 509	20 44 09.8	-10 43 24.5	0.0350	7.41 ± 0.03	Quiet	FOS
PG 2130+099	21 32 27.8	+10 08 19.5	0.0629	7.91 ± 0.05	Quiet	GHRS
NGC 7469	23 03 15.6	+08 52 26.4	0.0163	6.34 ± 0.04	Quiet	FOS

[a]Black hole masses are from Peterson et al. (2004), assuming $f = 1$.

[b]The radio type is either radio loud or radio quiet.

[c]The standard errors in M_{BH} for 3C 120 and PG 1613+658 were estimated by averaging the upper and lower uncertainties listed by Peterson et al. (2004).

sources are from the SDSS DR2 (Abazajian et al., 2004). Optical spectra for the 29 remaining $z < 0.8$ sources were obtained by us. Twenty-seven of the sources were observed at the Steward Observatory 2.3m Bok Telescope on Kitt Peak using the 600 lines mm^{-1} grating of the B&C Spectrograph; these sources had moderate spectral resolution (~ 5 $\mathring{A}$). The other two sources were observed at the Magellan Baade Telescope using the Inamori Magellan Areal Camera and Spectrograph (IMACS); long-slit spectra were obtained for these sources using a slit width of 0.9″ in long camera mode. We used the 600 lines mm^{-1} grating for PKS 1451-375 and the 300 lines mm^{-1} grating for PKS 2352-342, giving spectral resolutions of ~ 2 $\mathring{A}$ and ~ 5 $\mathring{A}$, respectively. The log of new spectra is displayed in Table 2.2, and they are shown in Figures 2.2–2.4. The spectra were reduced using the standard IRAF routines.

Table 2.2.   Log of New Observations

Object	RA (J2000)	DEC (J2000)	Redshift	Date	Instrument	Exp. Time (s)
Q 0003+0146	00 05 47.6	+02 03 02.2	0.234	Oct 06, 2002	Bok B&C	1800
Q 0017+0209	00 20 25.1	+02 26 25.3	0.401	Oct 06, 2002	Bok B&C	2700
PKS 0044+030	00 47 05.9	+03 19 54.9	0.624	Oct 06, 2002	Bok B&C	1800
Q 0100+0205	01 03 13.0	+02 21 10.4	0.394	Oct 06, 2002	Bok B&C	2700
Q 0115+027	01 18 18.5	+02 58 05.9	0.672	Oct 06, 2002	Bok B&C	3600
3C 57	02 01 57.2	-11 32 33.7	0.669	Oct 06, 2002	Bok B&C	1800
PKS 0214+10	02 17 07.7	+11 04 09.6	0.408	Oct 06, 2002	Bok B&C	2700
IR 0450-2958	04 52 30.0	-29 53 35.0	0.286	Oct 06, 2002	Bok B&C	207
3C 232	09 58 21.0	+32 24 02.2	0.533	May 15, 2002	Bok B&C	1800
4C 41.21	10 10 27.5	+41 32 39.1	0.611	May 15, 2002	Bok B&C	1800
B2 1028+313	10 30 59.1	+31 02 56.0	0.178	May 15, 2002	Bok B&C	1800
MC 1104+167	11 07 15.0	+16 28 02.4	0.632	May 15, 2002	Bok B&C	1800
Q 1132-0302	11 35 04.9	-03 18 52.5	0.237	May 14, 2002	Bok B&C	1800
Q 1150+497	11 53 24.5	+49 31 08.6	0.334	May 14, 2002	Bok B&C	1200
TEX 1156+213	11 59 26.2	+21 06 56.2	0.349	May 15, 2002	Bok B&C	2100
MRK 205	12 21 44.0	+75 18 38.1	0.070	May 15, 2002	Bok B&C	1800
Q 1230+0947	12 33 25.8	+09 31 23.0	0.420	Mar 20, 2004	Bok B&C	1800
Q 1317-0142	13 19 50.3	-01 58 04.6	0.225	May 14, 2002	Bok B&C	900
PG 1333+176	13 36 02.0	+17 25 13.0	0.554	May 15, 2002	Bok B&C	2100
Q 1340-0038	13 42 51.6	-00 53 46.0	0.326	May 14, 2002	Bok B&C	1200
Q 1401+0951	14 04 10.6	+09 37 45.5	0.441	May 15, 2002	Bok B&C	1800
Q 1425+2003	14 27 25.0	+19 49 52.3	0.111	May 14, 2002	Bok B&C	900
PKS 1451-375	14 54 27.4	-37 47 34.2	0.314	Jul 17, 2004	IMACS	2700
PG 1538+477	15 39 34.8	+47 35 31.6	0.770	May 14, 2002	Bok B&C	1500
PKS 2128-12	21 31 35.4	-12 07 05.5	0.501	Oct. 06, 2002	Bok B&C	1800
PKS 2135-147	21 37 45.2	-14 32 55.8	0.200	May 15, 2002	Bok B&C	900
Q 2141+175	21 43 35.6	+17 43 49.1	0.213	May 15, 2002	Bok B&C	1200
PKS 2243-123	22 46 18.2	-12 06 51.2	0.630	Oct 06, 2002	Bok B&C	1800
PKS 2352-342	23 55 25.6	-33 57 55.8	0.706	Jul 17, 2004	IMACS	1800
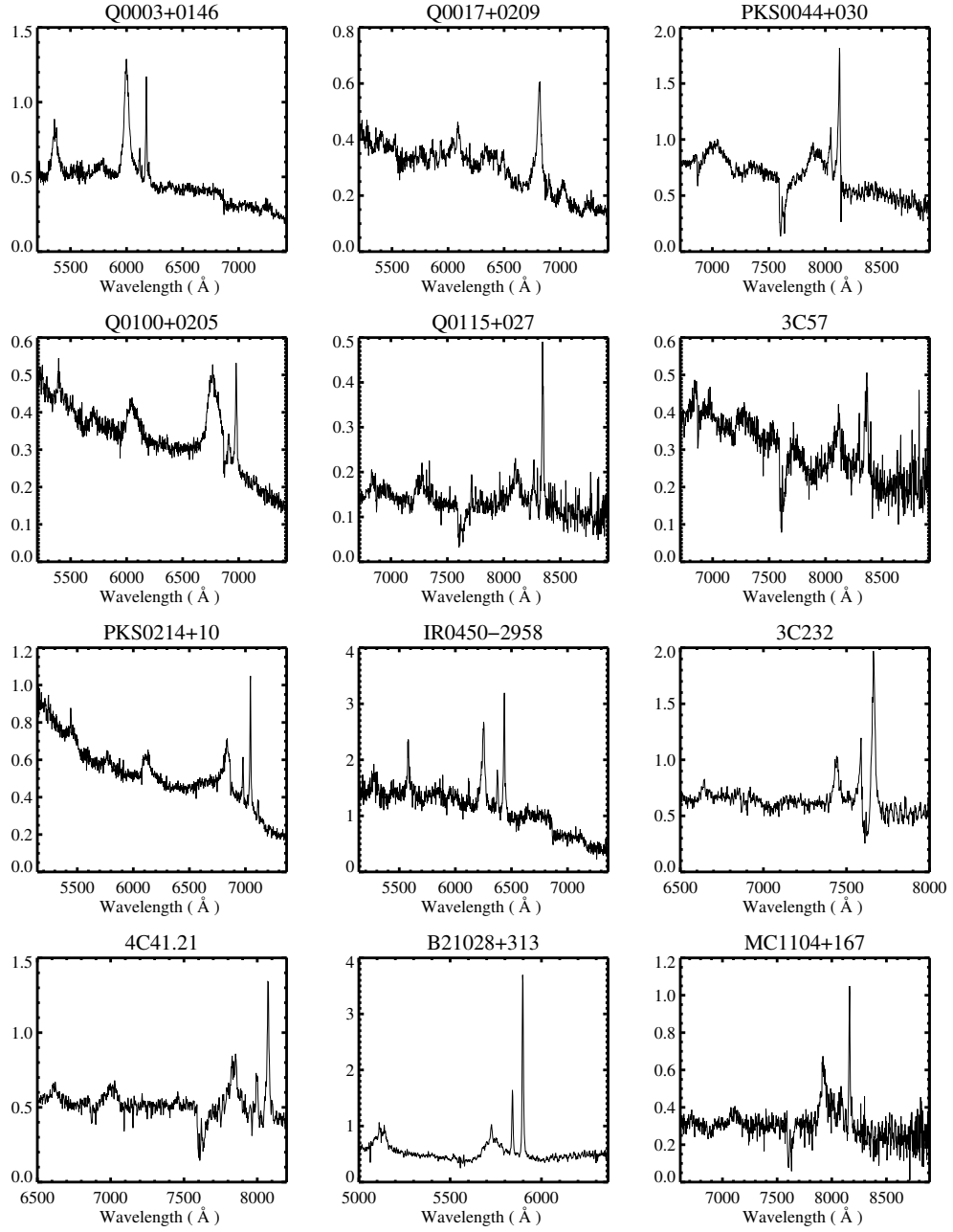
Figure 2.2 New optical spectra, shown in the observed frame. The fluxes are in units of 10^{-15} ergs cm^{-2} sec^{-1} $\mathring{A}^{-1}$. The absorption features at $\sim 6875\mathring{A}$ and $\sim 7600\mathring{A}$ are the A- and B-band atmospheric absorption lines. Note the strong iron emission in Q1340-0038.

Figure 2.3

PKS2128−12

PKS2135−147

Q2141+175

Wavelength (Å)

Wavelength (Å)

Wavelength (Å)

PKS2243−123

PKS2352−342

Wavelength (Å)

Wavelength (Å)
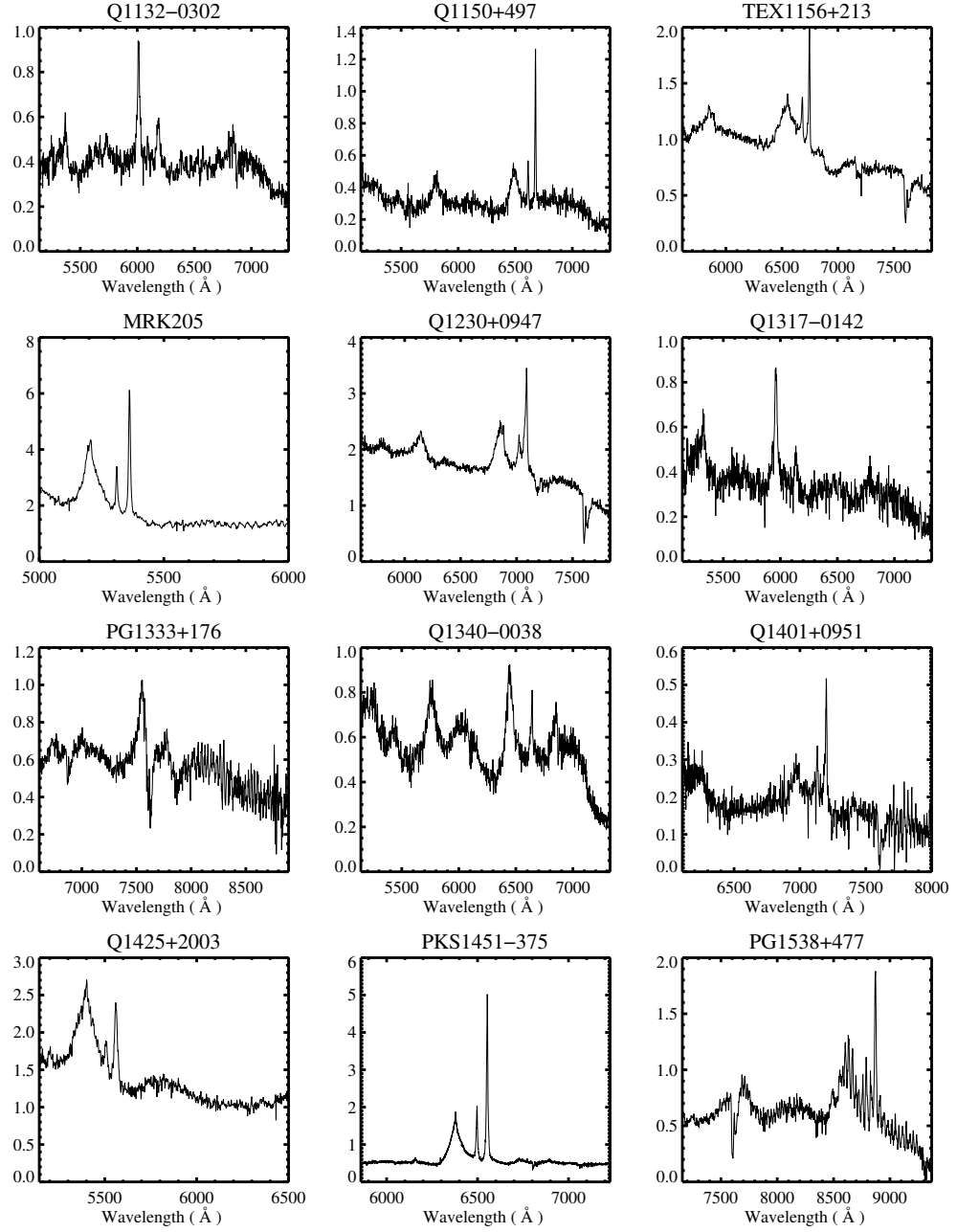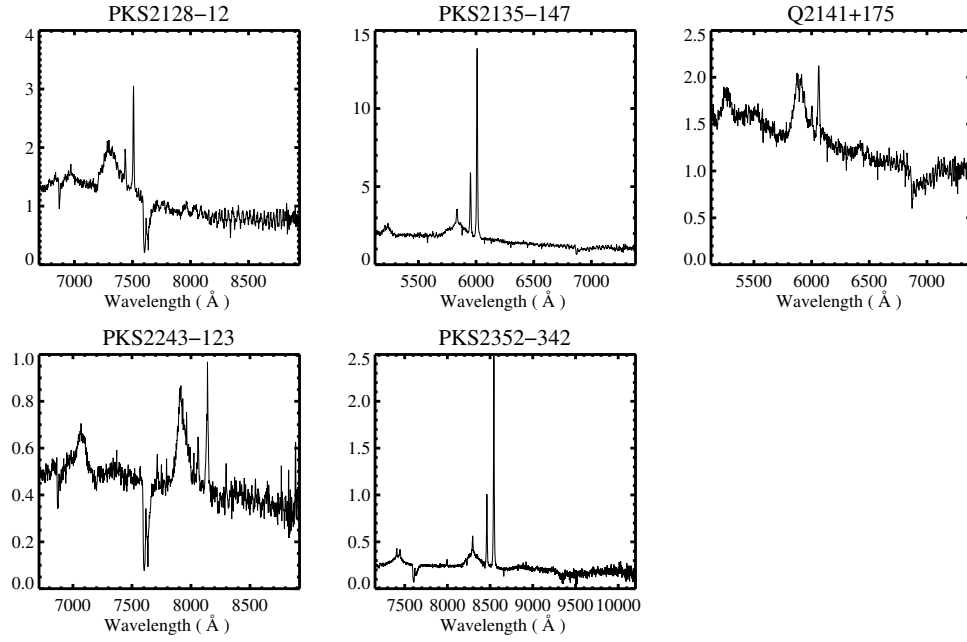
Figure 2.4

Rest-frame UV spectra for the eight sources with $z > 2$ are from Scott et al. (2000) and the SDSS. The continuum luminosities and Hβ $FWHM$ for the $z \sim 2.3$ sources were taken from McIntosh et al. (1999) and corrected to our adopted cosmology. The continuum fluxes for the two $z \sim 3.3$ sources were measured directly off the published spectra of Dietrich et al. (2002), and the values of Hβ $FWHM$ for them are from Dietrich & Hamann (2004).

2.5   LINE PROFILE PARAMETERS

2.5.1   Extracting The Line Profile

In order to extract the C IV emission line, it is necessary to subtract the continuum, Fe emission, and the He II $\lambda 1640$ and O III] $\lambda 1665$ emission lines. To remove the continuum and iron emission, we use a variation of the method outlined in Boroson & Green (1992). We model the continuum as a power law of the form $f_\nu \propto \nu^\alpha$. The Fe emission was modeled as a scaled and broadened form of the Fe emission template extracted from I Zw I by Vestergaard & Wilkes (2001). In constrast to most previous studies, we simultaneously fit the continuum and Fe emission parameters using the Levenberg-Marquardt method for nonlinear χ^2-minimization; this is similar to the method used by McIntosh et al. (1999). Performing the fits in this manner has the advantage of providing an estimate of the uncertainties in these parameters, given by the inverse of the curvature matrix of the χ^2 space. The set of possible windows used to fit the continuum and Fe emission are shown in Table 2.3. The actual continuum windows used to fit the continuum and Fe emission for any particular source depended on that source's available spectral range.

Having obtained an estimate of the continuum and Fe emission, we subracted these components from each spectrum. We then extracted the region within $-2 \times$

Table 2.3.   Continuum and Fe Emission Fitting Windows

	1	2	3	4	5	6
UV	$\lambda\lambda$1350–1365	$\lambda\lambda$1427–1500	$\lambda\lambda$1760–1860	$\lambda\lambda$1950–2300	$\lambda\lambda$2470–2755	$\lambda\lambda$2855–3010
Optical	$\lambda\lambda$3535–3700	$\lambda\lambda$4100–4200	$\lambda\lambda$4400–4700	$\lambda\lambda$5100–6200	$\lambda\lambda$6800–7534	$\cdots$

10^4 km s^{-1} and $3 \times 10^4 \text{ km s}^{-1}$ of 1549Å. Here, and throughout this work, we will use the convention that negative velocities are blueward of a given wavelength. Narrow absorption lines were removed and interpolated over. We then removed the He II λ1640 and O III] λ1665 emission lines from the wings of the C IV profile. This was done by modelling the C IV, He II, and O III] lines as a sum of Gaussians. In general, C IV was modelled as a sum of three Gaussians, He II two Gaussians, and O III] a single Gaussian, however this varied from source to source. The C IV extraction was done interactively in order to ensure accuracy of the fits. After obtaining estimates of the He II and O III] profiles, we subtracted these components. We did not fit the N IV] λ1486 emission line as this line is typically weak and lost in the C IV wings.

Extraction of the Hβ profile was done in a similar manner. For the optical Fe emission we used the I Zw I template of Véron-Cetty et al. (2004). After subtracting the continuum and Fe emission, we extracted the region within $\pm 2 \times 10^4 \text{ km s}^{-1}$ of 4861Å. The Hβ profile was modeled as a sum of 2–3 Gaussians. The [O III] $\lambda4959\text{Å}$ and [O III] $\lambda5007\text{Å}$ lines were modeled as a sum of 1–2 Gaussians, depending on the signal-to-noise of the lines. A sum of two Gaussians was used for the higher S/N lines because the [O III] lines are not exactly Gaussian, and not because the seperate Gaussians are considered to be physically distinct components. The widths of the narrow Gaussians for [O III] were fixed to

be equal. The [O III] lines and the narrow Hβ line were then subtracted to extract the Hβ line. As with the C IV profile, the entire extraction was done interactively.

2.5.2 Estimating The Line Profile Parameters

We measured the C IV line shift, $FWHM$, EW, and the first and second line moments. We define the line shift, Δv, as the location of the line peak relative to $1549 \mathring{A}$, in km s^{-1}. The location of $1549 \mathring{A}$ is determined from the redshifts listed on NED. We have checked the references for the redshifts to investigate how z was estimated, but not all of the references report this. For those that did specify, the redshifts were determined from the narrow emission lines (e.g., [O III] $\lambda 5007$.).

IUE observations were done using a very large aperture, so the values of Δv for the IUE sources may be biased. However, there is no noticeable difference between Δv estimated from the IUE spectra, and those estimated from the HST spectra. Furthermore, the Δv parameter does not enter into our final analysis, and so even if the IUE Δv parameters are significantly biased our conclusions remain unchanged.

The first moment of the C IV line is the centroid, μ_{CIV}, and the square root of the second central moment is the line dispersion, σ_*. The zeroth line moment is the line flux. The line moments are calculated as

$$F = \sum_{i=1}^{n} y_i \delta\lambda \tag{2.13}$$

$$\mu_{CIV} = \sum_{i=1}^{n} x_i y_i \bigg/ \sum_{i=1}^{n} y_i \tag{2.14}$$

$$\sigma_*^2 = \sum_{i=1}^{n} x_i^2 y_i \bigg/ \sum_{i=1}^{n} y_i - \mu_{CIV}^2 . \tag{2.15}$$

Here, F is the line flux, $\delta\lambda$ is the spacing between subsequent wavelengths, y is the observed spectral flux density, and x is the velocity relative to $1549 \mathring{A}$. In practice we do not perform the sums over all the n data points, but rather only over

those data points with $\hat{f}_i \geq 0.05(\max \hat{f})$, and assuming that the profile is mono-tonically decreasing blueward and redward of the peak. Here, $\hat{f}$ denotes the best-fit line profile, found from modelling the emission lines as a sum of Gaussians. This allows us to define the extent of the line profile. Although this gives us biased measurements of the line moments, it keeps these measurements stable. In particular, σ_* can be very sensitive to the profile wings, which have the highest uncertainty. By truncating the line profile we keep the estimate stable and less sensitive to errors in the Fe and continuum subtraction, as well as errors in the He II $\lambda 1640$ and O III] $\lambda 1665$ subtraction.

The line moments are calculated using the observed line profile, y, and not the best-fit to the spectral flux densities, $\hat{f}$. We do this because the line moments are relatively insensitive to a lack of smoothness. To be specific, consider a line profile, $f(x)$, and its Fourier transform, $\tilde{f}(k)$. For simplicity, we consider the continuous case here. The jth unnormalized line moment, μ_j, may be written in terms of the Fourier transform of the line profile:

$$\mu_j = \int_{-\infty}^{\infty} x^j f(x) dx = \frac{\tilde{f}^{(j)}(0)}{(-2\pi i)^j},$$

(2.16)

where $\tilde{f}^{(j)}(k)$ is the jth derivative of $\tilde{f}(k)$ and $i = \sqrt{-1}$. One can see from Equation (2.16) that the line moments only depend on $\tilde{f}(k)$ near $k = 0$, and are thus insensitive to the high frequency behavior of $f(x)$. A generic smoothing operator, such as a Gaussian fit, will shrink the high k components of $\tilde{f}(k)$ more than the low k, as it is generally the case that the high k components have lower signal-to-noise. However, because the line moments do not depend on the high k components, nothing is gained by enforcing smoothness. Because of this we just use the observed C IV profile, y, as it is an unbiased estimate of the true profile, f, whereas the best-fit estimate, $\hat{f}$, is a biased estimate.

There has been some discussion in the literature over whether $FWHM$ or σ_*

is the better width to use in estimating m (Fromerth & Melia, 2000; Peterson et al., 2004). Fromerth & Melia (2000) suggested using the line dispersion, σ_*, arguing that it provides a better estimate of the velocity dispersion for an arbitrary line profile. Other advantages of σ_* include its insensitivity to noise and narrow absorption lines. However, σ_* can be significantly biased due to its sensitivity to blending with other lines in the line wings, and to the removal of continuum and iron emission. These facts can be understood in light of Equation (2.16), which shows that the second moment of the line depends on the low-k behavior. Because of this, σ_* is relatively unaffected by information on small scales, such as noise and narrow absorption lines, but is significantly affected by information on large scales, such as line blending, truncation of the line profile, and continuum placement. The $FWHM$, on the other hand, is insensitive to line blends and errors in the continuum and iron subtraction. In addition, $FWHM$ is easily measured, however it probably provides a poor estimate of the velocity dispersion for irregularly shaped lines profiles. Peterson et al. (2004) compared the strengths and weaknesses of these two measurements and concluded that σ_* was the better parameter when measured in the RMS spectrum.

In order to choose the better SES estimate of the BLR velocity dispersion, we calculate the partial correlation between m and $FWHM$ and σ_*, respectively. The partial correlation coefficient describes the correlation between m and line width at a given luminosity; the line width with the higher partial correlation should give a better estimate of m. The partial linear correlation between $\log FWHM$ and m is 0.36, while the partial linear correlation between $\log \sigma_*$ and m is 0.31. Because the $FWHM$ has a moderately higher partial correlation, and because the $FWHM$ is not as affected by errors in the line deblending, continuum placement, etc., we use the $FWHM$ as an estimate of the velocity dispersion throughout the

rest of this analysis.

We only measure the $FWHM$ of the Hβ emission line. This is because we are only concerned with getting an estimate of M_{BH} from single-epoch Hβ based on the virial theorem for comparison with our C IV-based estimates, so the only line parameter of interest is the Hβ width.

The standard errors on the line parameter measurments are estimated using the bootstrap (Efron, 1979). In this method, we take our best-fit line profile spectral flux densities, $\hat{f}$, and generate $n_{\mathrm{boot}} = 128$ simulated observed line profiles by adding Gaussian noise to $\hat{f}$. We then estimate the line parameters of the simulated line profiles and calculate the variance in these parameters over the bootstrap samples.

In Figure 2.5 we plot $\log M_{BH}$ against $\log \lambda L_\lambda (1549\ \mathring{A})$, $\log FWHM$, $\log \sigma_*$, $\log EW$, Δv, μ_{CIV}, and continuum spectral slope. We report our measurements in Table 2.4. We have compared our measurements with Wang et al. (1996), Bechtold et al. (2002), Baskin & Laor (2004), and Kuraszkiewicz et al. (2004), and find them to be consistent after accounting for the different procedures used to measure these quantities. For clarity, we have removed three outliers in α from the plot of m against α. These sources were 3C 390.3, NGC 3516, and NGC 7469. 3C 390.3 is a broad-line radio galaxy with highly variable Balmer lines and double-peaked Hα (Corbett et al., 1998) and Hβ (Osterbrock et al., 1976) profiles; the C IV emission line also exhibits a double-peaked profile. The other two sources represent some of the faintest sources in our sample, and their spectra may have a contribution from their host galaxies; however, we notice nothing unusual about their spectra, save for their unusually soft values of α.
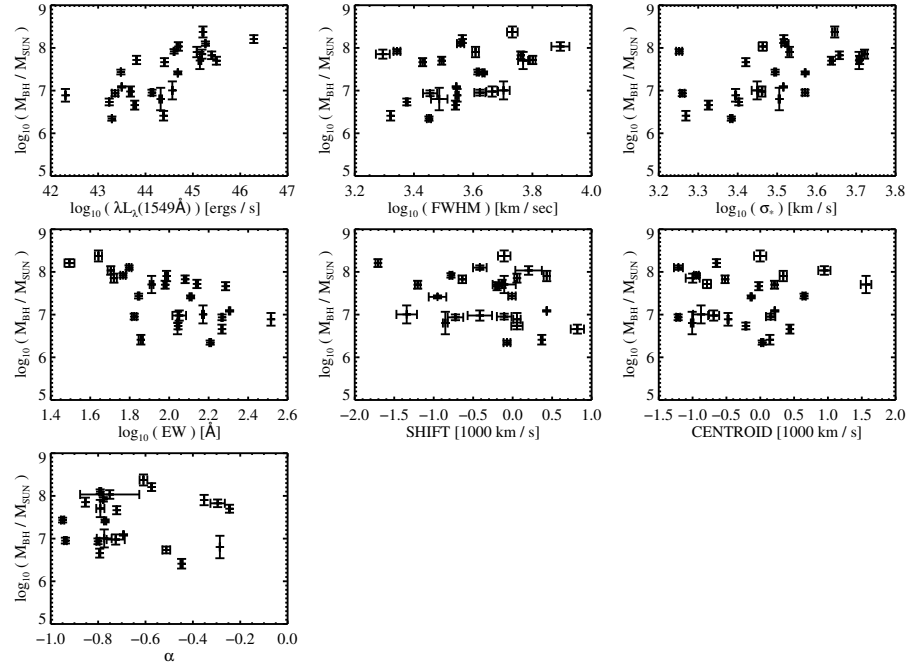
Figure 2.5 Plots showing m as a function of the measured quantities for the sample with reverberation mapping data. Three outliers in the m vs α plot have been removed so as to make the structure in the plot easier to see.

Table 2.4.   Continuum and C IV Emission Line Parameters

Object	$\log \lambda L_\lambda (1549\ Å)$	$FWHM$	σ_*	EW	Δv	μ	α
	ergs cm^{-2} sec^{-1}	1000 km s^{-1}	1000 km s^{-1}	Å	1000 km s^{-1}	1000 km s^{-1}	
MRK 335	44.38	2.094	1.856	71.82	0.368	0.141	-0.45
PG0026+12	45.19	1.971	5.263	52.45	0.054	-0.99	-0.85
0052+2509	45.39	5.762	4.544	120.4	-0.63	-0.51	-0.29
FAIRALL9	44.39	2.686	2.633	193.0	-0.19	-0.02	-0.72
MRK 590	43.35	2.846	1.818	185.2	-0.72	-1.21	-0.80
3C 120	44.56	5.040	2.813	148.3	-1.34	-0.87	-0.77
ARK 120	43.91	4.127	3.124	70.0	-0.00	0.647	-0.95
MRK 79	43.68	4.615	2.891	112.4	-0.41	-0.68	-0.73
PG 0804+761	45.26	3.608	3.307	62.63	-0.41	-1.20	-0.79
MRK 110	43.77	3.468	2.117	185.0	0.819	0.437	-0.79
PG0953+414	45.49	3.114	4.339	95.38	-1.20	0.210	-0.25
NGC 3516	42.31	3.512	2.480	327.9	0.054	-0.47	-2.05
NGC 3783	43.23	2.376	2.527	110.8	0.054	-0.21	-0.51
3C 273.0	46.28	3.659	3.285	31.24	-1.70	-0.64	-0.57
1307+0835	45.08	4.054	3.395	96.97	0.431	0.339	-0.35
MRK 279	44.32	3.057	3.201	110.2	-0.85	-1.00	-0.29
NGC 5548	43.49	3.490	3.278	202.2	0.431	0.214	-0.69
PG 1426+015	45.21	5.398	4.422	43.81	-0.10	-0.00	-0.61
MRK 817	44.13	4.192	3.719	66.52	-0.10	0.153	-0.94
PG 1613+658	45.14	5.869	5.099	81.43	-0.10	1.562	-0.79
PG 1617+175	44.69	7.842	2.906	50.60	0.201	0.947	-0.75
3C390.3	43.81	6.325	5.113	138.2	-0.13	-0.78	1.62
MRK509	44.68	4.313	3.722	128.3	-0.95	-0.13	-0.77
PG 2130+099	44.60	2.200	1.785	58.36	-0.78	-0.94	-0.78
NGC 7469	43.29	2.822	2.420	161.3	-0.07	0.028	-1.75

2.6   SELECTING THE IMPORTANT PARAMETERS

It is useful to investigate whether any of the additional parameters that we measured for C IV are significantly correlated with m, and thus contribute information about m in addition to that in l and $FWHM$. We used our sample of 25 sources with M_{BH} from reverberation mapping to test if including $\log EW, \alpha, \mu_{CIV}$, or Δv resulted in more accurate estimates of M_{BH}. However, we did not find any evidence to warrant the inclusion of these parameters.

In order to assess whether the data support including any additional parameters, we express m as a linear combination of all possible combinations of $\log \lambda L_\lambda, \log FWHM, \log EW, \alpha, \mu_{CIV}$, and Δv, a total of $2^6 = 64$ subsets. The regression coefficients are estimated via least-squares, and the result is a set of 64 linear regressions. In order to assess the relative merits of each of the regressions, and thus each of the respective parameters, we employ the Bayesian Information Criterion (BIC Schwartz, 1979). Using the BIC allows us to undertake a Bayesian comparison of the models without actually carrying out the full Bayesian prescription, considerably simplifying things. The BIC is easily obtained from the log-likelihood of the data as

$$BIC = 2\ell(\hat{\psi}) + d \ln n. \tag{2.17}$$

Here, $\ell(\hat{\psi})$ is the log-likelihood of the data evaluated at the maximum-likelihood estimate, ψ denotes the regression parameters, d is the number of parameters in the regression, and n is the number of data points.

The only parameter significantly correlated with m is l, with a posterior probability of $p_l = 0.975$. The data are ambiguous as to whether $FWHM$ is related to M_{BH} ($p_{FWHM} = 0.562$); however, when using the BIC we did not compare with regressions that assume $M_{BH} \propto FWHM_{CIV}^2$, and we perform a more careful

analysis in § 2.7.2. In addition, the data give weak evidence that the remaining parameters are unrelated to m. The posterior probabilities that m is correlated with these parameters are estimated to be 0.270, 0.210, 0.198, and 0.186 for $\log EW$, μ_{CIV}, Δv, and α, respectively.

2.7   REGRESSION ANALYSIS

Now that we have ruled out including any additional parameters for estimating M_{BH} from SES, we proceed to investigate the M_{BH}–L relationship and the C IV R–L relationship. Because $p(l|m)$ and $p(v|l,m)$ are statistically independent in their parameters, we can analyze each one seperately. Throughout this section we will be using our sample of 25 sources with black hole mass measurements from reverberation mapping. We begin by investigating the M_{BH}–L relationship.

2.7.1   The M_{BH}–L Relationship

We fit a linear relationship of the form $l = \alpha_0 + \alpha_m m$, assuming that the scatter about this relationship is independent of m and z. Because we are fitting the distribution of l at a given m, we use the BCES($Y|X$) (Akritas & Bershady, 1996) regression. The BCES methods take into account measurement errors in both coordinates by correcting their moments; however, the measurement errors are small compared to the variance in both m and l, so the correction is small. Based on the regression, we find

$$l = 35.72(\pm 1.67) + 1.17(\pm 0.22)m. \tag{2.18}$$

Here, $l = \log \lambda L_\lambda(1549\text{Å})$. The empirical value of $\hat{\alpha}_m = 1.17 \pm 0.22$ is consistent with $L \propto M_{BH}$, if one assumes that $L_\lambda \propto L$. The intrinsic scatter in Equation (2.18) is estimated to be $\hat{\sigma}_l = 0.61$ dex. The residuals and their cumulative distribution function (CDF) are shown in Figure 2.6. A Kolmogorov-Smirnov test found that

the regression residuals are not significantly different from a normal distribution, implying that $p(l|m)$ is normal with mean given by Equation (2.18) and standard deviation $\sigma_l \approx 0.61$ dex.

The value of $\hat{\alpha}_m$ found here is consistent with $M_{BH} \propto L^{0.9\pm0.15}$ found by Netzer (2003) and $M_{BH} \propto L^{0.97\pm0.16}$ found by Corbett et al. (2003). Peterson et al. (2004) used the 35 AGN with reverberation mapping data to estimate the M_{BH}– L relationship for L_λ at $5100\text{\AA}$. Using the BCES bisector regression, they find $M_{BH} \propto L_{5100}^{0.79\pm0.09}$. This is shallower than our result of $M_{BH} \propto L_{1549}^{1.17\pm0.22}$, although the two logarithmic slopes are consistent within the errors. The difference in the two values most likely results from the different regressions used. We used the BCES($Y|X$) because we are modelling $p(l|m)$, where as the bisector slope gives the regression that bisects the distribution of l at a given m, and of m at a given l.

Woo & Urry (2002) did not find any evidence for a correlation between M_{BH} and the bolometric luminosity. They used both a sample of reverberation-mapped AGNs and Seyfert galaxies with black hole masses derived from the stellar velocity dispersion. In neither case did they find evidence for a correlation, in constrast to the results found here and by others. Unfortunately, Woo & Urry (2002) did not perform a regression analysis or report a correlation coefficient, so it is difficult to do a quantitative comparison of their results with ours.

We have found here that $\hat{\alpha}_m = 1.17\pm0.22$ for the Peterson et al. (2004) sample. However, these sources are all at low redshift and have $\dot{m} = 0.01$–1 (Woo & Urry, 2002; Peterson et al., 2004), and thus this value of α_m may not be valid for sources with $\dot{m} \lesssim 0.01$ and $z \gtrsim 0.2$. Furthermore, most high z broad line quasars have luminosities and M_{BH} greater than that of the Peterson et al. (2004) sample, and it is possible that $\alpha_m \neq 1$ outside of the reverberation mapping sample range. It is also possible that the scatter about the M_{BH}–L relationship depends on z. The
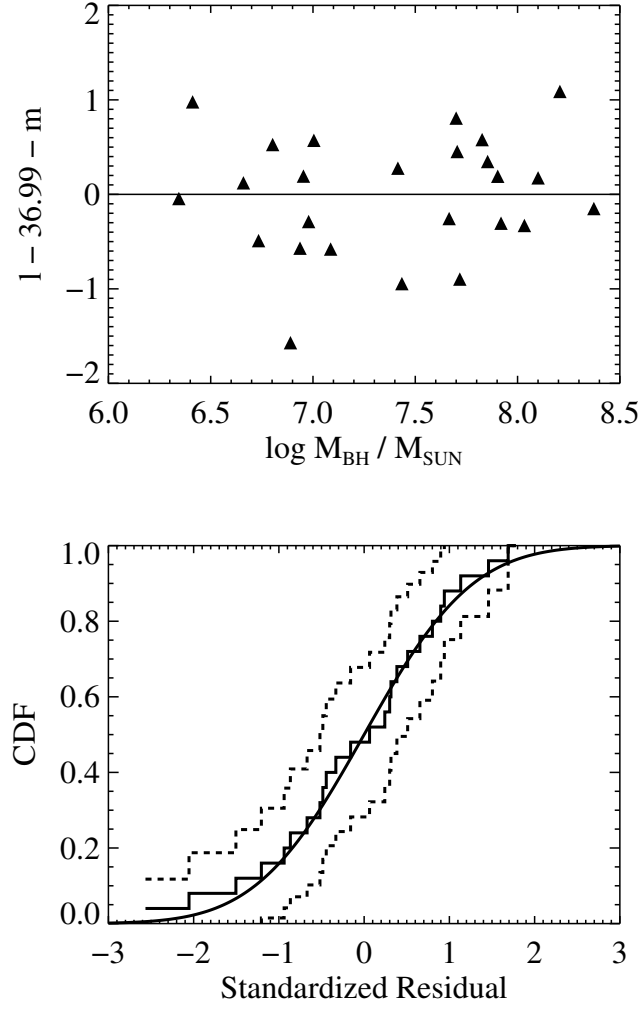
Figure 2.6 The residuals $l - \hat{l}$ as a function of m for the $M_{BH} - L$ relationship and the empirical CDF of the standardized residuals, $(l - \hat{l})/\hat{\sigma}_l$. The dashed lines define the 95% pointwise confidence interval of the empirical CDF, and the smooth solid line is the CDF of the standard normal density.

most likely source of a redshift dependence for ϵ_l is evolution of the distribution of $\dot{m}$ (e.g., Haiman & Menou, 2000; Merloni, 2004; Hopkins et al., 2006a; Steed & Weinberg, 2006), and possibly evolution of the bolometric correction. The average $\dot{m}$ has been observed to increase with increasing z (McLure & Dunlop, 2004), but this is likely due to selection effects. Because of these issues, one should exhibit caution when applying this M_{BH}–L relationship to high z sources.

2.7.2 Inferring the C IV R–L Relationship

In this subsection, we investigate the C IV R–L relationship using the reverberation mapping sample. We perform a linear regression to fit for the value of θ_l, $R \propto L^{\theta_l}$, by noting that $M_{BH} \propto L^{\theta_l}V^2$. Based on the results obtained here, we find $M_{BH} \propto L_{1549}^{0.41\pm0.07}FWHM_{CIV}^2$, consistent with $M_{BH} \propto L_{1549}^{1/2}FWHM_{CIV}^2$ expected from simple photoionization theory.

To estimate θ_l, we fit a linear relationship of the form $2v = m - m_0^{BL} - \theta_l l$ using the BCES($Y|X$) method. Here, the free parameters are m_0^{BL} and θ_l. The result is

$$\hat{m}_{CIV} = -17.82(\pm2.99) + 0.41(\pm0.07)l + 2\log FWHM_{CIV}. \qquad (2.19)$$

The intrinsic scatter about this relationship is $\hat{\sigma}_{CIV} = 0.40$ dex, where we have corrected for the measurement errors in m and $FWHM_{CIV}$. The correlation between the regression coefficients is $Corr(\hat{m}_0^{CIV}, \hat{\theta}_l) = -0.9996$. The regression results are consistent with the expectation from simple photoionization theory, $R \propto L_{1549}^{1/2}$ (cf. Eq.[2.7]). A Kolmogorov-Smirnov test found that the regression residuals are not significantly different from a normal distribution.

Performing the regression with v as the dependent variable ensures that the M_{BH}–L relationship does not 'absorb' into the regression coefficients, as the regression models the distribution of v at a given l and m. However, if we had performed the regression by fitting the distribution of m as a function of l and v,

the intrinsic correlation between m and l would have absorbed into the results. In this case we would be modelling $p(m|l, v)$, and the statistical model would be over-parameterized because we have only three variables. However, by seperating $p(m|l, v)$ into $p(l|m)$ and $p(v|l, m)$, as was done in § 2.3.2, we can analyze each distribution seperately and uniquely determine their parameters.

The residuals of the C IV broad line estimate, given by Equation (2.19), and their CDF are shown in Figure 2.7. In general, there is no obvious evidence for a violation of the regression assumptions. However, there appears to be a possible trend in the residuals with $FWHM$. To test this we calculated the linear correlation coefficient between the residuals and $\log FWHM$. This correlation was significant at only $\approx 1.7\sigma$. Therefore there is no significant evidence for a correlation in the residuals with $FWHM$, justifying the assumption $M_{BH} \propto L_{1549}^{\theta_l} FWHM^2$. Rank correlation tests gave similar results.

Very similar results were found by Vestergaard & Peterson (2006). However, our work differs from their's in that Vestergaard & Peterson (2006) assumed that the C IV and Hβ BLRs have the same dependence on L_{UV}, $R \propto L_{UV}^{0.53}$, and only fit the constant term. In contrast, we fit both the constant term and the coefficient for the dependence of line width on luminosity.

One can use Equation (2.9) to combine the mass estimates based on the M_{BH}-L with the C IV broad line mass estimate. Combining the two relationships, and taking $\sigma(z) \rightarrow \infty$, we find

$$\mu = -30.53 + 0.54l + 1.40 \log FWHM. \tag{2.20}$$

The intrinsic uncertainty in μ is reduced to $\sigma_\mu \approx 0.33$ dex, an improvement of $\approx 18\%$ over the broad line mass estimates. However, as mentioned in § 2.7.1, there is considerable systematic uncertainty on the behavior of the $M_{BH}-L$ relationship outside of the range probed by the reverberation mapping sample, and thus it
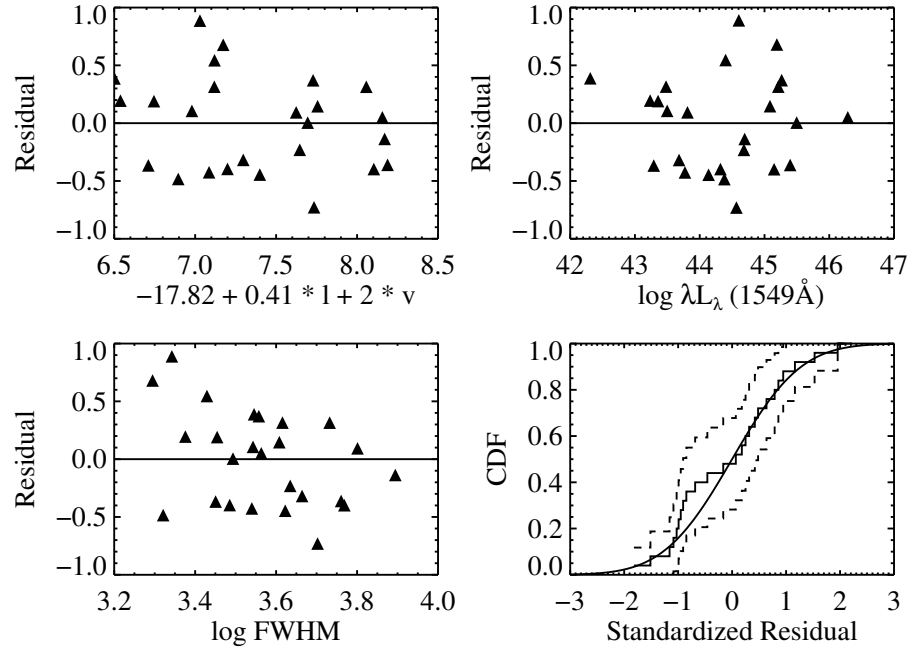
Figure 2.7 The residuals $m - \hat{m}_{CIV}$ for the C IV broad line mass estimate, shown as a function of $\hat{m}_{CIV}, l$, and $FWHM$. Also shown is the empirical CDF of the standardized residuals, $(m - \hat{m}_{CIV})/\hat{\sigma}_{CIV}$. The dashed lines define the 95% point-wise confidence interval of the empirical CDF, and the smooth solid line is the CDF of the standard normal density.

may be safer to just use $\hat{m}_{CIV}$.

There has been some concern over the effectiveness of using the C IV emission line for estimating quasar black hole masses (Shemmer et al., 2004; Baskin & Laor, 2005). If one ignores the C IV $FWHM$, and only uses the continuum luminosity to estimate M_{BH}, then the intrinsic scatter is $\sigma_l \approx 0.6$ dex (cf. § 2.7.1). However, the intrinsic uncertainty in an estimate of M_{BH} using the C IV R–L relationship and $FWHM$ is $\sigma_{CIV} \approx 0.4$ dex, a improvement over the M_{BH}–L relationship of $\sim 1/3$. This improvement is significant, and therefore we conclude that the C IV line may be used for estimating SMBH masses.

We can estimate σ_v^2 by comparing the C IV $FWHM$ for the RMS spectrum with that obtained from a single-epoch spectrum. To do this we use the sources from Peterson et al. (2004) with C IV $FWHM$ measured from the RMS spectrum, $FWHM_{RMS}$, excluding Fairall 9. We omitted Fairall 9 because its $FWHM_{RMS}$ was considered to be unreliable by Peterson et al. (2004). The remaining five data points are consistent with a 1:1 relationship between $FWHM_{RMS}$ and $FWHM_{SE}$. Estimating the intrinsic scatter using these five data points is difficult, particularly because the measurement errors are large. In fact, the observed scatter about $FWHM_{RMS} = FWHM_{SE}$ is consistent with entirely being the result of the measurement errors. This suggests that σ_v is small, and it is likely that $\sigma_v \lesssim 0.1$ dex.

If we assume that the only sources for scatter in the broad line estimates of m are from the R–L relationship and from using the single-epoch line width to estimate V, then a value of $\sigma_v \approx 0.1$ dex implies that the intrinsic scatter about the R–L relationship for C IV is $\hat{\sigma}_r \approx 0.35$ dex. This is about a factor of two larger than the value $\hat{\sigma}_r \approx 0.18$ dex found for Hβ by Peterson et al. (2004), where we have converted the value $\hat{\sigma}_r \approx 40\%$ to dex. This larger scatter in the C IV R–L

relationship implies that the Hβ mass estimates are more efficient than the C IV ones. However, Vestergaard & Peterson (2006) found that the Hβ mass estimates have a statistical scatter of $\hat{\sigma}_{\mathrm{H}\beta} \approx 0.43$ dex, and thus are just as accurate as the C IV-based ones. This larger value of $\sigma_{\mathrm{H}\beta}$ implies that either there is significant uncertainty in using the SES Hβ line width to estimate the BLR velocity dispersion, $\sigma_v^{\mathrm{H}\beta} \approx 0.36$, or that there are additional sources of uncertainty in the broad line mass estimates beyond the scatter in the R–L relationship and the uncertainty in using the SES line width instead of the variable component line width. If there are additional sources of scatter in the SES mass estimates, then $\hat{\sigma}_r \approx 0.35$ dex represents an upper bound on the C IV R–L relationship scatter.

Equation (2.7) implies that the scatter in the R–L relationship is the result of variations in the BLR ionization parameter, gas density, and average ionizing photon energy. While there are likely other sources of scatter in the R–L relationship, such as inclination, variability, and conversion between L_λ and L_{ion}, differences in BLR properties certainly contribute to this scatter and possibly dominate it. The fact that $\sigma_r^{CIV} \lesssim 1.6\sigma_r^{\mathrm{H}\beta}$ suggests that the magnitude of the dispersion in these properties is larger for the C IV emitting region. In addition, the regressions of (Peterson et al., 2004) are consistent with $R_{\mathrm{H}\beta} \propto L_{1450}^{1/2}$, similar to the result found here and by Vestergaard & Peterson (2006) for C IV.

2.8   COMPARING THE Hβ- AND C IV-BASED MASS ESTIMATES

It is worth comparing our estimates of m based on single-epoch C IV to those based on combining single-epoch Hβ with the $R_{\mathrm{H}\beta}$–L relationship. To do this, we took our sample of 100 sources that have data for both C IV and Hβ and calculated estimates of m from each. We compare the mass estimates obtained from both emission lines, and find that the two give mass estimates that are consistent

so long as one uses the UV continuum to estimate the BLR size.

For the Hβ emission line, we estimate the BLR size from both the optical and UV $R_{\mathrm{H}\beta}$–L relationships. The BLR size, $R_{\mathrm{H}\beta}$, is estimated as

$$\frac{R_{\mathrm{H}\beta}}{10 \ \mathrm{lt-dy}} = R_0 \left(\frac{\lambda L_\lambda}{10^{44} \mathrm{ergs \ s}^{-1}} \right)^\alpha. \tag{2.21}$$

For R_0 and α in the optical $R_{\mathrm{H}\beta}$–L relationship, we use the luminosity at $5100\mathring{A}$ and the average of the BCES bisector and FITEXY (Press et al., 1992) fits of Kaspi et al. (2005), $(\hat{R}_0, \hat{\alpha}) = (2.23\pm0.21, 0.69\pm0.05)$. For the UV $R_{\mathrm{H}\beta}$–L relationship, we use the luminosity at $1450\mathring{A}$ and $(R_0, \alpha) = (2.38, 0.5)$. We assume $R_{\mathrm{H}\beta} \propto L_{1450}^{1/2}$ to allow more direct comparison with the C IV-based mass estimates, and because the fits of Kaspi et al. (2005) are consistent with this form. Using the averaged values of R for the Balmer lines as listed by Kaspi et al. (2005), we recalculated the R–L_{1450} relationship for Hβ with α fixed at $\alpha = 0.5$, and found $R_0 = 2.38 \pm 0.25$. We then use $R_{\mathrm{H}\beta}$ with Hβ $FWHM$ to calculate $m_{\mathrm{H}\beta}$ from Equation (2.1), after converting the Hβ $FWHM$ to a velocity dispersion assuming $FWHM/V = 2$, based on the average $FWHM/\sigma_*$ from Peterson et al. (2004).

Based on the results of § 2.7.2, we calculate estimates of M_{BH} from the C IV line for these sources as $\hat{m}_{CIV} = -21.92 + 0.5l_{1549} + 2v_{CIV}$. This form was found using the same method as described in § 2.7.2, but with θ_l fixed at $\theta_l = 1/2$. In Table 2.5 we report the broad line mass estimates based on the Hβ and C IV lines for these sources, as well as a weighted average of the two.

Table 2.5.   Black Hole Mass Estimates

RA (J2000)	DEC (J2000)	z	$\log M_{BH}^{\mathrm{H}\beta}/M_\odot$ [a]	$\log M_{BH}^{CIV}/M_\odot$	$\log M_{BH}^{BL}/M_\odot$ [b]
00 05 47.6	+02 03 02.2	0.234	7.32 ± 0.43	7.61 ± 0.40	7.47 ± 0.29
00 05 59.2	+16 09 49.1	0.450	8.38 ± 0.43	8.18 ± 0.40	8.27 ± 0.29
00 06 19.5	+20 12 10.3	0.025	6.72 ± 0.43	6.90 ± 0.40	6.82 ± 0.29
00 20 25.1	+02 26 25.3	0.401	7.23 ± 0.43	8.33 ± 0.40	7.82 ± 0.29
00 29 13.7	+13 16 03.8	0.142	7.30 ± 0.43	7.17 ± 0.40	7.23 ± 0.29
00 47 05.9	+03 19 54.9	0.624	9.07 ± 0.43	8.67 ± 0.40	8.86 ± 0.29
00 52 02.4	+01 01 29.3	2.270	9.47 ± 0.44	9.41 ± 0.40	9.44 ± 0.29
00 52 33.7	+01 40 40.6	2.307	9.25 ± 0.46	9.55 ± 0.40	9.42 ± 0.30
00 54 52.1	+25 25 39.3	0.155	7.89 ± 0.43	8.34 ± 0.40	8.13 ± 0.29
01 03 13.0	+02 21 10.4	0.394	8.47 ± 0.43	8.16 ± 0.40	8.30 ± 0.29
01 18 18.5	+02 58 05.9	0.672	8.38 ± 0.44	7.61 ± 0.40	7.96 ± 0.29
01 23 45.7	-58 48 21.8	0.047	7.40 ± 0.43	7.16 ± 0.40	7.27 ± 0.29
01 26 42.8	+25 59 01.3	2.370	9.24 ± 0.44	9.25 ± 0.40	9.25 ± 0.30
01 57 35.0	+74 42 43.2	2.338	9.10 ± 0.44	9.20 ± 0.41	9.15 ± 0.30
02 01 57.2	-11 32 33.7	0.669	8.01 ± 0.47	8.85 ± 0.40	8.49 ± 0.30
02 17 07.7	+11 04 09.6	0.408	8.69 ± 0.43	8.20 ± 0.40	8.42 ± 0.29
02 59 05.6	+00 11 21.9	3.366	8.95 ± 0.43	9.03 ± 0.40	8.99 ± 0.29
03 04 49.9	-00 08 13.4	3.294	8.90 ± 0.43	8.69 ± 0.40	8.79 ± 0.29
03 51 28.6	-14 29 09.1	0.616	8.92 ± 0.43	9.02 ± 0.40	8.98 ± 0.29
04 05 34.0	-13 08 14.1	0.571	8.56 ± 0.43	8.05 ± 0.40	8.29 ± 0.29
04 07 48.4	-12 11 36.0	0.574	8.25 ± 0.43	8.41 ± 0.40	8.34 ± 0.29
04 17 16.8	-05 53 45.9	0.781	9.04 ± 0.45	8.96 ± 0.40	9.00 ± 0.30
04 41 17.3	-43 13 43.7	0.593	8.08 ± 0.43	7.84 ± 0.40	7.95 ± 0.29
04 52 30.0	-29 53 35.0	0.286	7.37 ± 0.43	7.86 ± 0.40	7.64 ± 0.29
04 56 08.9	-21 59 09.4	0.534	8.73 ± 0.43	8.49 ± 0.40	8.60 ± 0.29
07 45 41.7	+31 42 55.7	0.462	8.27 ± 0.43	8.54 ± 0.40	8.41 ± 0.29
08 40 47.6	+13 12 23.7	0.684	8.06 ± 0.43	8.09 ± 0.40	8.07 ± 0.29
08 53 34.2	+43 49 01.0	0.513	7.84 ± 0.43	8.41 ± 0.40	8.15 ± 0.29
09 19 57.7	+51 06 10.0	0.553	8.26 ± 0.43	8.58 ± 0.40	8.43 ± 0.29
09 50 48.4	+39 26 51.0	0.206	8.53 ± 0.43	7.88 ± 0.40	8.18 ± 0.29
09 56 52.4	+41 15 23.0	0.239	7.70 ± 0.43	7.85 ± 0.40	7.78 ± 0.29
09 58 21.0	+32 24 02.2	0.533	7.48 ± 0.43	8.87 ± 0.40	8.22 ± 0.30

Table 2.5—Continued

RA (J2000)	DEC (J2000)	z	$\log M_{BH}^{\mathrm{H}\beta}/M_\odot$ [a]	$\log M_{BH}^{CIV}/M_\odot$	$\log M_{BH}^{BL}/M_\odot$ [b]
10 04 02.6	+28 55 35.0	0.329	6.86 ± 0.43	8.25 ± 0.40	7.61 ± 0.29
10 04 20.1	+05 13 00.0	0.161	7.60 ± 0.43	7.40 ± 0.40	7.49 ± 0.29
10 10 27.5	+41 32 39.1	0.611	7.99 ± 0.43	8.38 ± 0.40	8.20 ± 0.29
10 30 59.1	+31 02 56.0	0.178	7.97 ± 0.44	7.59 ± 0.40	7.76 ± 0.30
10 51 51.5	+00 51 18.1	0.357	7.88 ± 0.43	8.18 ± 0.40	8.04 ± 0.29
11 04 13.9	+76 58 58.2	0.311	8.44 ± 0.43	7.92 ± 0.40	8.16 ± 0.29
11 06 31.8	+00 52 53.4	0.425	8.67 ± 0.43	8.43 ± 0.40	8.54 ± 0.29
11 06 33.5	-18 21 24.0	2.319	8.89 ± 0.45	9.65 ± 0.40	9.31 ± 0.30
11 07 15.0	+16 28 02.4	0.632	8.04 ± 0.45	8.65 ± 0.40	8.38 ± 0.30
11 18 30.3	+40 25 55.0	0.154	7.82 ± 0.43	7.93 ± 0.40	7.88 ± 0.29
11 19 08.7	+21 19 18.0	0.176	7.55 ± 0.43	8.13 ± 0.40	7.87 ± 0.29
11 24 39.2	+42 01 45.2	0.234	7.61 ± 0.43	7.47 ± 0.40	7.53 ± 0.29
11 35 04.9	-03 18 52.5	0.237	7.07 ± 0.43	7.87 ± 0.40	7.50 ± 0.29
11 39 57.1	+65 47 49.4	0.652	7.60 ± 0.43	8.22 ± 0.40	7.93 ± 0.29
11 41 21.7	+01 48 03.3	0.383	7.20 ± 0.45	8.23 ± 0.40	7.77 ± 0.30
11 47 18.0	-01 32 07.7	0.382	8.08 ± 0.43	6.68 ± 0.43	7.37 ± 0.30
11 53 24.5	+49 31 08.6	0.334	7.68 ± 0.43	8.02 ± 0.40	7.86 ± 0.29
11 58 39.9	+62 54 28.1	0.594	7.45 ± 0.51	8.32 ± 0.40	7.99 ± 0.32
11 59 26.2	+21 06 56.2	0.349	8.69 ± 0.43	8.03 ± 0.40	8.33 ± 0.29
12 04 42.2	+27 54 12.0	0.165	8.24 ± 0.43	7.26 ± 0.40	7.71 ± 0.29
12 14 17.7	+14 03 12.3	0.080	6.72 ± 0.43	6.95 ± 0.40	6.84 ± 0.29
12 19 20.9	+06 38 38.4	0.334	8.00 ± 0.43	7.95 ± 0.40	7.97 ± 0.29
12 21 44.0	+75 18 38.1	0.070	7.89 ± 0.43	7.59 ± 0.40	7.73 ± 0.29
12 31 20.6	+07 25 52.8	2.391	9.14 ± 0.49	9.19 ± 0.40	9.17 ± 0.31
12 33 25.8	+09 31 23.0	0.420	7.98 ± 0.43	8.18 ± 0.40	8.09 ± 0.29
13 01 12.9	+59 02 06.9	0.472	7.91 ± 0.43	9.01 ± 0.40	8.50 ± 0.29
13 05 33.0	-10 33 20.4	0.286	8.19 ± 0.43	8.38 ± 0.40	8.29 ± 0.29
13 09 47.0	+08 19 49.8	0.155	8.54 ± 0.43	7.87 ± 0.40	8.18 ± 0.29
13 12 17.7	+35 15 21.0	0.184	7.62 ± 0.43	7.43 ± 0.40	7.52 ± 0.29
13 19 50.3	-01 58 04.6	0.225	7.05 ± 0.43	7.23 ± 0.40	7.15 ± 0.29
13 23 49.5	+65 41 48.0	0.168	7.11 ± 0.43	7.60 ± 0.40	7.37 ± 0.29
13 36 02.0	+17 25 13.0	0.554	7.72 ± 0.43	8.79 ± 0.40	8.29 ± 0.29

Table 2.5—Continued

RA (J2000)	DEC (J2000)	z	$\log M_{BH}^{\mathrm{H}\beta}/M_\odot$[a]	$\log M_{BH}^{CIV}/M_\odot$	$\log M_{BH}^{BL}/M_\odot$[b]
13 42 51.6	-00 53 46.0	0.326	8.27 ± 0.43	8.20 ± 0.40	8.23 ± 0.29
13 57 04.5	+19 19 06.6	0.719	7.12 ± 0.44	7.99 ± 0.40	7.60 ± 0.30
14 04 10.6	+09 37 45.5	0.441	9.06 ± 0.45	8.49 ± 0.41	8.74 ± 0.30
14 05 16.2	+25 55 33.6	0.164	7.35 ± 0.43	8.19 ± 0.40	7.80 ± 0.29
14 17 00.9	+44 56 06.0	0.114	7.61 ± 0.43	7.68 ± 0.40	7.65 ± 0.29
14 19 03.9	-13 10 45.0	0.129	7.56 ± 0.43	7.76 ± 0.40	7.67 ± 0.29
14 27 25.0	+19 49 52.3	0.111	8.03 ± 0.43	8.30 ± 0.40	8.17 ± 0.29
14 27 35.7	+26 32 15.0	0.366	7.85 ± 0.43	8.80 ± 0.40	8.36 ± 0.29
14 29 43.1	+47 47 26.0	0.221	7.65 ± 0.43	7.64 ± 0.40	7.64 ± 0.29
14 42 07.5	+35 26 22.9	0.077	7.01 ± 0.43	7.26 ± 0.40	7.15 ± 0.29
14 46 45.9	+40 35 07.1	0.267	7.38 ± 0.43	8.42 ± 0.40	7.94 ± 0.29
14 54 27.4	-37 47 34.2	0.314	7.62 ± 0.43	7.93 ± 0.40	7.79 ± 0.29
15 14 43.5	+36 50 51.0	0.371	7.84 ± 0.44	8.28 ± 0.40	8.08 ± 0.30
15 39 34.8	+47 35 31.6	0.770	8.58 ± 0.46	9.01 ± 0.40	8.82 ± 0.30
15 45 30.3	+48 46 07.9	0.400	7.73 ± 0.43	8.55 ± 0.40	8.17 ± 0.29
15 47 43.5	+20 52 16.4	0.264	8.39 ± 0.43	8.21 ± 0.40	8.29 ± 0.29
16 14 13.2	+26 04 16.2	0.131	7.05 ± 0.43	7.78 ± 0.40	7.44 ± 0.29
16 20 21.8	+17 36 24.0	0.555	8.22 ± 0.43	8.69 ± 0.40	8.47 ± 0.29
16 27 56.1	+55 22 31.0	0.133	8.49 ± 0.43	7.89 ± 0.40	8.17 ± 0.29
16 42 58.8	+39 48 36.9	0.595	8.01 ± 0.43	8.45 ± 0.40	8.24 ± 0.29
17 04 41.3	+60 44 30.0	0.371	7.37 ± 0.44	8.15 ± 0.40	7.80 ± 0.30
18 21 59.4	+64 21 07.5	0.297	8.85 ± 0.43	8.53 ± 0.40	8.68 ± 0.29
19 27 48.5	+73 58 02.0	0.302	8.13 ± 0.43	8.19 ± 0.40	8.16 ± 0.29
20 44 09.8	-10 43 24.5	0.035	7.64 ± 0.43	7.69 ± 0.40	7.66 ± 0.29
21 31 35.4	-12 07 05.5	0.501	8.13 ± 0.43	8.64 ± 0.40	8.40 ± 0.29
21 37 45.2	-14 32 55.8	0.200	8.86 ± 0.43	7.99 ± 0.43	8.42 ± 0.31
21 43 35.6	+17 43 49.1	0.213	8.24 ± 0.43	8.20 ± 0.41	8.22 ± 0.30
22 03 15.0	+31 45 37.7	0.297	8.08 ± 0.43	8.76 ± 0.40	8.45 ± 0.29
22 46 18.2	-12 06 51.2	0.630	8.41 ± 0.43	8.13 ± 0.40	8.26 ± 0.29
22 54 05.8	-17 34 55.0	0.068	8.15 ± 0.43	7.56 ± 0.40	7.83 ± 0.29
22 54 10.4	+11 36 38.9	0.323	7.59 ± 0.43	8.13 ± 0.40	7.88 ± 0.29
23 03 43.5	-68 07 37.1	0.512	7.62 ± 0.43	7.77 ± 0.40	7.70 ± 0.29

Table 2.5—Continued

RA (J2000)	DEC (J2000)	z	$\log M_{BH}^{\mathrm{H}\beta}/M_\odot$ [a]	$\log M_{BH}^{CIV}/M_\odot$	$\log M_{BH}^{BL}/M_\odot$ [b]
23 11 17.8	+10 08 16.2	0.432	8.48 ± 0.43	8.36 ± 0.40	8.42 ± 0.29
23 46 36.9	+09 30 46.0	0.672	7.94 ± 0.43	7.99 ± 0.40	7.97 ± 0.29
23 51 56.0	-01 09 13.7	0.174	8.52 ± 0.43	8.09 ± 0.40	8.29 ± 0.29
23 55 25.6	-33 57 55.8	0.706	6.51 ± 0.44	8.39 ± 0.40	7.53 ± 0.30

Note. — The 1σ uncertainties include the contributions from measurement error in the line widths and continuum luminosities, and from the intrinsic uncertainty in $\hat{m}$.

[a]The Hβ-based mass estimates are for the UV $R_{\mathrm{H}\beta}$–L relationship. The 1σ errors were calculated assuming $\sigma_{\mathrm{H}\beta} = 0.43$ dex (Vestergaard & Peterson, 2006).

[b]M_{BH}^{BL} is a weighted average of the Hβ- and C IV-based mass estimates.

In Figure 2.8 we show $\hat{m}_{\mathrm{H}\beta}$ vs $\hat{m}_{\mathrm{CIV}}$ for both the optical and UV $R_{\mathrm{H}\beta}$–L relationship. We also show the 99% (2.6σ) confidence region for the parameters of a BCES bisector fit to $\hat{m}_{CIV} = A + B\hat{m}_{\mathrm{H}\beta}$. As can be seen, the values of $\hat{m}_{\mathrm{H}\beta}$ calculated using the UV $R_{\mathrm{H}\beta}$–L relationship are more consistent with the C IV-based ones. In addition, a pure 1:1 relationship falls outside of the 99% confidence region on the BCES bisector parameters for the optical-based $\hat{m}_{\mathrm{H}\beta}$ values, but it is contained within the 99% confidence region for the UV-based estimates. In fact, the BCES bisector fit for the UV-based mass estimates differs from a 1:1 relationship at a significance level of only 1.9σ. Our results are in agreement with Warner et al. (2003) and Dietrich & Hamann (2004), who also compared the C IV- and Hβ-based mass estimates and found them to be consistent.

That the C IV- and Hβ-based mass estimates are consistent is in contrast to the conclusions of Baskin & Laor (2005) and Shemmer et al. (2004). Baskin & Laor (2005) suggested that C IV-based mass estimates may be problematic because they find that $\log(FWHM_{\mathrm{CIV}}/FWHM_{\mathrm{H}\beta})$ is significantly anti-correlated
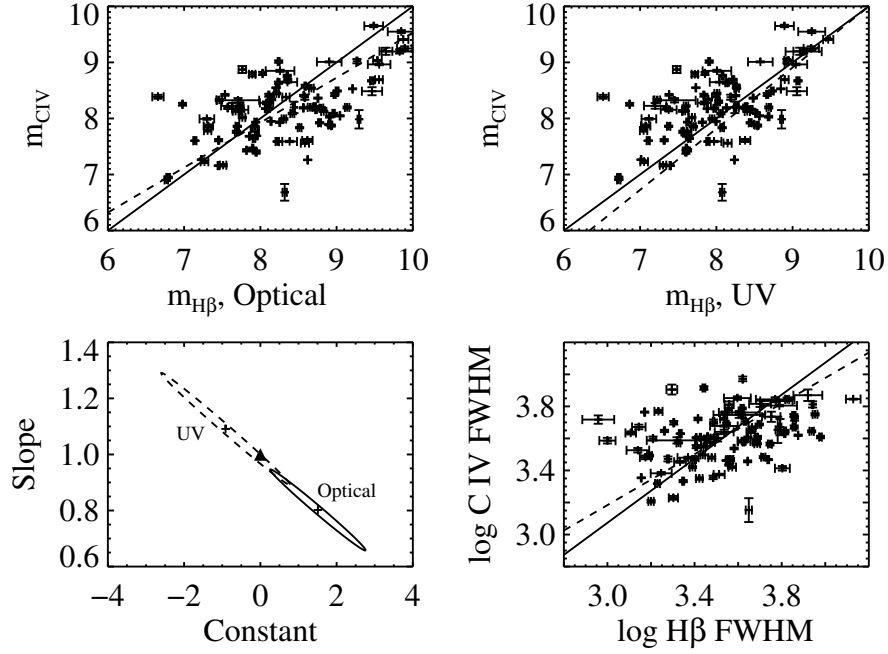
Figure 2.8 Plots comparing the Hβ-based mass estimates with the C IV-based ones. The values of $\hat{m}_{H\beta}$ for the top left plot are calculated from the optical R–L relationship, while the values of $\hat{m}_{H\beta}$ for the top right plot assume $R_{H\beta} \propto L_{1450}^{1/2}$. The error bars denote the propagated measurement errors in $\hat{m}$ resulting from the measurement errors in the line widths and continuum luminosities. The solid lines are a 1:1 relationship and the dashed lines are the BCES bisector fits. The bottom left plot shows the 99% (2.6σ) confidence regions for the BCES bisector fits comparing $\hat{m}_{CIV}$ with $\hat{m}_{H\beta}$. The solid contour correspond to $\hat{m}_{H\beta}$ calculated from the optical $R_{H\beta}$–L relationship, and the dashed contour when $\hat{m}_{H\beta}$ is calculated from the UV one. The crosses mark the best-fit values, and the triangle marks the values expected for a 1:1 relationship. The bottom right plot shows a comparison between the Hβ and C IV $FWHM$. The solid line is the best fit $FWHM_{H\beta} \propto FWHM_{CIV}$ relationship, and the dashed line is the BCES bisector fit.

with $\log FWHM_{\text{H}\beta}$. However this is expected even when $v_{\text{H}\beta} \propto v_{\text{CIV}}$, as an anti-correlation between the ratio of the line widths and the Hβ $FWHM$ is expected by construction; i.e., an anti-correlation between $\log FWHM_{\text{CIV}} - \log FWHM_{\text{H}\beta}$ and $\log FWHM_{\text{H}\beta}$ is expected so long as the deviations of $\log FWHM_{\text{CIV}}$ and $\log FWHM_{\text{H}\beta}$ from there respective means are not strongly correlated. In addition, none of the sources in the Baskin & Laor (2005) sample have Hβ broader than C IV for $FWHM_{\text{H}\beta} < 4000\,\text{km s}^{-1}$. To a lesser extant, a similar effect is seen in Figure 7 of Shemmer et al. (2004) and Figure 4 of Warner et al. (2003). In Figure 2.8 we compare the C IV and Hβ $FWHM$ for our sources. We find that most of the sources with $FWHM_{\text{H}\beta} \lesssim 2000\,\text{km s}^{-1}$ tend to have broader C IV lines. While the line widths certainly do not have a 1:1 relationship, the BCES bisector fit found that on average $FWHM_{\text{H}\beta}$ is approximately proportional to $FWHM_{CIV}$, $FWHM_{CIV} \propto FWHM_{\text{H}\beta}^{0.79\pm0.06}$. Baskin & Laor (2005) and Shemmer et al. (2004) did not perform a regression analysis, so we are unable to make a more quantitative comparison.

If the line widths are set by the virial relationship, we would expect $FWHM_{CIV} \propto FWHM_{\text{H}\beta}$. While the divergence from $FWHM_{CIV} \propto FWHM_{\text{H}\beta}$ is likely real (3.5σ significance), it is small, especially when compared with the intrinsic scatter in the $FWHM$ plot. Furthermore, this divergence from $FWHM_{CIV} \propto FWHM_{\text{H}\beta}$ does not appear to have a significant effect on the C IV mass estimates, as the masses inferred from the two emission lines differ only at the level of 1.9σ. As can by seen from the plot in Figure 2.8, any systematic difference between the C IV- and Hβ-based mass estimates is small compared to the statistical scatter in $\hat{m}$.

Vestergaard & Peterson (2006) performed a reanalysis of the Baskin & Laor (2005) sample and concluded that the poor correlation between the C IV $FWHM$ and the Hβ $FWHM$ seen by Baskin & Laor (2005) was due to the inclusion of

a large number of Narrow Line Seyfert 1s (NLS1) and the subtraction of a C IV narrow component. Vestergaard (2004) noted that for these sources the C IV line profile is unlikely to be suitable for estimating M_{BH}, as there may be a strong component from an outflowing wind. Upon removing the NLS1s and the sources with poor IUE data, Vestergaard & Peterson (2006) found the correlation between C IV $FWHM$ and Hβ $FWHM$ to be significantly better.

In addition, Baskin & Laor (2005) compared the correlation of C IV EW with $L/L_{edd} \propto L/M_{BH}$. A similar investigation was performed by Shemmer et al. (2004), where they compared the metallicity indicator N V/C IV with L/L_{edd}. Both authors found a stronger correlation when M_{BH} was estimated from the Hβ line as compared to C IV, and suggested that the C IV line may give a less efficient and possibly biased estimate of M_{BH}. However, as noted in § 2.3.3, correlation coefficients inferred from the estimated black hole masses must be interpreted with caution. In particular, the SES estimate of m, $\hat{m}$, merely defines the centroid of the probability density of $m|l, v, z$, and is in general *not* the actual m for an object. Therefore, this difference in the correlation coefficient between the Hβ- and C IV-based mass estimates may be just the result of random sampling, and it is unclear whether one can say that the correlation coefficients inferred from the two different mass estimates are inconsistent with those mass estimates being drawn from the same parent distribution.

The most direct test of the effectiveness of using C IV to estimate M_{BH} is found by comparing those values of $\hat{m}$ estimated using C IV with the actual reverberation mapping values of m. This was done in § 2.7.2, where inclusion of the C IV $FWHM$ resulted in a reduction of the error in $\hat{m}$ of $\sim 1/3$. In addition, Figure 2.7 compares the C IV-based estimate of m with the reverberation mapping values. As can be seen, assuming $M_{BH} \propto FWHM_{CIV}^2$ is consistent with the reverbera-

tion mapping sample, and therefore there is no reason to assume that the C IV line gives a biased estimate of M_{BH}. If $\hat{m}_{CIV}$ does give a biased estimate of m, then this bias is likely negligible compared to the variance in $\hat{m}_{CIV}$.

2.9 SUMMARY

In this work we have undertaken a statistical investigation of the best method to estimate SMBH mass based on the single-epoch C IV line and AGN continuum at $1549\mathring{A}$. The main conclusions are :

1. Estimating AGN black hole masses from a single-epoch spectrum is a considerably different problem than in the reverberation-mapping case. Because of this, one is not estimating M_{BH} directly, but calculating a probability distribution of M_{BH} given the observed luminosity, line width, and redshift, $p(m|l, v, z) \propto p(v|l, m, z)p(l|m, z)p(m|z)$.

2. Combining the information in L from both an intrinsic M_{BH}–L correlation and the R–L relationship results in improved black hole mass estimates. However, because of the current systematic uncertainties in the M_{BH}–L relationship, estimates based on its inclusion should be viewed with caution. In addition, incorporating information from the intrinsic distribution of M_{BH} also results in more accurate estimates, on average.

3. The distribution of M_{BH} inferred from the broad line mass estimates, $\hat{m}_{BL}$, is broader than the intrinsic distribution of M_{BH}, as $\hat{m}_{BL}$ are estimates of m contaminated by 'measurement error.' In addition, it is necessary to estimate or assume $p(l|m, z)$ in order to estimate a survey's completeness in M_{BH}.

4. The best estimates of M_{BH} based on the C IV emission line are obtained

using the UV continuum luminosity and $FWHM$. We find evidence that the C IV line shift, centroid, EW, and spectral slope of the UV continuum do not contribute any additional information to estimating M_{BH} for a given L_{1549} and $FWHM_{CIV}$.

5. Using the reverberation mapping sample, we find an M_{BH}–L relationship of the form

$$l = 35.72(\pm 1.67) + 1.17(\pm 0.22)m. \tag{2.22}$$

The intrinsic scatter about this relationship is $\sigma_l \approx 0.61$ dex. Combining mass estimates based on this relationship with the broad line mass estimates results in a reduction in the statistical error of $\approx 18\%$.

6. We estimate a C IV R–L relationship of the form $R \propto L_{1549}^{0.41 \pm 0.07}$. This is consistent with an R–L relationship of the form $R \propto L^{1/2}$ predicted from simple photoionization physics or if the BLR size is set by the dust sublimation radius. The scatter about the C IV R–L relationship inferred from the reverberation mapping estimates of M_{BH} is $\sigma_r \approx 0.35$ dex.

7. A broad line estimate of m based on the C IV emission line may be calculated as

$$\log \hat{M}_{BH}/M_\odot = \hat{m}_{CIV} = -17.82(\pm 2.99) + 0.41(\pm 0.07) \log \lambda L_\lambda(1549 \text{\AA}) + 2 \log FWHM_{CIV}. \tag{2.23}$$

Here, λL_λ is in units of ergs s^{-1} and $FWHM$ is in units of km s^{-1}. The correlation between regression coefficients is $Corr(\hat{m}_0^{CIV}, \hat{\theta}_l) = -0.9996$. The intrinsic scatter in m about Equation (2.23) is ≈ 0.40 dex.

8. The C IV- and Hβ-based mass estimates are consistent if one assumes $R \propto L_{UV}^{1/2}$ for both emission lines. The two emission lines give estimates of M_{BH}

with comparable accuracy.

9. We find evidence that the R–L relationships for C IV and Hβ are similar in their dependence on L_{UV}. For both, $R \propto L_{UV}^{1/2}$ is consistent with the available data. Also, the scatter in the C IV R–L relationship is a factor $\lesssim 1.6$ larger. This suggests that the C IV broad line region gas has a larger dispersion in its properties (e.g., density, ionization parameter, inclination) than the Hβ BLR gas.

CHAPTER 3

SOME ASPECTS OF MEASUREMENT ERROR IN LINEAR REGRESSION OF
ASTRONOMICAL DATA

3.1   CHAPTER ABSTRACT

I describe a Bayesian method to account for measurement errors in linear regression of astronomical data. The method allows for heteroscedastic and possibly correlated measurement errors, and intrinsic scatter in the regression relationship. The method is based on deriving a likelihood function for the measured data, and I focus on the case when the intrinsic distribution of the independent variables can be approximated using a mixture of Gaussians. I generalize the method to incorporate multiple independent variables, non-detections, and selection effects (e.g., Malmquist bias). A Gibbs sampler is described for simulating random draws from the probability distribution of the parameters, given the observed data. I use simulation to compare the method with other common estimators. The simulations illustrate that the Gaussian mixture model outperforms other common estimators and can effectively give constraints on the regression parameters, even when the measurement errors dominate the observed scatter, source detection fraction is low, or the intrinsic distribution of the independent variables is not a mixture of Gaussians. I conclude by using this method to fit the X-ray spectral slope as a function of Eddington ratio using a sample of 39 $z \lesssim 0.8$ radio-quiet quasars. I confirm the correlation seen by other authors between the radio-quiet quasar X-ray spectral slope and the Eddington ratio, where the X-ray spectral slope softens as the Eddington ratio increases. IDL routines are made available for performing the regression.

3.2   CHAPTER INTRODUCTION

Linear regression is one of the most common statistical techniques used in astronomical data analysis. In general, linear regression in astronomy is characterized by intrinsic scatter about the regression line, and measurement errors in both the independent and dependent variables. The source of intrinsic scatter is variations in the physical properties of astronomical sources that are not completely captured by the variables included in the regression. It is important to correctly account for both measurement error and intrinsic scatter, as both aspects can have a non-negligible effect on the regression results. In particular, ignoring the intrinsic scatter and weighting the data points solely by the measurement errors can result in the higher-precision measurements being given disproportionate influence on the regression results. Furthermore, when the independent variable is measured with error, the ordinary least squares (OLS) estimate of the regression slope is biased toward zero (e.g., Fuller, 1987; Akritas & Bershady, 1996; Fox, 1997). When there are multiple independent variables, measurement error can have an even stronger and more unpredictable effect (Fox, 1997). In addition, the existence of non-detections, referred to as 'censored data', in the data set will result in additional complications (e.g., Isobe et al., 1986). Therefore, when performing regression, it is essential to correctly account for the measurement errors and intrinsic scatter in order to ensure that the data analysis, and thus the scientific conclusions based on it, are trustworthy.

Many methods have been proposed for performing linear regression when intrinsic scatter is present and both variables are measured with error. These include methods that correct the observed moments of the data (e.g., Fuller, 1987; Akritas & Bershady, 1996; Freedman et al., 2004), minimize an 'effective' χ^2 statistic (e.g., Clutton-Brock, 1967; Barker & Diana, 1974; Press et al., 1992; Tremaine

et al., 2002), assume a probability distribution for the true independent variable values (so-called 'structural equation models', e.g., Schafer, 1987, 2001; Roy & Banerjee, 2006); Bayesian approaches to these models have also been developed (e.g., Zellner, 1971; Gull, 1989; Dellaportas & Stephens, 1995; Carroll et al., 1999; Scheines et al., 1999). In addition, methods have been proposed to account for measurement error in censored regression (e.g., Stapleton & Young, 1984; Weiss, 1993). The most commonly used methods in astronomy are the BCES estimator (Akritas & Bershady, 1996) and the 'FITEXY' estimator (Press et al., 1992). Both methods have their advantages and disadvantages, some of which have been pointed out by Tremaine et al. (2002). However, neither method is applicable when the data contain non-detections.

In this work I describe a Bayesian method for handling measurement errors in astronomical data analysis. My approach starts by computing the likelihood function of the complete data, i.e., the likelihood function of both the unobserved true values of the data and the measured values of the data. The measured data likelihood is then found by integrating the likelihood function for the complete data over the unobserved true values (e.g., Little & Rubin, 2002; Gelman et al., 2004). This approach is known as 'structural equation modelling' of measurement error problems, and has been studied from both a frequentist approach (e.g., Fuller, 1987; Carroll et al., 1995; Schafer, 2001; Aitken & Rocci, 2002) and a Bayesian approach (e.g., Müller & Roeder, 1997; Richardson & Leblond, 1997). In this work, I extend the statistical model of Carroll et al. (1999) to allow for measurement errors of different magnitudes (i.e., 'heteroscedastic' errors), non-detections, and selection effects, so long as the selection function can be modelled mathematically. Our method models the distribution of independent variables as a weighted sum of Gaussians. The mixture of Gaussians model allows flexibility

when estimating the distribution of the true values of the independent variable, thus increasing its robustness against model mispecification (e.g., Huang et al., 2006). The basic idea is that one can use a suitably large enough number of Gaussians to accurately approximate the true distribution of independent variables, even though in general the individual Gaussians have no physical meaning.

The paper is organized as follows. In § 3.3 we summarize some notation, and in § 3.4 I review the effects of measurement error on the estimates for the regression slope and correlation coefficient. In § 3.5 I describe the statistical model and derive the likelihood functions, and in § 3.6 I describe how to incorporate knowledge of the selection effects and account for non-detections. In § 3.7.1 I describe the prior distribution for this model, and in § 3.7.2 I describe a Gibbs sampler for sampling from the posterior distributions. In § 3.8 I use simulation to illustrate the effectiveness of this structural model and compare with the OLS, BCES($Y|X$), and FITEXY estimators. Finally, in § 3.9 I illustrate the method using astronomical data by performing a regression of the X-ray photon index, Γ_X, on the Eddington ratio using a sample of 39 $z < 0.83$ radio-quiet quasars. Sections 3.5, 3.6, and 3.7 are somewhat technical, and the reader who is uninterested in the mathematical and computational details may skip to them.

3.3 NOTATION

I will use the common statistical notation that an estimate of a quantity is denoted by placing a 'hat' above it; e.g., $\hat{\theta}$ is an estimate of the true value of the parameter θ. In general, greek letters will denote the true value of a quantity, while roman letters will denote the contaminated measured value. I will frequently refer to the 'bias' of an estimator. The bias of an estimator is $E(\hat{\theta}) - \theta_0$, where $E(\hat{\theta})$ is the expectation value of the estimator $\hat{\theta}$, and θ_0 is the true value of θ. An unbiased

estimator is one such that $E(\hat{\theta}) = \theta_0$.

I will denote a normal density with mean μ and variance σ^2 as $N(\mu, \sigma^2)$, and I will denote as $N_p(\mu, \Sigma)$ a multivariate normal density with p-element mean vector μ and $p \times p$ covariance matrix Σ. If I want to explicitly identify the argument of the Gaussian function, I will use the notation $N(x|\mu, \sigma^2)$, which should be understood to be a Gaussian with mean μ and variance σ^2 as a function of x. Following Gelman et al. (2004), I denote the scaled inverse-χ^2 density as Inv-$\chi^2(\nu, s^2)$, where ν is the degrees of freedom and s^2 is the scale parameter, and we denote the inverse-Wishart as Inv-Wishart$_\nu(S)$, where ν is the degrees of freedom and S is the scale matrix. The inverse-Wishart distribution can be thought of as a multivariate generalization of the scaled inverse-χ^2 distribution. I will often use the common statistical notation where "$\sim$" means "is drawn from" or "is distributed as". For example, $x \sim N(\mu, \sigma^2)$ states that x is drawn from a normal density with mean μ and variance σ^2.

3.4   EFFECT OF MEASUREMENT ERROR ON CORRELATION AND REGRESSION

It is well known that measurement error can attenuate the estimate of the regression slope and correlation coefficient (e.g., Fuller, 1987; Fox, 1997). For completeness, I give a brief review of the effect of measurement error on correlation and regression analysis for the case of one independent variable.

Denote the independent variable as ξ and the dependent variable as η; ξ and η are also referred to as the 'covariate' and the 'response', respectively. I assume that ξ is a random vector of n data points drawn from some probability distribution. The dependent variable, η, depends on ξ according to the usual additive

model:

$$\eta_i = \alpha + \beta\xi_i + \epsilon_i \tag{3.1}$$

Here, ϵ_i is a random variable representing the intrinsic scatter in η_i about the regression relationship, and (α, β) are the regression coefficients. The mean of ϵ is assumed to be zero, and the variance of ϵ is assumed to be constant and is denoted as σ^2. We do not observe the actual values of (ξ, η), but instead observe values (x, y) which are measured with error. The measured values are assumed to be related to the actual values as

$$x_i = \xi_i + \epsilon_{x,i} \tag{3.2}$$

$$y_i = \eta_i + \epsilon_{y,i}, \tag{3.3}$$

where $\epsilon_{x,i}$ and $\epsilon_{y,i}$ are the random measurement errors on x_i and y_i, respectively. In general, the errors are normally distributed with known variances $\sigma_{x,i}^2$ and $\sigma_{y,i}^2$, and covariance $\sigma_{xy,i}$. For simplicity, throughout the rest of this section I assume that σ_x^2, σ_y^2, and σ_{xy} are the same for each data point.

When the data are measured without error, the least-squares estimate of the regression slope, $\hat{\beta}_{OLS}$, and the estimated correlation coefficient, $\hat{\rho}$, are

$$\hat{\beta}_{OLS} = \frac{Cov(\xi, \eta)}{Var(\xi)} \tag{3.4}$$

$$\hat{\rho} = \frac{Cov(\xi, \eta)}{\sqrt{Var(\xi)Var(\eta)}} = \hat{\beta}_{OLS}\sqrt{\frac{Var(\xi)}{Var(\eta)}}. \tag{3.5}$$

Here, $Cov(\xi, \eta)$ is the sample covariance between ξ and η, and $Var(\xi)$ is the sample variance of ξ. When the data are measured with error, the least-squares estimate of the regression slope, $\hat{b}_{OLS}$, and the estimated correlation coefficient, $\hat{r}$, become

$$\hat{b}_{OLS} = \frac{Cov(x, y)}{Var(x)} = \frac{Cov(\xi, \eta) + \sigma_{xy}}{Var(\xi) + \sigma_x^2} \tag{3.6}$$

$$\hat{r} \;=\; \frac{Cov(x,y)}{\sqrt{Var(x)Var(y)}} = \frac{Cov(\xi,\eta)+\sigma_{xy}}{\sqrt{(Var(\xi)+\sigma_x^2)(Var(\eta)+\sigma_y^2)}}. \tag{3.7}$$

From these equations it is apparent that the estimated slope and correlation are biased when the data are measured with error.

It is informative to assess the effect of measurement error in terms of the ratios $R_x = \sigma_x^2/Var(x)$, $R_y = \sigma_y^2/Var(y)$, $R_{xy} = \sigma_{xy}/Cov(x,y)$, as these quantities can be calculated from the data. The fractional bias in the estimated slope and correlation may then be expressed as

$$\frac{\hat{b}}{\hat{\beta}} \;=\; \frac{1-R_x}{1-R_{xy}} \tag{3.8}$$

$$\frac{\hat{r}}{\hat{\rho}} \;=\; \frac{\sqrt{(1-R_x)(1-R_y)}}{1-R_{xy}}. \tag{3.9}$$

From Equations (3.8) and (3.9) it is apparent that measurement errors have the following effects. First, covariate measurement error reduces the magnitude of the observed correlation between the independent variable and the response, as well as biasing the estimate of the slope towards zero. Second, measurement error in the response also reduces the magnitude of the observed correlation between the variables. Third, if the measurement errors are correlated the effects depend on the sign of this correlation. If the measurement error correlation has the same sign as the intrinsic correlation between ξ and η, then the measurement errors cause a spurious increase in the observed correlation; otherwise the measurement errors cause a spurious decrease in the observed correlation. The magnitude of these effects depend on how large the measurement errors are compared to the observed variance in x and y.

In Figure 3.1 I plot the fractional bias in the correlation coefficient, $(\hat{\rho} - \hat{r})/\hat{\rho}$, as a function of R_x and R_y when the errors are uncorrelated. As can be seen, measurement error can have a significant effect on the estimation of the linear
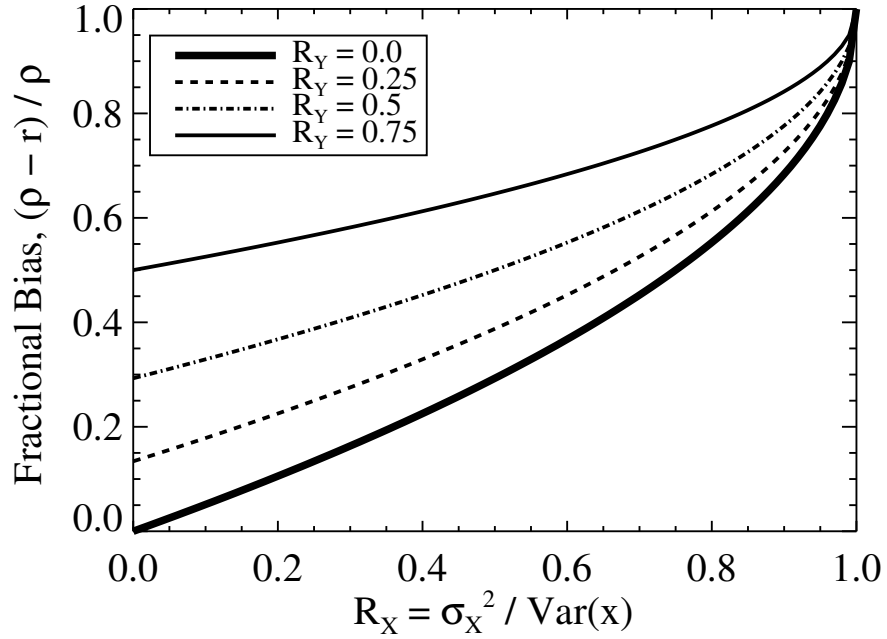
Figure 3.1 The fractional bias in the correlation coefficient when the data are contaminated with measurement error. The fractional bias is shown as a function of the contribution of measurement error to the observed variance in both x and y, for uncorrelated measurement errors. When the measurement errors make up $\sim 50\%$ of the observed variance in both x and y, the observed correlation coefficient is reduced by about $\sim 50\%$.

correlation coefficient. For example, when $R_x \approx 0.5$ and $R_y \approx 0.5$, the estimated correlation is $\approx 50\%$ lower than the true correlation. Therefore, interpretation of correlation coefficients and regression slopes must be approached with caution when the data have been contaminated by measurement error. To ensure accurate results, it is necessary to employ statistical methods that correct for the measurement errors.

3.5   THE STATISTICAL MODEL

3.5.1   Regression with One Independent Variable

I assume that the independent variable, ξ, is drawn from a probability distribution $p(\xi|\psi)$, where ψ denotes the parameters for this distribution. The dependent variable is then drawn from the conditional distribution of η given ξ, denoted as $p(\eta|\xi, \theta)$; θ denotes the parameters for this distribution. The joint distribution of ξ and η is then $p(\xi, \eta|\psi, \theta) = p(\eta|\xi, \theta)p(\xi|\psi)$. In this work I assume the normal linear regression model given by Equation (3.1), and thus $p(\eta|\xi, \theta)$ is a normal density with mean $\alpha + \beta\xi$ and variance σ^2, and $\theta = (\alpha, \beta, \sigma^2)$.

Since the data are a randomly observed sample, we can derive the likelihood function for the measured data. The likelihood function of the measured data, $p(x, y|\theta, \psi)$, is obtained by integrating the complete data likelihood over the missing data, ξ and η (e.g., Little & Rubin, 2002; Gelman et al., 2004):

$$p(x, y|\theta, \psi) = \int\int p(x, y, \xi, \eta|\theta, \psi) \, d\xi \, d\eta. \tag{3.10}$$

Here, $p(x, y, \xi, \eta|\theta, \psi)$ is the complete data likelihood function. Because of the hierarchical structure inherent in the measurement error model, it is helpful to decompose the complete data likelihood into conditional probability densities:

$$p(x, y|\theta, \psi) = \int\int p(x, y|\xi, \eta)p(\eta|\xi, \theta)p(\xi|\psi) \, d\xi \, d\eta. \tag{3.11}$$

The density $p(x, y|\xi, \eta)$ describes the joint distribution of the measured values x and y at a given ξ and η, and depends on the assumed distribution of the measurement errors, ϵ_x and ϵ_y. In this work I assume Gaussian measurement error, and thus $p(x_i, y_i|\xi_i, \eta_i)$ is a multivariate normal density with mean (ξ_i, η_i) and covariance matrix Σ_i, where $\Sigma_{11,i} = \sigma_{y,i}^2$, $\Sigma_{22,i} = \sigma_{x,i}^2$, and $\Sigma_{12,i} = \sigma_{xy,i}$. The statistical model may then be conveniently expressed hierarchically as

$$\xi_i \quad \sim \quad p(\xi|\psi) \tag{3.12}$$

$$\eta_i | \xi_i \quad \sim \quad N(\alpha + \beta \xi_i, \sigma^2) \tag{3.13}$$

$$y_i, x_i | \eta_i, \xi_i \quad \sim \quad N_2([\eta_i, \xi_i], \Sigma_i) \tag{3.14}$$

Note that if x_i is measured without error, then $p(x_i | \xi_i)$ is a Dirac delta function, and $p(x_i, y_i | \xi_i, \eta_i) = p(y_i | \eta_i) \delta(x_i - \xi_i)$. An equivalent result holds if y_i is measured without error.

Equation (3.11) may be used to obtain the observed data likelihood function for any assumed distribution of ξ. In this work, I model $p(\xi | \psi)$ as a mixture of K Gaussians,

$$p(\xi_i | \psi) = \sum_{k=1}^{K} \frac{\pi_k}{\sqrt{2\pi\tau_k^2}} \exp\left\{ -\frac{1}{2} \frac{(\xi_i - \mu_k)^2}{\tau_k^2} \right\}, \tag{3.15}$$

where $\sum_{k=1}^{K} \pi_k = 1$. Note that, π_k may be interpreted as the probability of drawing a data point from the kth Gaussian. I will use the convenient notation $\pi = (\pi_1, \ldots, \pi_K), \mu = (\mu_1, \ldots, \mu_K)$, and $\tau^2 = (\tau_1^2, \ldots, \tau_K^2)$; note that $\psi = (\pi, \mu, \tau^2)$. It is useful to model $p(\xi | \psi)$ using this form because it is flexible enough to adapt to a wide variety of distributions, but is also conjugate for the regression relationship (Eq.[3.1]) and the measurement error distribution, thus simplifying the mathematics.

Assuming the Gaussian mixture model for $p(\xi | \psi)$, the measured data likelihood for the ith data point can be directly calculated using Equation (3.11). Denoting the measured data as $\mathbf{z} = (y, x)$, the measured data likelihood function for the ith data point is then a mixture of bivariate normal distributions with weights π, means $\zeta = (\zeta_1, \ldots, \zeta_K)$, and covariance matrices $V_i = (V_{1,i}, \ldots, V_{K,i})$. Because the data points are statistically independent, the full measured data likelihood is then the product of the likelihood functions for the individual data points:

$$p(x, y | \theta, \psi) = \prod_{i=1}^{n} \sum_{k=1}^{K} \frac{\pi_k}{2\pi |V_{k,i}|^{1/2}} \exp\left\{ -\frac{1}{2} (\mathbf{z}_i - \zeta_k)^T V_{k,i}^{-1} (\mathbf{z}_i - \zeta_k) \right\} \tag{3.16}$$

$$\zeta_k = (\alpha + \beta\mu_k, \mu_k) \tag{3.17}$$

$$V_{k,i} = \begin{pmatrix} \beta^2 \tau_k^2 + \sigma^2 + \sigma_{y,i}^2 & \beta \tau_k^2 + \sigma_{xy,i} \\ \beta \tau_k^2 + \sigma_{xy,i} & \tau_k^2 + \sigma_{x,i}^2 \end{pmatrix}. \tag{3.18}$$

Here, $\mathbf{z}^T$ denotes the transpose of $\mathbf{z}$. Equation (3.16) may be maximized to compute the maximum-likelihood estimate (MLE). When $K > 1$, the expectation-maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977) is probably the most efficient tool for calculating the MLE. Roy & Banerjee (2006) describe an EM algorithm when $p(\xi)$ is assumed to be a mixture of normals and the measurement error distribution is multivariate t, and their results can be extended to the statistical model described in this work.

It is informative to decompose the measured data likelihood, $p(x_i, y_i | \theta, \psi) = p(y_i | x_i, \theta, \psi) p(x_i | \psi)$, as this representation is useful when the data contain non-detections (cf., § 3.6.2). The marginal distribution of x_i is

$$p(x_i | \psi) = \sum_{k=1}^{K} \frac{\pi_k}{\sqrt{2\pi(\tau_k^2 + \sigma_{x,i}^2)}} \exp \left\{ -\frac{1}{2} \frac{(x_i - \mu_k)^2}{\tau_k^2 + \sigma_{x,i}^2} \right\}, \tag{3.19}$$

and the conditional distribution of y_i given x_i is

$$p(y_i | x_i, \theta, \psi) = \sum_{k=1}^{K} \frac{\gamma_k}{\sqrt{2\pi Var(y_i | x_i, k)}} \exp \left\{ -\frac{1}{2} \frac{[y_i - E(y_i | x_i, k)]^2}{Var(y_i | x_i, k)} \right\} \tag{3.20}$$

$$\gamma_k = \frac{\pi_k N(x_i | \mu_k, \tau_k^2 + \sigma_{x,i}^2)}{\sum_{j=1}^{K} \pi_j N(x_i | \mu_j, \tau_j^2 + \sigma_{x,i}^2)} \tag{3.21}$$

$$E(y_i | x_i, k) = \alpha + \left(\frac{\beta \tau_k^2 + \sigma_{xy,i}}{\tau_k^2 + \sigma_{x,i}^2} \right) x_i + \left(\frac{\beta \sigma_{x,i}^2 - \sigma_{xy,i}}{\tau_k^2 + \sigma_{x,i}^2} \right) \mu_k \tag{3.22}$$

$$Var(y_i | x_i, k) = \beta^2 \tau_k^2 + \sigma^2 + \sigma_{y,i}^2 - \frac{(\beta \tau_k^2 - \sigma_{xy,i})^2}{\tau_k^2 + \sigma_{x,i}^2}. \tag{3.23}$$

Here, γ_k can be interpreted as the probability that the ith data point was drawn from the kth Gaussian given x_i, $E(y_i | x_i, k)$ gives the expectation value of y_i at x_i, given that the data point was drawn from the kth Gaussian, and $Var(y_i | x_i, k)$ gives the variance in y_i at x_i, given that the data point was drawn from the kth Gaussian.

3.5.2 Relationship between Uniformly Distributed Covariates and Effective χ^2 Estimators

It is informative to investigate the case where the distribution of ξ is assumed to be uniform, $p(\xi) \propto 1$. Interpreting $p(\xi)$ as a 'prior' on ξ, one may be tempted to consider assuming $p(\xi) \propto 1$ as a more objective alternative to the normal distribution. A uniform distribution for ξ may be obtained as the limit $\tau^2 \to \infty$, and thus the likelihood function for $p(\xi) \propto 1$ can be calculated from Equation (3.20) by taking $\tau^2 \to \infty$ and $K = 1$. When the measurement errors are uncorrelated, the likelihood for uniform $p(\xi)$ is

$$p(y|x, \theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi(\sigma^2 + \sigma_{y,i}^2 + \beta^2 \sigma_{x,i}^2)}} \exp\left\{ -\frac{1}{2} \frac{(y_i - \alpha - \beta x_i)^2}{\sigma^2 + \sigma_{y,i}^2 + \beta^2 \sigma_{x,i}^2} \right\}. \tag{3.24}$$

The argument of the exponential is the FITEXY goodness of fit statistic, χ_{EXY}^2, as modified by Tremaine et al. (2002) to account for intrinsic scatter; this fact has also been recognized by Weiner et al. (2006). Despite this connection, minimizing χ_{EXY}^2 is not the same as maximizing the conditional likelihood of y given x, as both β and σ^2 appear in the normalization of the likelihood function as well.

For a given value of σ^2, minimizing χ_{EXY}^2 can be interpreted as minimizing a weighted sum of squared errors, where the weights are given by the variances in y_i at a given x_i, and one assumes a uniform distribution for ξ. Unfortunately, this is only valid for a fixed value of σ^2. Moreover, little is known about the statistical properties of the FITEXY estimator, such as its bias and variance, although bootstrapping (e.g., Efron, 1979; Davison & Hinkley, 1997) may be used to estimate them. Furthermore, it is ambiguous how to calculate the FITEXY estimates when there is an intrinsic scatter term. The FITEXY goodness-of-fit statistic, χ_{EXY}^2, cannot be simultaneously minimized with respect to α, β, and σ^2, as χ_{EXY}^2 is a strictly decreasing function of σ^2. As such, it is unclear how to proceed in the

optimization beyond an *ad hoc* approach. Many authors have followed the approach adopted by Tremaine et al. (2002) and increase σ^2 until $\chi^2_{EXY}/(n-2) = 1$, or assume $\sigma^2 = 0$ if $\chi^2_{EXY}/(n-2) < 1$.

Despite the fact that minimizing χ^2_{EXY} is not the same as maximizing Equation (3.24), one may still be tempted to calculate a MLE based on Equation (3.24). However, it can be shown that if one assumes $p(\xi) \propto 1$, and if all of the x and y have the same respective measurement error variances, σ^2_x and σ^2_y, the MLE estimates for α and β are just the ordinary least squares estimates (Zellner, 1971). While this is not necessarily true when the magnitudes of the measurement errors vary between data points, one might expect that the MLE will behave similarly to the OLS estimate. I confirm this fact using simulation in § 3.8.1. Unfortunately, this implies that the MLE for $p(\xi) \propto 1$ inherits the bias in the OLS estimate, and thus nothing is gained. Furthermore, as argued by Gull (1989), one can easily be convinced that assuming $p(\xi) \propto 1$ is incorrect by examining a histogram of x.

3.5.3 Regression with Multiple Independent Variables

The formalism developed in § 3.5.1 can easily be generalized to multiple independent variables. In this case Equation (3.1) becomes

$$\eta_i = \alpha + \beta^T \xi_i + \epsilon_i, \tag{3.25}$$

where β is now a p-element vector and ξ_i is a p-element vector containing the values of the independent variables for the ith data point. Similar to before, we assume that the distribution of ξ_i can be approximated using a mixture of K multivariate normal densities with p-element mean vectors $\mu = (\mu_1, \ldots, \mu_K)$, $p \times p$ covariance matrices $T = (T_1, \ldots, T_K)$, and weights $\pi = (\pi_1, \ldots, \pi_K)$. The measured value of ξ_i is the p-element vector $\mathbf{x}_i$, and the Gaussian measurement errors

on $(y_i, \mathbf{x}_i)$ have $(p+1) \times (p+1)$ covariance matrix Σ_i. The statistical model is then

$$\xi_i \ \sim \ \sum_{k=1}^{K} \pi_k N_p(\mu_k, T_k) \tag{3.26}$$

$$\eta_i | \xi_i \ \sim \ N(\alpha + \beta^T \xi_i, \sigma^2) \tag{3.27}$$

$$y_i, \mathbf{x}_i | \eta_i, \xi_i \ \sim \ N_{p+1}([\eta_i, \xi_i], \Sigma_i). \tag{3.28}$$

$$\tag{3.29}$$

Denoting $\mathbf{z}_i = (y_i, \mathbf{x}_i)$, the measured data likelihood is

$$p(x, y | \theta, \psi) \ = \ \prod_{i=1}^{n} \sum_{k=1}^{K} \frac{\pi_k}{(2\pi)^{(p+1)/2} |V_{k,i}|^{1/2}} \exp\left\{ -\frac{1}{2} (\mathbf{z}_i - \zeta_k)^T V_{k,i}^{-1} (\mathbf{z}_i - \zeta_k) \right\} \tag{3.30}$$

$$\zeta_k \ = \ (\alpha + \beta^T \mu_k, \mu_k) \tag{3.31}$$

$$V_{k,i} \ = \ \begin{pmatrix} \beta^T T_k \beta + \sigma^2 + \sigma_{y,i}^2 & \beta^T T_k + \sigma_{xy,i}^T \\ T_k \beta + \sigma_{xy,i} & T_k + \Sigma_{x,i} \end{pmatrix}. \tag{3.32}$$

Here, ζ_k is the $(p+1)$-element mean vector of $\mathbf{z}_i$ for Gaussian k, $V_{k,i}$ is the $(p+1) \times (p+1)$ covariance matrix of $\mathbf{z}_i$ for Gaussian k, $\sigma_{y,i}^2$ is the variance in the measurement error on y_i, $\sigma_{xy,i}$ is the p-element vector of covariances between the measurement errors on y_i and $\mathbf{x}_i$, and $\Sigma_{x,i}$ is the $p \times p$ covariance matrix of the measurement errors on $\mathbf{x}_i$.

Similar to the case for one independent variable, the measured data likelihood can be decomposed as $p(x, y | \theta, \psi) = p(y | x, \theta, \psi) p(x | \psi)$, where $p(\mathbf{x}_i | \psi) = \sum_{k=1}^{K} \pi_k N_p(\mathbf{x}_i | \mu_k, T_k + \Sigma_{x,i})$ and

$$p(y_i | \mathbf{x}_i, \theta, \psi) \ = \ \sum_{k=1}^{K} \frac{\gamma_k}{\sqrt{2\pi Var(y_i | \mathbf{x}_i, k)}} \exp\left\{ -\frac{1}{2} \frac{[y_i - E(y_i | \mathbf{x}_i, k)]^2}{Var(y_i | \mathbf{x}_i, k)} \right\} \tag{3.33}$$

$$\gamma_k \ = \ \frac{\pi_k N(\mathbf{x}_i | \mu_k, T_k + \Sigma_{x,i})}{\sum_{j=1}^{K} \pi_j N(\mathbf{x}_i | \mu_j, T_j + \Sigma_{x,i})} \tag{3.34}$$

$$E(y_i | \mathbf{x}_i, k) \ = \ \alpha + \beta^T \mu_k + (\beta^T T_k + \sigma_{xy,i}^T)(T_k + \Sigma_{x,i})^{-1}(\mathbf{x}_i - \mu_k) \tag{3.35}$$

$$Var(y_i | \mathbf{x}_i, k) \ = \ \beta^T T_k \beta + \sigma^2 + \sigma_{y,i}^2 - (\beta^T T_k + \sigma_{xy,i}^T)(T_k + \Sigma_{x,i})^{-1}(T_k \beta + \sigma_{xy,i}) \tag{3.36}$$

3.6 DATA COLLECTION ISSUES: SELECTION EFFECTS AND NON-DETECTIONS

There are several issues common in the collection of astronomical data that violate the simple assumptions made in § 3.5. Astronomical data collection consists almost entirely of passive observations, and thus selection effects are a common concern. Instrumental detection limits often result in the placement of upper or lower limits on quantities, and astronomical surveys are frequently flux-limited. In this section I modify the likelihood functions described in § 3.5 to include the effects of data collection.

General methods for dealing with missing data are described in Little & Rubin (2002) and Gelman et al. (2004), and I apply the methodology described in these references to the measurement error model developed here. Although in this work I focus on linear regression, many of these results can be applied to more general statistical models, such as estimating luminosity functions.

3.6.1 Selection Effects

Suppose that one collects a sample of n sources out of a possible N sources. One is interested in understanding how the observable properties of these sources are related, but is concerned about the effects of the selection procedure on the data analysis. For example, one may perform a survey that probes some area of the sky. There are N sources located within this solid angle, where N is unknown. Because of the survey's selection method, the sample only includes n sources. In this case the astronomer is interested in how measurement error and the survey's selection method affect statistical inference.

I investigate selection effects within the framework of our statistical model by introducing an indicator variable, I, which denotes whether a source is included in the sample. If the ith source is included in the sample, then $I_i = 1$, otherwise

$I_i = 0$. In addition, I assume that the selection function only depends on the measured values, x and y. Under this assumption, the selection function of the sample is the probability of including a source with a given x and y, $p(I|x, y)$. This is commonly the case in astronomy, where sources are collected based on their measured properties. For example, one may select sources for a sample based on their measured properties as reported in the literature. In addition, if one performs a flux-limited survey then a source will only be considered detected if its measured flux falls above some set flux limit. If a sample is from a survey with a simple flux limit, then $p(I_i = 1|y_i) = 1$ if the measured source flux y_i is above the flux limit, and $p(I_i = 1|y_i) = 0$ if the measured source flux is below the flux limit. Since the selection function depends on the measured flux value, and not the true flux value, sources with true flux values above the flux limit can be missed by the survey, and sources with true flux below the limit can be detected by the survey. This effect is well-known in astronomy and is commonly referred to as Malmquist bias (e.g., Landy & Szalay, 1992).

Including the variable I, the complete data likelihood can be written as

$$p(x, y, \xi, \eta, I|\theta, \psi) = p(I|x, y)p(x, y|\xi, \eta)p(\eta|\xi, \theta)p(\xi|\psi). \tag{3.37}$$

Equation (3.37) is valid for any number of independent variables, and thus x_i and ξ_i may be either scalar or vector. Integrating Equation (3.37) over the missing data, the observed data likelihood is

$$p(x_{obs}, y_{obs}|\theta, \psi, N) \propto \binom{N}{n} \prod_{i \in \mathcal{A}_{obs}} p(x_i, y_i|\theta, \psi)$$

$$\times \prod_{j \in \mathcal{A}_{mis}} \int p(I_j = 0|x_j, y_j)p(x_j, y_j|\xi_j, \eta_j)p(\eta_j|\xi_j, \theta)p(\xi_j|\psi) \, dx_j \, dy_j \, d\xi_j \, d\eta_j \tag{3.38}$$

Here, $\binom{N}{n}$ is the binomial coefficient, $\mathcal{A}_{obs}$ denotes the set of n included sources,

x_{obs} and y_{obs} denote the values of x and y for the included sources, and $\mathcal{A}_{mis}$ denotes the set of $N - n$ missing sources. In addition, I have omitted terms that do not depend on θ, ψ, or N. Note that N is unknown and is thus also a parameter of the statistical model. The binomial coefficient is necessary because it gives the number of possible ways to select a sample of n sources from a set of N sources.

It is apparent from Equation (3.38) that statistical inference on the regression parameters is unaffected if the selection function is independent of y and x. (e.g., Little & Rubin, 2002; Gelman et al., 2004). In this case the selection function may be ignored.

3.6.1.1 Selection Based on Measured Independent Variables

It is commonly the case that a sample is selected based only on the measured independent variables. For example, suppose one performs a survey in which all sources with measured optical flux greater than some threshold are included. Then, these optically selected sources are used to fit a regression in order to understand how the X-ray luminosity of these objects depends on their optical luminosity and redshift. In this case, the probability of including a source only depends on the measured values of the optical luminosity and redshift, and is thus independent of the X-ray luminosity.

When the sample selection function is independent of y, given x, then $p(I|x, y) = p(I|x)$. Because we are primarily interested in the regression parameters, θ, I model the distributions of ξ for the included and missing sources seperately, with the parameters for the distribution of included sources denoted as ψ_{obs}. In addition, I assume that the measurement errors between y and x are statistically independent. Then the $N - n$ integrals over y and η for the missing sources in Equation (3.38) are equal to unity, and we can write the observed data likelihood

as

$$p(x_{obs}, y_{obs}|\theta, \psi_{obs}) \propto \prod_{i=1}^{n} \int \int p(x_i|\xi_i)p(y_i|\eta_i)p(\eta_i|\xi_i, \theta)p(\xi_i|I_i = 1, \psi_{obs}) \; d\xi_i \; d\eta_i,$$

(3.39)

where $p(\xi_i|I_i = 1, \psi_{obs})$ is the distribution of those ξ included in one's sample. Here I have omitted terms depending on N because one is primarily interested in inference on the regression parameters, θ. Equation (3.39) is identical to Equation (3.11), with the exception that $p(\xi|\psi)$ now only models the distribution of those ξ that have been included in one's sample, and I have now assumed that the measurement errors on y and x are independent. In particular, for the Gaussian mixture models described in § 3.5.1 and § 3.5.3, the observed data likelihood is given by Equations (3.16) and (3.30), where π, μ, and τ^2 (or T) should be understood as referring to the parameters for the distribution of the observed ξ. As is evident from the similarity between Equations (3.39) and (3.11), *if the sample is selected based on the measured independent variables, and if the measurement errors on the dependent and independent variables are statistically independent, then inference on the regression parameters, θ, is unaffected by selection effects.*

3.6.1.2   Selection Based on Measured Dependent and Independent Variables

If the method in which a sample is selected depends on the measured dependent variable, y, or if the measurement error in x and y are correlated, the observed data likelihood becomes more complicated. As an example, one might encounter this situation if one uses an X-ray selected sample to investigate the dependence of X-ray luminosity on optical luminosity and redshift. In this case, the selection function of the sample depends on both the X-ray luminosity and redshift, and is thus no longer independent of the dependent variable. Such data sets are said to be 'truncated'.

If the selection function depends on y, or if the measurement errors on y and x

are not independent, one cannot simply ignore the terms depending on N, since the $N - n$ integrals in Equation (3.38) depend on θ. However, we can eliminate the dependence of Equation (3.38) on the unknown N by applying a Bayesian approach. The posterior distribution of θ, ψ, and N is related to the observed data likelihood function as $p(\theta, \psi, N | x_{obs}, y_{obs}) \propto p(\theta, \psi, N) p(x_{obs}, y_{obs} | \theta, \psi, N)$, where $p(\theta, \psi, N)$ is the prior distribution of (θ, ψ, N). If we assume a uniform prior on θ, ψ, and $\log N$, then one can show (e.g., Gelman et al., 2004) that the posterior distribution of θ and ψ is

$$p(\theta, \psi | x_{obs}, y_{obs}) \propto [p(I = 1 | \theta, \psi)]^{-n} \prod_{i=1}^{n} p(x_i, y_i | \theta, \psi). \tag{3.40}$$

Here, $p(x_i, y_i | \theta, \psi)$ is given by Equation (3.11), and $p(I = 1 | \theta, \psi)$ is the probability of including a source in one's sample, given the model parameters, θ and ψ:

$$p(I = 1 | \theta, \psi) = \int \int p(I = 1 | x, y) p(x, y | \theta, \psi) \, dx \, dy. \tag{3.41}$$

I have left off the subscripts for the data points in Equation (3.41) because the integrals are the same for each $(x_j, y_j, \xi_j, \eta_j)$. If one assumes the Gaussian mixture model of Sections 3.5.1 and 3.5.3, then $p(x_i, y_i | \theta, \psi)$ is given by Equations (3.16) or (3.30). The posterior mode can then be used as an estimate of θ and ψ, which is found by maximizing Equation (3.40).

3.6.2 Non-detections

In addition to issues related to the sample selection method, it is common in astronomical data to have non-detections. Such non-detections are referred to as 'censored' data, and the standard procedure is to place an upper and/or lower limit on the censored data point. Methods of data analysis for censored data have been reviewed and proposed in the astronomical literature, (e.g., Feigelson & Nelson, 1985; Schmitt, 1985; Marshall, 1992; Akritas & Siebert, 1996), and Isobe

et al. (1986) describe censored regression when the variables are measured without error. See Feigelson (1992) for a review of censored data in astronomy.

To facilitate the inclusion of censored data, I introduce an additional indicator variable, D, indicating whether a data point is censored or not on the dependent variable. If y_i is detected, then $D_i = 1$, else if y_i is censored then $D_i = 0$. It is commonly the case that a source is considered 'detected' if its measured flux falls above some multiple of the background noise level, say 3σ. Then, in this case, the probability of detecting the source given the measured source flux y_i is $p(D_i = 1|y_i) = 1$ if $y_i > 3\sigma$, and $p(D_i = 0|y_i) = 1$ if $y_i < 3\sigma$. Since source detection depends on the measured flux, some sources with intrinsic flux η above the flux limit will have a measured flux y that falls below the flux limit. Similarly, some sources with intrinsic flux below the flux limit will have a measured flux above the flux limit.

I assume that a sample is selected based on the independent variables, i.e., $p(I|x, y) = p(I|x)$. It is difficult to imagine obtaining a censored sample if the sample is selected based on its dependent variable, as some of the values of y are censored and thus unknown. Therefore, I only investigate the effects of censoring on y when the probability that a source is included in the sample is independent of y, given x. In addition, I do not address the issue of censoring on the independent variable. Although such methods can be developed, it is probably simpler to just omit such data as inference on the regression parameters is unaffected when a sample is selected based only on the independent variables (cf., § 3.6.1).

The observed data likelihood for an x-selected sample is given by Equation (3.39). We can modify this likelihood to account for censored y by including the indicator variable D and again integrating over the missing data:

$$p(x_{obs}, y_{obs}, D|\theta, \psi_{obs}) \propto$$

$$\prod_{i \in \mathcal{A}_{det}} p(x_i, y_i | \theta, \psi_{obs}) \prod_{j \in \mathcal{A}_{cens}} p(x_j | \psi_{obs}) \int p(D_j = 0 | y_j, x_j) p(y_j | x_j, \theta, \psi_{obs}) \, dy_j \qquad (3.42)$$

Here, the first product is over the set of data points with detections, $\mathcal{A}_{det}$, and the second product is over the set of data points with non-detections, $\mathcal{A}_{cens}$. The conditional distribution $p(y_j | x_j, \theta, \psi_{obs})$ and the marginal distribution $p(x_j | \psi_{obs})$ for the Gaussian mixture model are both given in § 3.5.1 and § 3.5.3. If the data points are measured without error and one assumes the normal regression model, $p(\eta | \xi, \theta) = N(\eta | \alpha + \beta \xi, \sigma^2)$, then Equation 3.42 becomes the censored data likelihood function described in Isobe et al. (1986). A MLE for censored regression with measurement errors is then obtained by maximizing Equation (3.42).

3.7   COMPUTATIONAL METHODS

In this section I describe a Bayesian method for computing estimates of the regression parameters, θ, and their uncertainties. The Bayesian approach calculates the posterior probability distribution of the model parameters, given the observed data, and therefore is accurate for both small and large sample sizes. The posterior distribution follows from Baye's formula as $p(\theta, \psi | x, y) \propto p(\theta, \psi) p(x, y | \theta, \psi)$, where $p(\theta, \psi)$ is the prior distribution of the parameters. I describe some Markov Chain methods for drawing random variables from the posterior, which can then be used to estimate quantities such as standard errors and confidence intervals on θ and ψ. Gelman et al. (2004) is a good reference on Bayesian methods, and Loredo (1992) gives a review of Bayesian methods intended for astronomers. Further details of Markov Chain simulation, including methods for making the simulations more efficient, can be found in Gelman et al. (2004).

3.7.1   The Prior Density

In order to ensure a proper posterior for the Gaussian mixture model, it is necessary to invoke a proper prior density on the mixture parameters (Roeder &

Wasserman, 1997). I adopt a uniform prior on the regression parameters $(\alpha, \beta, \sigma^2)$, and take $\pi_1, \ldots, \pi_K \sim \text{Dirichlet}(1, \ldots, 1)$. The Dirichlet density is a multivariate extension of the Beta density, and the $\text{Dirichlet}(1, \ldots, 1)$ prior adopted in this work is equivalent to a uniform prior on π, under the constraint $\sum_{k=1}^{K} \pi_k = 1$.

The prior on μ and τ^2 (or T) adopted in this work is very similar to that advocated by Roeder & Wasserman (1997) and Carroll et al. (1999). I adopt a normal prior on the individual μ_k with mean μ_0 and variance u^2 (or covariance matrix U). This reflects our prior belief that the distribution of ξ is more likely to be fairly unimodal, and thus that we expect it to be more likely that the individual Gaussians will be close together than far apart. If there is only one covariate, then I adopt a scaled inverse-χ^2 prior on the individual τ_k^2 with scale parameter w^2 and one degree of freedom, otherwise if there are $p > 1$ covariates I adopt an inverse-Wishart prior on the individual T_k with scale matrix W and p degrees of freedom. This reflects our prior expectation that the variances for the individual Gaussian components should be similar, but the low number of degrees of freedom accomodates a large range of scales. Both the Gaussian means and variances are assumed to be independent in their prior distribution, and the 'hyper-parameters' μ_0, u^2 (or U), and w^2 (or W) are left unspecified. By leaving the parameters for the prior distribution unspecified, they becomes additional parameters in the statistical model, and therefore are able to adapt to the data.

Since the hyper-parameters are left as free parameters they also require a prior density. I assume a uniform prior on μ_0 and w^2 (or W). If there is one covariate, then I assume a scaled inverse-χ^2 prior for u^2 with scale parameter w^2 and one degree of freedom, otherwise if there are multiple covariate we assume a inverse-Wishart prior for U with scale matrix W and p degrees of freedom. The prior on u^2 (or U) reflects the prior expectation that the dispersion of the Gaussian

components about their mean μ_0 should be on the order of the typical dispersion of each individual Gaussian. The prior density for one covariate is then $p(\theta, \psi, \mu_0, u^2, w^2) \propto p(\pi)p(\mu|\mu_0, u^2)p(\tau^2|w^2)p(u^2|w^2)$ and is summarized hierarchically as

$$\alpha, \beta \sim \text{Uniform}(-\infty, \infty) \tag{3.43}$$

$$\sigma^2 \sim \text{Uniform}(0, \infty) \tag{3.44}$$

$$\pi \sim \text{Dirichlet}(1, \ldots, 1) \tag{3.45}$$

$$\mu_1, \ldots, \mu_K|\mu_0, u^2 \sim N(\mu_0, u^2) \tag{3.46}$$

$$\tau_1^2, \ldots, \tau_K^2, u^2|w^2 \sim \text{Inv-}\chi^2(1, w^2) \tag{3.47}$$

$$\mu_0 \sim \text{Uniform}(-\infty, \infty) \tag{3.48}$$

$$w^2 \sim \text{Uniform}(0, \infty). \tag{3.49}$$

The prior density for multiple covariates is just the multivariate extension of Equations (3.43)–(3.49).

3.7.2 Markov Chains for Sampling from the Posterior Distribution

The posterior distribution summarizes our knowledge about the parameters in the statistical model, given the observed data and the priors. Direct computation of the posterior distribution is too computationally intensive for the model described in this work. However, we can obtain any number of random draws from the posterior using Markov chain monte carlo (MCMC) methods. In MCMC methods, we simulate a Markov chain that performs a random walk through the parameter space, saving the locations of the walk at each iteration. Eventually, the Markov chain converges to the posterior distribution, and the saved parameter values can be treated as a random draw from the posterior. The random draws can then be used to estimate posterior medians, standard errors, of plot histogram estimates of the posterior.

3.7.2.1 Gibbs Sampler for the Gaussian Model

The easiest method for sampling from the posterior is to construct a Gibbs sampler. The basic idea behind the Gibbs sampler is to construct a Markov Chain, where new values of the model parameters and missing data are simulated at each iteration, conditional on the values of the observed data and the current values of the model parameters and the missing data. Within the context of the measurement error model considered in this work, the Gibbs Sampler undergoes four different stages.

The first stage of the Gibbs sampler simulates values of the missing data, given the measured data and current parameter values, a process known as data augmentation. In this work the missing data are η, ξ, and any non-detections. In addition, I introduce an additional latent variable, $\mathbf{G}_i$, which gives the class membership for the ith data point. The vector $\mathbf{G}_i$ has K elements, where $G_{ik} = 1$ if the ith data point comes from the kth Gaussian, and $G_{ij} = 0$ if $j \neq k$. I will use G to refer to the set of n vectors $\mathbf{G}_i$. Noting that π_k gives the probability of drawing a data point from the kth Gaussian, the mixture model for ξ may then be expressed hierarchically as

$$\mathbf{G}_i|\pi \quad \sim \quad \text{Multinom}(1, \pi_1, \ldots, \pi_K) \tag{3.50}$$

$$\xi_i|G_{ik} = 1, \mu_k, \tau_k^2 \quad \sim \quad N(\mu_k, \tau_k^2). \tag{3.51}$$

Here, $\text{Multinom}(m, p_1, \ldots, p_K)$ is a multinomial distribution with m trials, where p_k is the probability of success for the kth class on any particular trial. The vector $\mathbf{G}_i$ is also considered to be missing data, and is introduced to simplify construction of the Gibbs sampler.

The new values of the missing data simulated in the data augmentation step are then used to simulate new values of the regression and Gaussian mixture

parameters. The second stage of the Gibbs sampler simulates values of the regression parameters, θ, given the current values of xi and η. The third stage simulates values of the mixture parameters, ψ, given the current values of ξ and η. The fourth stage uses the new values of θ and ψ to update the parameters of the prior density. The values of the parameters are saved, and the process is repeated, creating a Markov Chain. After a large number of iterations, the Markov Chain converges, and the saved values of θ and ψ from the latter part of the algorithm may then be treated as a random draw from the posterior distribution, $p(\theta, \psi | x, y)$.

Methods for simulating random variables from the distributions used for this Gibbs sampler are described in various works (e.g., Ripley, 1987; Press et al., 1992; Gelman et al., 2004).

A Gibbs sampler for the Gaussian mixture model is

1. Start with initial guesses for η, G, θ, ψ, and the prior parameters.

2. If there are any non-detections, then draw y_i for the censored data points from $p(y_i | \eta_i, D_i = 0) \propto p(D_i = 0 | y_i) p(y_i | \eta_i)$. This may be done by first drawing y_i from $p(y_i | \eta_i)$:

$$y_i | \eta_i \sim N(\eta_i, \sigma_{y,i}^2). \tag{3.52}$$

   One then draws a random variable u_i, uniformly-distributed on $[0, 1]$. If $u_i < p(D_i = 0 | y_i)$ then the value of y_i is kept, otherwise one draws a new value of y_i and u_i until $u_i < p(D_i = 0 | y_i)$.

3. Draw values of ξ from $p(\xi | x, y, \eta, G, \theta, \psi)$. The distribution $p(\xi | x, y, \eta, G, \theta, \psi)$ can be derived from Equations (3.12)–(3.14) or (3.26)–(3.28) and the properties of the multivariate normal distribution:

(a) If there is only one independent variable then ξ_i is updated as:

$$\xi_i | x_i, y_i, \eta_i, \mathbf{G}_i, \theta, \psi \quad \sim \quad N(\hat{\xi}_i, \sigma_{\hat{\xi},i}^2) \tag{3.53}$$

$$\hat{\xi}_i \quad = \quad \sum_{k=1}^{K} G_{ik} \hat{\xi}_{ik} \tag{3.54}$$

$$\hat{\xi}_{ik} \quad = \quad \sigma_{\hat{\xi},i}^2 \left[\frac{\hat{\xi}_{xy,i}}{\sigma_{x,i}^2(1 - \rho_{xy,i}^2)} + \frac{\beta(\eta_i - \alpha)}{\sigma^2} + \frac{\mu_k}{\tau_k^2} \right] \tag{3.55}$$

$$\hat{\xi}_{xy,i} \quad = \quad x_i + \frac{\sigma_{xy,i}}{\sigma_{y,i}^2}(\eta_i - y_i) \tag{3.56}$$

$$\sigma_{\hat{\xi},i}^2 \quad = \quad \sum_{k=1}^{K} G_{ik} \sigma_{\hat{\xi},ik}^2 \tag{3.57}$$

$$\sigma_{\hat{\xi},ik}^2 \quad = \quad \left[\frac{1}{\sigma_{x,i}^2(1 - \rho_{xy,i}^2)} + \frac{\beta^2}{\sigma^2} + \frac{1}{\tau_k^2} \right]^{-1}. \tag{3.58}$$

Here, $\rho_{xy,i} = \sigma_{xy,i}/(\sigma_{x,i}\sigma_{y,i})$ is the correlation between the measurement errors on x_i and y_i. Note that ξ_i is updated using only information from the kth Gaussian, since $G_{ij} = 1$ only for $j = k$ and $G_{ij} = 0$ otherwise.

(b) If there are multiple independent variables, I have found it easier and computationally faster to update the values of ξ_i using a scalar Gibbs sampler. In this case, the p elements of ξ_i are updated individually. I denote ξ_{ij} to be the value of the jth independent variable for the ith data point, and x_{ij} to be the measured value of ξ_{ij}. In addition, I denote $\xi_{i,-j}$ to be the $(p-1)$-element vector obtained by removing ξ_{ij} from ξ_i, i.e., $\xi_{i,-j} = (\xi_{i1}, \ldots, \xi_{i(j-1)}, \xi_{i(j+1)}, \ldots, \xi_{ip})$. Similarly, β_{-j} denotes the $(p-1)$-element vector of regression coefficients obtained after removing β_j from β. Then, ξ_{ij} is updated as

$$\xi_{ij} | \mathbf{x}_i, y_i, \mathbf{G}_i, \xi_{i,-j}, \eta_i, \theta, \psi \quad \sim \quad N(\hat{\xi}_{ij}, \sigma_{\hat{\xi},ij}^2) \tag{3.59}$$

$$\hat{\xi}_{ij} \quad = \quad \sum_{k=1}^{K} G_{ik} \hat{\xi}_{ijk} \tag{3.60}$$

$$\hat{\xi}_{ijk} \quad = \quad \frac{(\Sigma_i^{-1} \mathbf{z}_i^*)_{j+1} + (T_k^{-1} \mu_{ik}^*)_j + \beta_j(\eta_i - \alpha - \beta_{-j}^T \xi_{i,-j})/\sigma^2}{(\Sigma_i^{-1})_{(j+1)(j+1)} + (T_k^{-1})_{jj} + \beta_j^2/\sigma^2} \tag{3.61}$$

$$(\mathbf{z}_i^*)_l = \begin{cases} y_i - \eta_i \text{ if } l = 1 \\ x_{il} \text{ if } l = j+1 \\ x_{il} - \xi_{il} \text{ if } l \neq j+1 \end{cases} \tag{3.62}$$

$$(\mu_{ik}^*)_l = \begin{cases} (\mu_k)_l \text{ if } l = j \\ (\mu_k)_l - \xi_{il} \text{ if } l \neq j \end{cases} \tag{3.63}$$

$$\sigma_{\hat{\xi},ij}^2 = \sum_{k=1}^{K} G_{ik} \sigma_{\hat{\xi},ijk}^2 \tag{3.64}$$

$$\sigma_{\hat{\xi},ijk}^2 = \left[(\Sigma_i^{-1})_{(j+1)(j+1)} + (T_k^{-1})_{jj} + \frac{\beta_j^2}{\sigma^2} \right]^{-1}. \tag{3.65}$$

Here, $\mathbf{z}_i^*$ is a $(p+1)$-element vector obtained by subtracting (η_i, ξ_i) from $\mathbf{z}_i = (y_i, \mathbf{x}_i)$, with the exception of the jth element of ξ_i; instead, the $(j+1)^{\text{th}}$ element of $\mathbf{z}_i^*$ is just x_{ij}. The p-element vector μ_{ik}^* is obtained in an equivalent manner. The $(p+1) \times (p+1)$ matrix Σ_i is the covariance matrix of the measurement errors on $\mathbf{z}_i$. The term $(\Sigma_i^{-1} \mathbf{z}_i^*)_{(j+1)}$ denotes the $(j+1)^{\text{th}}$ element of the vector $\Sigma_i^{-1} \mathbf{z}_i^*$, and likewise for $(T_k^{-1} \mu_{ik}^*)_j$. The terms $(\Sigma_i^{-1})_{(j+1)(j+1)}$ and $(T_k^{-1})_{jj}$ denote the $(j+1)^{\text{th}}$ and j^{th} elements of the diagonals of Σ_i^{-1} and T_k^{-1}, respectively. This step is repeated until all p independent variables have been updated for each data point.

If any of the ξ_i are measured without error, then one simply sets $\xi_i = x_i$ for those data points.

4. Draw values of η from $p(\eta|x, y, \xi, \theta)$. Similar to ξ, the distribution $p(\eta|x, y, \xi, \theta)$ can be derived from Equations (3.12)–(3.14) or (3.26)–(3.28) and the properties of the multivariate normal distribution.

   (a) If there is only one covariate then η is updated as

$$\eta_i | x_i, y_i, \xi_i, \theta \sim N(\hat{\eta}_i, \sigma_{\hat{\eta},i}^2) \tag{3.66}$$

$$\hat{\eta}_i = \sigma_{\hat{\eta},i}^2 \left[\frac{y_i + \sigma_{xy,i}(\xi_i - x_i)/\sigma_{x,i}^2}{\sigma_{y,i}^2(1 - \rho_{xy,i}^2)} + \frac{\alpha + \beta\xi_i}{\sigma^2} \right] \quad (3.67)$$

$$\sigma_{\hat{\eta},i}^2 = \left[\frac{1}{\sigma_{y,i}^2(1 - \rho_{xy,i}^2)} + \frac{1}{\sigma^2} \right]^{-1}. \quad (3.68)$$

(b) If there are multiple covariates then η is updated as

$$\eta_i | \mathbf{x}_i, y_i, \xi_i, \theta \sim N(\hat{\eta}_i, \sigma_{\hat{\eta},i}^2) \quad (3.69)$$

$$\hat{\eta}_i = \frac{(\Sigma_i^{-1}\mathbf{z}_i^*)_1 + (\alpha + \beta^T\xi_i)/\sigma^2}{(\Sigma_i^{-1})_{11} + 1/\sigma^2} \quad (3.70)$$

$$\sigma_{\hat{\eta},i}^2 = \left[(\Sigma_i^{-1})_{11} + \frac{1}{\sigma^2} \right]^{-1} \quad (3.71)$$

$$\mathbf{z}_i^* = (y_i, \mathbf{x}_i - \xi_i). \quad (3.72)$$

Here, $(\Sigma_i^{-1}\mathbf{z}_i^*)_1$ is the first element of the vector $\Sigma_i^{-1}\mathbf{z}_i^*$, $\mathbf{z}_i^*$ is a $(p+1)$-element vector whose first element is y_i and remaining elements are $\mathbf{x}_i - \xi_i$, and $(\Sigma_i^{-1})_{11}$ is the first diagonal element of Σ_i^{-1}.

If any of the η are measured without error, then one sets $\eta = y$ for those data points.

5. Draw new values of the Gaussian labels, G. The conditional distribution of $\mathbf{G}_i$ is Multinomial with number of trials $m = 1$ and group probabilities $q_k = p(G_{ik} = 1 | \xi_i, \psi)$:

$$\mathbf{G}_i | \xi_i, \psi \sim \text{Multinom}(1, q_1, \ldots, q_K) \quad (3.73)$$

$$q_k = \frac{\pi_k N_p(\xi_i | \mu_k, T_k)}{\sum_{j=1}^K \pi_j N_p(\xi_i | \mu_j, T_j)}. \quad (3.74)$$

Note that if there is only one covariate then $p = 1$ and $T_k = \tau_k^2$.

6. Draw (α, β) from $p(\alpha, \beta | \xi, \eta, \sigma^2)$. Given ξ, η, and σ^2, the distribution of α and β is obtained by ordinary regression:

$$\alpha, \beta | \xi, \eta, \sigma^2 \sim N_{p+1}(\hat{\mathbf{c}}, \Sigma_{\hat{\mathbf{c}}}) \quad (3.75)$$

$$\hat{\mathbf{c}} = (X^T X)^{-1} X^T \eta \tag{3.76}$$

$$\Sigma_{\hat{\mathbf{c}}} = (X^T X)^{-1} \sigma^2. \tag{3.77}$$

Here, X is a $n \times (p+1)$ matrix, where the first column is a column of ones, the second column contains the n values of ξ_i for the first independent variable, the third column contains the n values of ξ_i for the second independent variable, etc.

7. Draw a new value of σ^2 from $p(\sigma^2|\xi, \eta, \alpha, \beta)$. The distribution $p(\sigma^2|\xi, \eta, \alpha, \beta)$ is derived by noting that given α, β and ξ_i, η_i is normally distributed with mean $\alpha + \beta^T \xi_i$ and variance σ^2. Re-expressing this distribution in terms of σ^2 instead of η, and taking the product of the distributions for each data point, it follows that σ^2 has a scaled inverse-χ^2 distribution:

$$\sigma^2|\xi, \eta, \alpha, \beta \sim \text{Inv-}\chi^2(\nu, s^2) \tag{3.78}$$

$$\nu = n - 2 \tag{3.79}$$

$$s^2 = \frac{1}{n-2} \sum_{i=1}^{n} (\eta_i - \alpha - \beta^T \xi_i)^2. \tag{3.80}$$

8. Draw new values of the group proportions, π. Given G, π follows a Dirichlet distribution:

$$\pi|G \sim \text{Dirichlet}(n_1 + 1, \ldots, n_K + 1) \tag{3.81}$$

$$n_k = \sum_{i=1}^{n} G_{ik}. \tag{3.82}$$

Note that n_k is the number of data points that belong to the kth Gaussian.

9. Draw a new value of μ_k from $p(\mu_k|\xi, G, T_k, \mu_0, U)$. If there is only one independent variable, then $T_k = \tau_k^2$ and $U = u^2$. The new value of μ_k is simulated as

$$\mu_k|\xi, G, T_k, \mu_0, U \sim N_p(\hat{\mu}_k, \Sigma_{\hat{\mu}_k}) \tag{3.83}$$

$$\hat{\mu}_k \quad = \quad (U^{-1} + n_k T_k^{-1})^{-1}(U^{-1}\mu_0 + n_k T_k^{-1}\bar{\xi}_k) \qquad (3.84)$$

$$\bar{\xi}_k \quad = \quad \frac{1}{n_k}\sum_{i=1}^{n} G_{ik}\xi_i \qquad (3.85)$$

$$\Sigma_{\hat{\mu}_k} \quad = \quad (U^{-1} + n_k T_k^{-1})^{-1}. \qquad (3.86)$$

10. Draw a new value of τ_k^2 or T_k. The distribution of $\tau^2|\xi, \mu$ or $T_k|\xi, \mu$ is derived in a manner similar to $\sigma^2|\xi, \eta, \alpha, \beta$, and noting that the prior is conjugate for this likelihood. The distribution of $\tau_k^2|\xi, \mu$ is a scaled inverse-χ^2 distribution, and the distribution of $T_k|\xi, \mu$ is an inverse-Wishart distribution:

    (a) If there is only one independent variable then draw

$$\tau_k^2|\xi, G, \mu_k, w^2 \quad \sim \quad \text{Inv-}\chi^2(\nu_k, t_k^2) \qquad (3.87)$$

$$\nu_k \quad = \quad n_k + 1 \qquad (3.88)$$

$$t_k^2 \quad = \quad \frac{1}{n_k + 1}\left[w^2 + \sum_{i=1}^{n} G_{ik}(\xi_i - \mu_k)^2\right]. \qquad (3.89)$$

    (b) If there are multiple independent variables then draw

$$T_k|\xi, G, \mu_k, W \quad \sim \quad \text{Inv-Wishart}_{\nu_k}(S_k) \qquad (3.90)$$

$$\nu_k \quad = \quad n_k + p \qquad (3.91)$$

$$S_k \quad = \quad W + \sum_{i=1}^{n} G_{ik}(\xi_i - \mu_k)(\xi_i - \mu_k)^T. \qquad (3.92)$$

11. Draw a new value for $\mu_0|\mu, U$. Noting that conditional on μ_0 and U, $\mu_1, \ldots, \mu_K$ are independently distributed as $N_p(\mu_0, U)$, it is straight-forward to show that

$$\mu_0|\mu, U \quad \sim \quad N_p(\bar{\mu}, U/K) \qquad (3.93)$$

$$\bar{\mu} \quad = \quad \frac{1}{K}\sum_{k=1}^{K} \mu_k. \qquad (3.94)$$

If there is only one covariate then $p = 1$ and $U = u^2$.

12. Draw a new value for u^2 or U, given μ_0, μ, and w^2 (or W). Similar to the case for τ_k^2 or T_k, the conditional distribution of u^2 or U is scaled inverse-χ^2 or inverse-Wishart.

   (a) If there is only one covariate then

$$u^2|\mu_0, \mu, w^2 \quad \sim \quad \text{Inv-}\chi^2(\nu_u, \hat{u}^2) \tag{3.95}$$

$$\nu_u \quad = \quad K+1 \tag{3.96}$$

$$\hat{u}^2 \quad = \quad \frac{1}{\nu_u}\left[w^2 + \sum_{k=1}^{K}(\mu_k - \mu_0)^2\right]. \tag{3.97}$$

   (b) If there are multiple covariates then

$$U|\mu_0, \mu, W \quad \sim \quad \text{Inv-Wishart}_{\nu_U}(\hat{U}) \tag{3.98}$$

$$\nu_U \quad = \quad K+p \tag{3.99}$$

$$\hat{U} \quad = \quad W + \sum_{k=1}^{K}(\mu_k - \mu_0)(\mu_k - \mu_0)^T. \tag{3.100}$$

13. Finally, draw a new value of $w^2|u^2, \tau^2$ or $W|U, T$:

   (a) If there is only one covariate then $w^2|u^2, \tau^2$ is drawn from a Gamma distribution. This can be derived by noting that $p(w^2|u^2, \tau^2) \propto p(u^2|w^2)p(\tau^2|w^2)$ has the form of a Gamma distribution as a function of w^2. The new value of w^2 is then simulated as

$$w^2|u^2, \tau^2 \quad \sim \quad \text{Gamma}(a, b) \tag{3.101}$$

$$a \quad = \quad \frac{1}{2}(K+3) \tag{3.102}$$

$$b \quad = \quad \frac{1}{2}\left[\frac{1}{u^2} + \sum_{k=1}^{K}\frac{1}{\tau_k^2}\right]. \tag{3.103}$$

   (b) If there are multiple covariates then $W|U, T$ is drawn from a Wishart distribution. This can be derived by noting that $p(W|U, T) \propto p(U|W)p(T|W)$

has the form of a Wishart distribution as a function of W. The new value of W is then simulated as

$$W|U,T \quad \sim \quad \text{Wishart}_{\nu_W}(\hat{W}) \tag{3.104}$$

$$\nu_W \quad = \quad (K+2)p+1 \tag{3.105}$$

$$\hat{W} \quad = \quad (U^{-1} + \sum_{k=1}^{K} T_k^{-1})^{-1}. \tag{3.106}$$

After completing steps 2–13 above, an iteration of the Gibbs sampler is complete. One then uses the new simulated values of ξ, η, θ, ψ, and the prior parameters, and repeats steps 2–13. The algorithm is repeated until convergence, and the values of θ and ψ at each iteration are saved. Upon reaching convergence, one discards the values of θ and ψ from the beginning of the simulation, and the remaining values of $\alpha, \beta, \sigma^2, \mu$, and τ^2 (or T) may be treated as a random draw from the posterior distribution, $p(\theta, \psi|x, y)$. One can then use these values to calculate estimates of the parameters, and their corresponding variances and confidence intervals. The posterior distribution of the parameters can also be estimated from these values of θ and ψ using histogram techniques. Techniques for monitering convergence of the Markov Chains can be found in Gelman et al. (2004).

The output from the Gibbs sampler may be used to perform Bayesian inference on other quantities of interest. In particular, the Pearson linear correlation coefficient, ρ, is often used in assessing the strength of a relationship between the x and y. A random draw from the posterior distribution for the correlation between η and ξ_j, denoted as ρ_j, can be calculated from Equation (3.5) for each draw from the Gibbs sampler. For the Gaussian mixture model, the variance $Var(\eta)$ and covariance matrix $\Sigma_\xi \equiv Var(\xi)$ are

$$Var(\eta) \quad = \quad \beta^T \Sigma_\xi \beta + \sigma^2 \tag{3.107}$$

$$\Sigma_\xi \quad = \quad \sum_{k=1}^{K} \pi_k (T_k + \mu_k \mu_k^T) - \bar{\xi}\bar{\xi}^T \tag{3.108}$$

$$\bar{\xi} = \sum_{k=1}^{K} \pi_k \mu_k, \tag{3.109}$$

and $Var(\xi_j)$ is the jth diagonal element of Σ_ξ. The simplification for one covariate is self-evident.

If there is considerable posterior probability near $\sigma^2 \approx 0$ or $\tau_k^2 \approx 0$, then the Gibbs sampler can get 'stuck'. For example, if $\tau_k^2 \approx 0$, then step 3a of the Gibbs sampler will draw values of $\xi | G \approx \mu_k$. Then, step 9 will produce a new value of μ_k that is almost identical to the previous iteration, step 10a will produce a new value of $\tau_k^2 \approx 0$, and so on. The Gibbs sampler will eventually get 'unstuck', but this can take a long time and result in very slow convergence. In particular, it is very easy for the Gibbs sampler to get stuck if the measurement errors are large relative to σ^2 or τ_k^2, or if the number of data points is small. In this situation I have found it useful to use the Metropolis-Hastings algorithm instead.

3.7.2.2 Metropolis-Hastings Algorithm

If the selection function is not independent of y, given the independent variables (cf. Eq.[3.40]), or if the selection function depends on x and the measurement errors are correlated, then posterior simulation based on the Gibbs sampler is more complicated. In addition, if the measurement errors are large compared to the intrinsic dispersion in the data, or if the sample size is small, then the Gibbs sampler can become stuck and extremely inefficient. In both of these cases one can use the Metropolis-Hastings algorithm (Metropolis & Ulam, 1949; Metropolis et al., 1953; Hastings, 1970) to sample from the posterior distribution, as the Metropolis-Hasting algorithm can avoid constructing markov chains for ξ and η. For a description of the Metropolis-Hastings algorithm, we refer the reader to Chib & Greenberg (1995) or Gelman et al. (2004).

3.8   SIMULATIONS

In this section I perform simulations to illustrate the effectiveness of the Gaussian structural model for estimating the regression parameters, even in the presence of severe measurement error and censoring. In addition, I compare the OLS, BCES$(Y|X)$, and FITEXY estimators with a maximum-likelihood estimator based on the Gaussian mixture model with $K = 1$ Gaussian.

3.8.1   Simulation Without Non-Detections

The first simulation I performed is for a simple regression with one independent variable. I generated 2.7×10^5 data sets by first drawing n values of the independent variable, ξ, from a distribution of the form

$$p(\xi) \propto e^\xi \left(1 + e^{2.75\xi}\right)^{-1}. \tag{3.110}$$

The distribution of ξ is shown in Figure 3.2, along with the best-fitting one and two Gaussian approximations. In this case the two Gaussian mixture is nearly indistinguishable from the actual distribution of ξ, and thus should provide an excellent approximation to $p(\xi)$. The values for ξ had a mean of $\mu = -0.493$ and a dispersion of $\tau = 1.200$. I varied the number of data points in the simulated data sets as $n = 25, 50$, and 100. I then simulated values of η according to Equation (3.1), with $\alpha = 1.0$ and $\beta = 0.5$. The intrinsic scatter, ϵ, had a normal distribution with mean zero and standard deviation $\sigma = 0.75$, and the correlation between η and ξ was $\rho \approx 0.62$. The joint distribution of ξ and η for one simulated data set with $n = 50$ is shown in Figure 3.3.

Measured values for ξ and η were simulated according to Equations (3.2) and (3.3). The measurement errors had a zero mean normal distribution of varying dispersion and were independent for x and y. The variances in the measurement errors, $\sigma_{x,i}^2$ and $\sigma_{y,i}^2$, were different for each data point and drawn from a
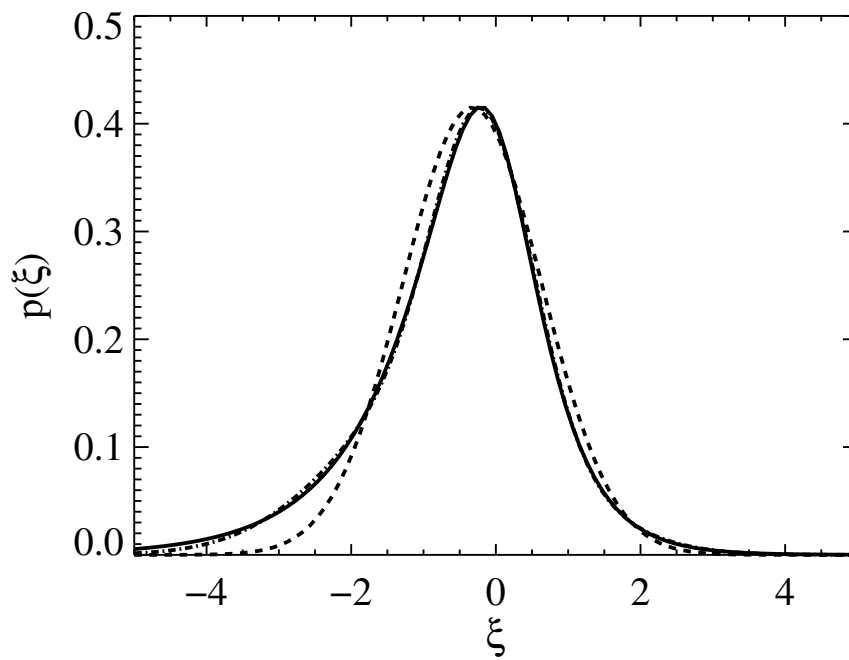
Figure 3.2 The actual distribution of ξ (solid line) for the simulations, compared with the best-fitting one (dashed line) and two (dashed-dotted line) Gaussian fit. The two Gaussian fit is nearly indistinguishable from the true $p(\xi)$. Although the one Gaussian fit provides a reasonable approximation to the distribution of ξ, it is not able to pick up the asymmetry in $p(\xi)$.
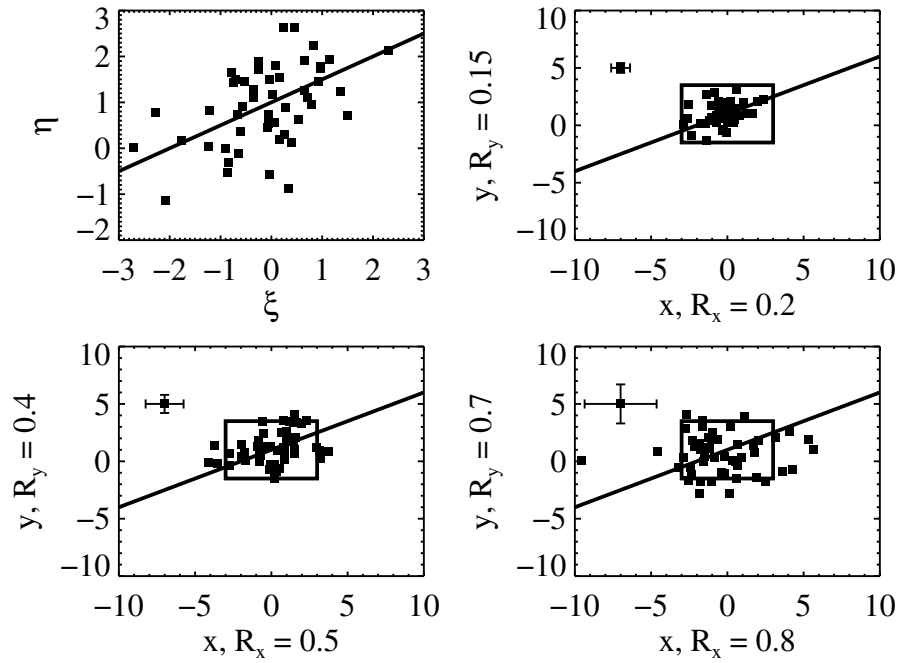
Figure 3.3 Distributions of the simulated data for various levels of measurement error (cf., § 3.8.1). The top left panel shows the distribution of η as a function of ξ for one simulated data set; the solid line is the true value of the regression line. The remaining panels show the distributions of the observed values, y and x, for various levels of measurement error. The data point with error bars in each panel is a fictitious data point and is used to illustrate the median values of the error bars. The box outlines the bounds of the plot of η against ξ. As can be seen, large measurement errors wash out any visual evidence for a correlation between the variables.

scaled inverse-χ^2 distribution. The degrees of freedom for the inverse-χ^2 distri-
bution was $\nu = 5$, and the scale parameters are denoted as t and s for the x and
y measurement error variances, respectively. The scale parameters dictate the
typical size of the measurements errors, and were varied as $t = 0.5\tau, \tau, 2\tau$ and
$s = 0.5\sigma, \sigma, 2\sigma$. These values corresponded to values of $R_x \sim 0.2, 0.5, 0.8$ and
$R_y \sim 0.15, 0.4, 0.6$ respectively. I simulated 10^4 data sets for each grid point of t, s,
and n, giving a total of 2.7×10^5 simulated data sets. The joint distributions of x
and y for varying values of t/τ and s/σ are also shown in Figure 3.3. These values
of x and y are the 'measured' values of the simulated data set shown in the plot
of η as a function of ξ.

For each simulated data set, I calculated the maximum-likelihood estimate,
found by maximizing Equation (3.16). For simplicity, I only use $K = 1$ Gaussian.
I also calculated the OLS, BCES($Y|X$), and FITEXY estimates for comparison. I
calculated a OLS estimate of σ^2 by subtracting the average σ_y^2 from the variance
in the regression residuals. If the OLS estimate of σ^2 was negative, I set $\hat{\sigma}_{OLS} = 0$.
Following Fuller (1987), I estimate σ^2 for a BCES($Y|X$)-type estimator as $\hat{\sigma}_{BCES}^2 =$
$Var(y) - \bar{\sigma}_y^2 - \hat{\beta}_{\mathrm{BCES}} Cov(x, y)$, where $\bar{\sigma}_y^2$ is the average measurement error variance
in y, and $\hat{\beta}_{\mathrm{BCES}}$ is the BCES($Y|X$) estimate of the slope. If $\hat{\sigma}_{BCES}^2$ is negative, I set
$\hat{\sigma}_{BCES} = 0$. Following Tremaine et al. (2002), I compute a FITEXY estimate of σ by
increasing σ^2 until $\chi_{EXY}^2/(n-2) = 1$, or assume $\sigma^2 = 0$ if $\chi_{EXY}^2/(n-2) < 1$. The
sampling distributions of the slope and intrinsic scatter estimators for $n = 50$ are
shown in Figures 3.4 and 3.5 as a function of t/τ, and the results of the simulations
are summarized in Table 3.1.

The bias of the OLS estimate is apparent, becoming more severe as the mea-
surement errors in the independent variable increase. In addition, the variance in
the OLS slope estimate decreases as the measurement errors in ξ increase, giving
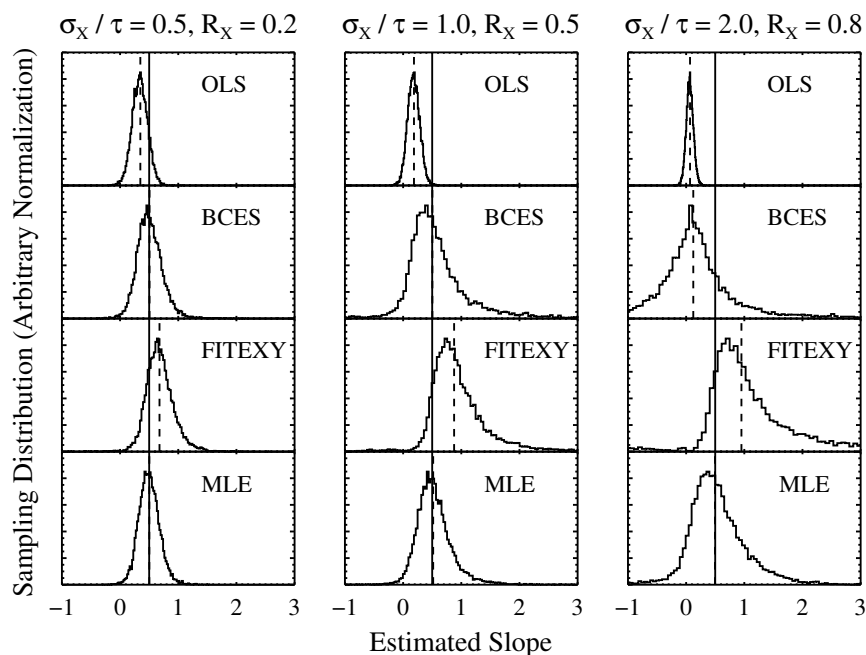
Figure 3.4 The sampling distributions of the slope estimators as a function of covariate measurement error magnitude for $n = 50$ data points and $\sigma_y \sim \sigma$, inferred from simulations (cf., § 3.8.1). The estimators are the ordinary least-squares estimator (OLS), the BCES($Y|X$) estimator, the FITEXY estimator, and the maximum-likelihood estimator (MLE) of the $K = 1$ gaussian structural model. The solid vertical lines mark the true value of $\beta = 0.5$, and the dashed vertical lines mark the median values of each respective estimator. The OLS estimator is biased toward zero, while the FITEXY estimator is biased away from zero; in both cases, the bias gets worse for larger measurement errors. The BCES($Y|X$) estimator is, in general, unbiased, but can become biased and highly variable if the measurement errors becomes large. The MLE of the Gaussian model performs better than the other estimators, as it is approximately unbiased and less variable.
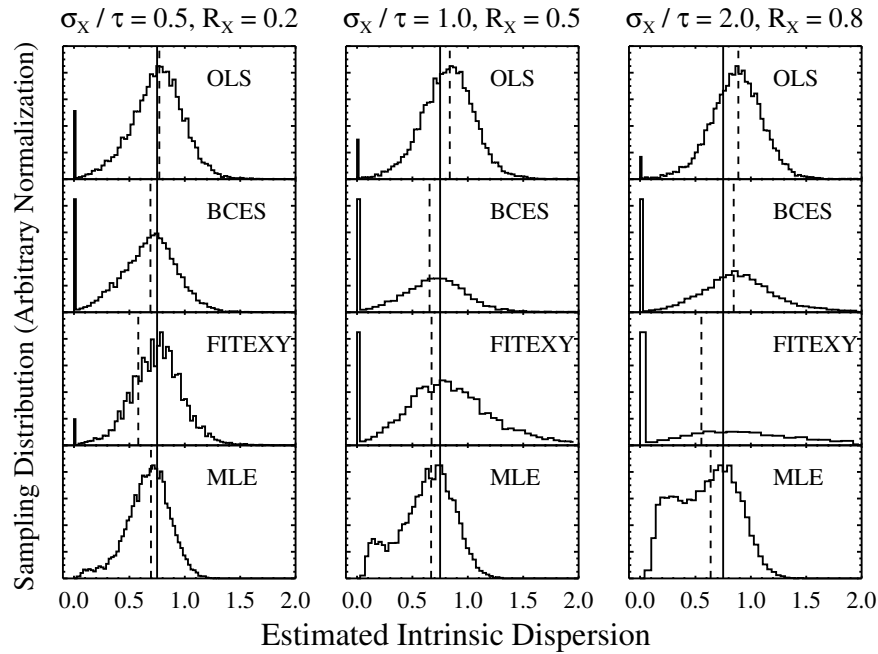
Figure 3.5 Same as Figure 3.4, but for the standard deviation of the intrinsic scatter, σ. The solid vertical lines mark the true value of $\sigma = 0.75$, and the dashed vertical lines mark the median values of each respective estimator. All of the estimators exhibit some bias, and the BCES and FITEXY estimators can exhibit significant variance. Moreover, the BCES and FITEXY estimators both commonly have values of $\hat{\sigma} = 0$, misleading one into concluding that there is no intrinsic scatter; this occasionally occurs for the OLS estimate as well. In contrast, the MLE based on the Gaussian model does not suffer from this problem, at least for these simulations.

Table 3.1. Dependence of the Estimator Sampling Distributions on

Measurement Error and Sample Size

		OLS		BCES($Y\|X$)		FITEXY		MLE	
$t/\tau = s/\sigma$[a]	n[b]	$\hat{\beta}$[c]	$\hat{\sigma}$[d]	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\beta}$	$\hat{\sigma}$
0.5	25	$0.357^{+0.242}_{-0.246}$	$0.784^{+0.717}_{-0.608}$	$0.518^{+0.513}_{-0.349}$	$0.687^{+0.714}_{-0.650}$	$0.896^{+1.127}_{-0.425}$	$0.855^{+1.616}_{-0.757}$	$0.513^{+0.393}_{-0.315}$	$0.677^{+0.663}_{-0.580}$
	50	$0.355^{+0.166}_{-0.164}$	$0.801^{+0.591}_{-0.528}$	$0.510^{+0.306}_{-0.233}$	$0.716^{+0.601}_{-0.540}$	$0.898^{+0.572}_{-0.298}$	$0.873^{+1.042}_{-0.668}$	$0.506^{+0.242}_{-0.212}$	$0.717^{+0.555}_{-0.507}$
	100	$0.354^{+0.117}_{-0.114}$	$0.810^{+0.488}_{-0.447}$	$0.506^{+0.197}_{-0.164}$	$0.743^{+0.494}_{-0.466}$	$0.895^{+0.352}_{-0.218}$	$0.885^{+0.786}_{-0.587}$	$0.504^{+0.162}_{-0.149}$	$0.732^{+0.456}_{-0.429}$
1.0	25	$0.190^{+0.255}_{-0.239}$	$0.798^{+1.047}_{-0.798}$	$0.442^{+2.763}_{-2.167}$	$0.610^{+1.418}_{-0.610}$	$0.827^{+2.293}_{-1.687}$	$0.727^{+2.899}_{-0.727}$	$0.524^{+0.907}_{-0.576}$	$0.572^{+0.903}_{-0.564}$
	50	$0.191^{+0.172}_{-0.164}$	$0.839^{+0.869}_{-0.752}$	$0.519^{+1.816}_{-0.707}$	$0.643^{+1.023}_{-0.643}$	$0.870^{+1.195}_{-0.459}$	$0.814^{+1.754}_{-0.814}$	$0.519^{+0.552}_{-0.370}$	$0.669^{+0.745}_{-0.643}$
	100	$0.189^{+0.121}_{-0.116}$	$0.862^{+0.726}_{-0.640}$	$0.520^{+0.913}_{-0.348}$	$0.687^{+0.784}_{-0.687}$	$0.895^{+0.665}_{-0.329}$	$0.855^{+1.246}_{-0.788}$	$0.502^{+0.337}_{-0.242}$	$0.714^{+0.623}_{-0.604}$
2.0	25	$0.066^{+0.243}_{-0.228}$	$0.565^{+1.797}_{-0.565}$	$0.036^{+2.761}_{-2.944}$	$0.663^{+2.544}_{-0.663}$	$0.443^{+3.793}_{-2.836}$	$0.000^{+2.994}_{-0.000}$	$0.366^{+1.468}_{-1.395}$	$0.381^{+1.223}_{-0.362}$
	50	$0.067^{+0.164}_{-0.158}$	$0.768^{+1.525}_{-0.768}$	$0.116^{+2.878}_{-2.951}$	$0.743^{+2.271}_{-0.743}$	$0.634^{+3.276}_{-3.027}$	$0.258^{+2.983}_{-0.258}$	$0.426^{+1.055}_{-0.918}$	$0.559^{+1.082}_{-0.529}$
	100	$0.065^{+0.113}_{-0.106}$	$0.843^{+1.293}_{-0.843}$	$0.209^{+2.936}_{-2.962}$	$0.743^{+1.932}_{-0.743}$	$0.765^{+2.492}_{-2.024}$	$0.627^{+2.928}_{-0.627}$	$0.444^{+0.698}_{-0.548}$	$0.673^{+0.921}_{-0.621}$

Note. — The values given for $\hat{\beta}$, and $\hat{\sigma}$ are the median and interval containing 90% of the estimates over the simulations. For example, when $t/\tau = s/\sigma = 0.5$ and $n = 25$, the median value of the OLS slope estimator is 0.357, and 90% of the values of $\hat{\beta}_{OLS}$ are contained within $0.357^{+0.242}_{-0.246}$.

[a] Typical value of the measurement error magnitude for the simulations.

[b] The number of data points in the simulated data sets.

[c] The estimate of the slope, β. The true value is $\beta = 0.5$.

[d] The estimate of the dispersion in the intrinsic scatter, σ. The true value is $\sigma = 0.75$.

one the false impression that one's estimate of the slope is more precise when the measurement errors are large. This has the effect of concentrating the OLS estimate of β around $\hat{\beta}_{OLS} \sim 0$, thus effectively erasing any evidence of a relationship between the two variables. When the measurement errors are large, the OLS estimate of the intrinsic scatter, $\hat{\sigma}^2_{OLS}$, is occasionally zero.

The BCES($Y|X$) estimator performs better than the OLS and FITEXY estimators, being approximately unbiased when the measurement errors are $\sigma_x/\tau \lesssim 1$. However, the BCES estimate of the slope, $\hat{\beta}_{\text{BCES}} = Cov(x,y)/(Var(x) - \bar{\sigma}^2_x)$, suffers some bias when the measurement errors are large and/or the sample size is small. In addition, the variance in $\hat{\beta}_{BCES}$ is larger than the MLE, and $\hat{\beta}_{BCES}$ becomes considerably unstable when the measurement errors on ξ are large. This instability results because the denominator in the equation for $\hat{\beta}_{\text{BCES}}$ is $Var(x) - \bar{\sigma}^2_x$. If $\bar{\sigma}^2_x \approx Var(x)$, then the denominator is ≈ 0, and $\hat{\beta}_{\text{BCES}}$ can become very large. Similar to the OLS and FITEXY estimates, the estimate of the intrinsic variance for the BCES-type estimator is often zero when the measurement errors are large, suggesting the false conclusion that there is no intrinsic scatter about the regression line.

The FITEXY estimator performed poorly in the simulations, being both biased and highly variable. The bias of the FITEXY estimator is such that $\hat{\beta}_{EXY}$ tends to overestimate β, the severity of which tends to increase as R_y decreases. This upward bias in $\hat{\beta}_{EXY}$ has been noted by Weiner et al. (2006), who also performed simulations comparing $\hat{\beta}_{EXY}$ with $\hat{\beta}_{BCES}$. They note that when one minimizes χ^2_{EXY} alternatively with respect to β and σ^2, and iterates until convergence, then the bias in $\hat{\beta}_{EXY}$ can be improved. I have tested this and also find that the bias in $\hat{\beta}_{EXY}$ is reduced, but at the cost of a considerable increase in variance in $\hat{\beta}_{EXY}$. In general, our simulations imply that the variance of the FITEXY estimator is

comparable to that of the BCES$(Y|X)$ estimator if one does not iterate the minimization of χ^2_{EXY}, and the variance of $\hat{\beta}_{EXY}$ is larger if one does iterate. However, since $\hat{\beta}_{BCES}$ is approximately unbiased when R_x is not too large, $\hat{\beta}_{BCES}$ should be preferred over $\hat{\beta}_{EXY}$. In addition, when the measurement errors are large the FITEXY estimate of σ is commonly $\hat{\sigma}_{EXY} = 0$, similar to the BCES-type estimate of the intrinsic dispersion.

The maximum-likelihood estimator based on the Gaussian structural model performs better than the OLS, BCES, and FITEXY estimators, and gives fairly consistent estimates even in the presence of severe measurement error and low sample size. The MLE is approximately unbiased, in spite of the fact that the MLE incorrectly assumes that the independent variables are normally distributed. The variance in the MLE of the slope, $\hat{\beta}_{MLE}$, is smaller than that of $\hat{\beta}_{BCES}$ and $\hat{\beta}_{EXY}$, particularly when R_x is large. In contrast to the OLS estimate of the slope, the dispersion in $\hat{\beta}_{MLE}$ increases as the measurement errors increases, reflecting the additional uncertainty in $\hat{\beta}_{MLE}$ caused by the measurement errors. Finally, in contrast to the other estimators, the MLE of the intrinsic variance is always positive, and the probability of obtaining $\hat{\sigma}_{MLE} = 0$ is negligible for these simulations.

I argued in § 3.5.1 that assuming a uniform distribution on ξ does not lead to better estimates than the usual OLS case. I also used these simulations to estimate the sampling density of the MLE assuming $p(\xi) \propto 1$. The results were nearly indistinguishable from the OLS estimator, supporting our conjecture that assuming $p(\xi) \propto 1$ does not offer an improvement over OLS.

While it is informative to compare the sampling distribution of our proposed maximum-likelihood estimator with those of the OLS, BCES$(Y|X)$, and FITEXY estimators, I do not derive the uncertainties in the regression parameters from the sampling distribution of the MLE. As described in § 3.7.2, we derive the uncer-

tainties in the regression parameters by simulating draws from the posterior distribution, $p(\theta, \psi | x, y)$. This allows a straight-forward method of interpreting the parameter uncertainties that does not rely on large-sample approximations, as the posterior distribution is the probability distribution of the parameters, given the observed data. The posterior distributions of ρ, β, and σ for a simulated data set with $n = 50, \sigma_x \sim \tau$, and $\sigma_y \sim \sigma$ is shown in Figure 3.6. When estimating these posteriors, I used $K = 2$ Gaussians in the mixture model. As can be seen from Figure 3.6, the true values of ρ, β, and σ are contained within the regions of non-negligible posterior probability. I have estimated posteriors for other simulated data sets, varying the number of data points and the degree of measurement error. As one would expect, the uncertainties in the regression parameters, represented by the widths of the posterior distributions, increase as the size of the measurement errors increase and the sample size decreases.

A common frequentist approach is to compute the covariance matrix of the MLE by inverting the estimated Fisher information matrix, evaluated at the MLE. Then, under certain regularity conditions, the MLE of the parameters is asympotically normally distributed with mean equal to the true value of the parameters and covariance matrix equal to the inverse of the Fisher information matrix. Furthermore, under these regularity conditions the posterior distribution and sampling distribution of the MLE are asymptotically the same. Figure 3.7 compares the posterior distribution of the slope for a simulated data set with that inferred from the MLE. The posterior and MLE was calculated assuming $K = 1$ Gaussian. As can be seen, the posterior distribution for β is considerably different from the approximation based on the MLE of β, and thus the two have not converged for this sample. In particular, the posterior is more skewed and heavy-tailed, placing more probability on values of $\beta > 0$ than does the distribution approximated by
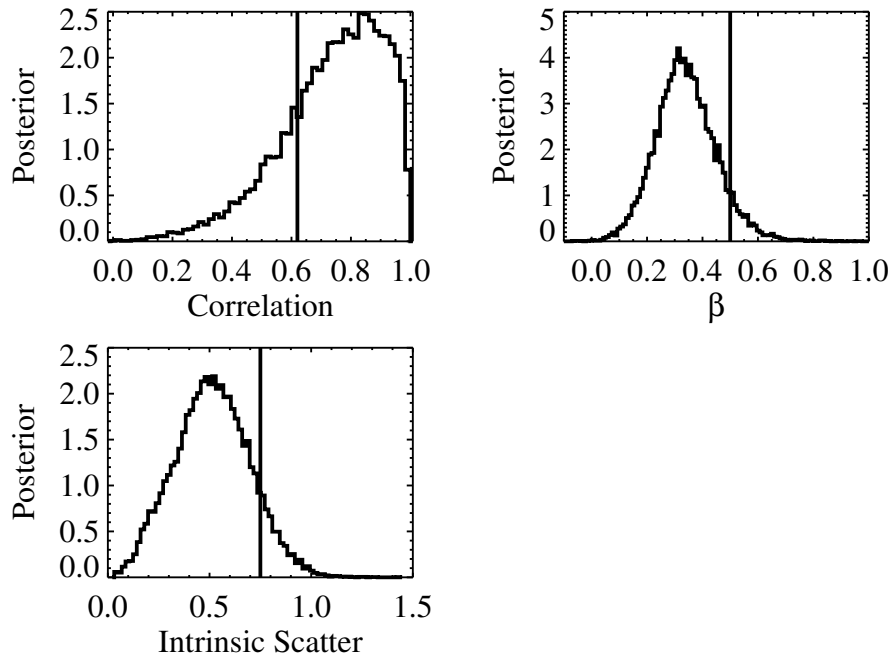
Figure 3.6 The marginal posterior distributions of the linear correlation coefficient, the regression slope, and the intrinsic dispersion for a simulated data set of $n = 50$ data points with $\sigma_x \sim \tau$ and $\sigma_y \sim \sigma$. The vertical lines mark the true values of the parameters. The true values of the regression parameters are contained within the spread of the marginal posteriors, implying that bounds on the regression parameters inferred from the posterior are trustworthy.
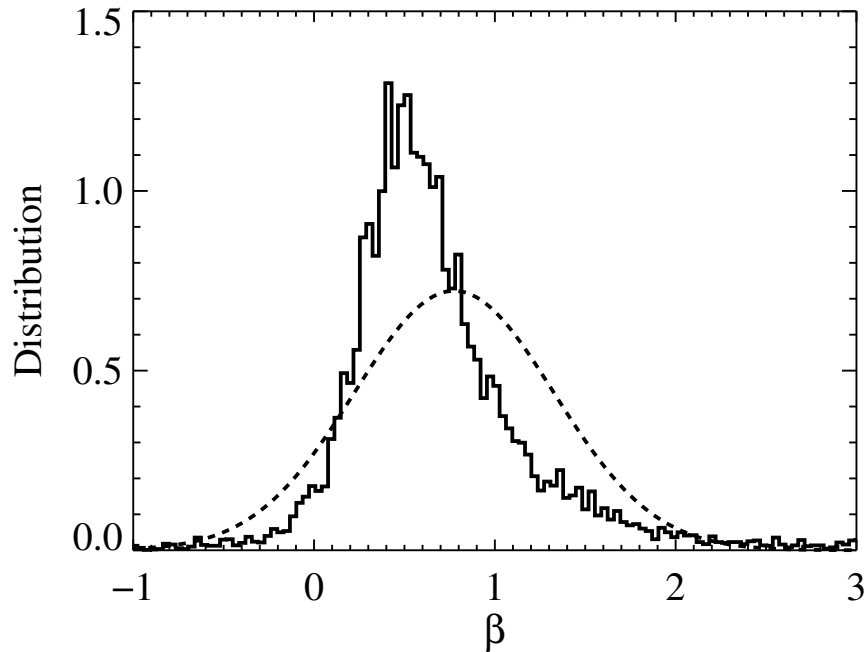
Figure 3.7 The posterior distributions of the slope (solid histogram), compared with the posterior approximated from the MLE and Fisher information matrix (dashed line), for a simulated data set of $n = 50$ data points with $\beta = 0.5, \sigma_x \sim \tau$, and $\sigma_y \sim \sigma$. The two distributions have not converged and the bayesian and frequentist inference differ in this case, with the bayesian approach placing more probability near $\beta \approx 0.5$ and on positive value of β.

the MLE. Therefore, uncertainties in the MLE should be interpreted with caution if using the asymptotic approximation to the sampling distribution of the MLE.

3.8.2   Simulation With Non-Detections

To assess the effectiveness of the Gaussian structural model in dealing with censored data sets with measurement error, I introduced non-detections into the simulations. The simulations were performed in an identical manner as that described in § 3.8.1, but now I only consider sources to be 'detected' if $y > 1.5$. For

those sources that were 'censored' ($y < 1.5$), I placed an upper limit on them of $y = 1.5$.

I focus on the results for a simulated data set with $n = 100$ data points and measurement errors similar to the intrinsic dispersion in the data, $\sigma_y \sim \sigma$ and $\sigma_x \sim \tau$. The detection threshold of $y > 1.5$ resulted in a detection fraction of $\sim 30\%$. This simulation represents a rather extreme case of large measurement errors and low detection fraction, and provides an interesting test of the method. In Figure 3.8 I show the distribution of ξ and η, as well as the distribution of their measured values, for one of the simulated data sets. For this particular data set, there were 29 detections and 71 non-detections. As can be seen, the significant censoring and large measurement errors have effectively erased any visual evidence for a relationship between the two variables.

I estimated the posterior distribution of the regression parameters for this data set using the Gibbs sampler (cf, § 3.7.2) with $K = 2$ Gaussians. The posterior median of the regression line, as well as the 95% (2σ) pointwise confidence intervals[1] on the regression line are shown in Figure 3.8. The posterior distributions for $\rho, \beta,$ and σ are shown in Figure 3.9. As can be seen, the true value of the parameters is contained within the 95% probability regions, although the uncertainty is large. For this particular data set, we can put limits on the value of the correlation coefficient as $0.2 \lesssim \rho \lesssim 1$ and the slope as $0 \lesssim \beta \lesssim 2.0$. For comparison, the usual maximum-likelihood estimate that ignores the measurement error (e.g., Isobe et al., 1986) concludes $\hat{\beta} = 0.229 \pm 0.077$. This estimate is biased and differs from the true value of β at a level of 3.5σ.

The posterior constraints on the regression parameters are broad, reflecting

---

[1]Technically, these are called 'credibility intervals', as I am employing a Bayesian approach. These intervals contain 95% of the posterior probability. While the difference between confidence intervals and credibility intervals is not purely semantical, I do not find the difference to be significant within the context of my work, so I use the more familiar term 'confidence interval'.
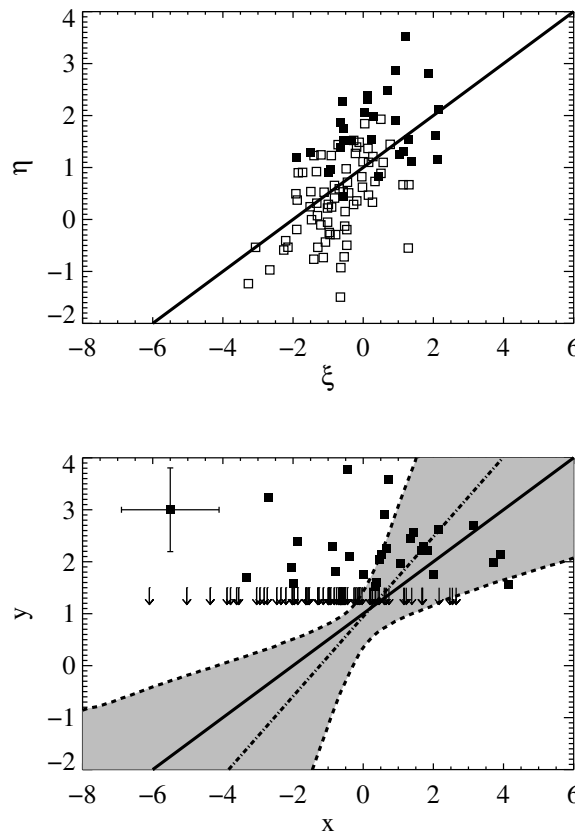
Figure 3.8 Distribution of η and ξ (top), and the measured values of y and x (bottom), from a simulated censored data set of $n = 50$ data points, $\sigma_x \sim \tau$, and $\sigma_y \sim \sigma$ (cf., § 3.8.2). In the plot of η and ξ, the solid squares denote the values of ξ and η for the detected data points, and the hollow squares denote the values of ξ and η for the undetected data points. The solid line in both plots is the true regression line. In the plot of y and x, the squares denote the measured values of x and y for the detected data points, and the arrows denote the 'upper limits' on y for the undetected data points. The fictitious data point with error bars illustrates the median values of the error bars. The dashed-dotted line shows the best fit regression line, as calculated from the posterior median of α and β, and the filled region defines the approximate 95% (2σ) pointwise confidence intervals on the regression line. The true values of the regression line are contained within the 95% confidence intervals.
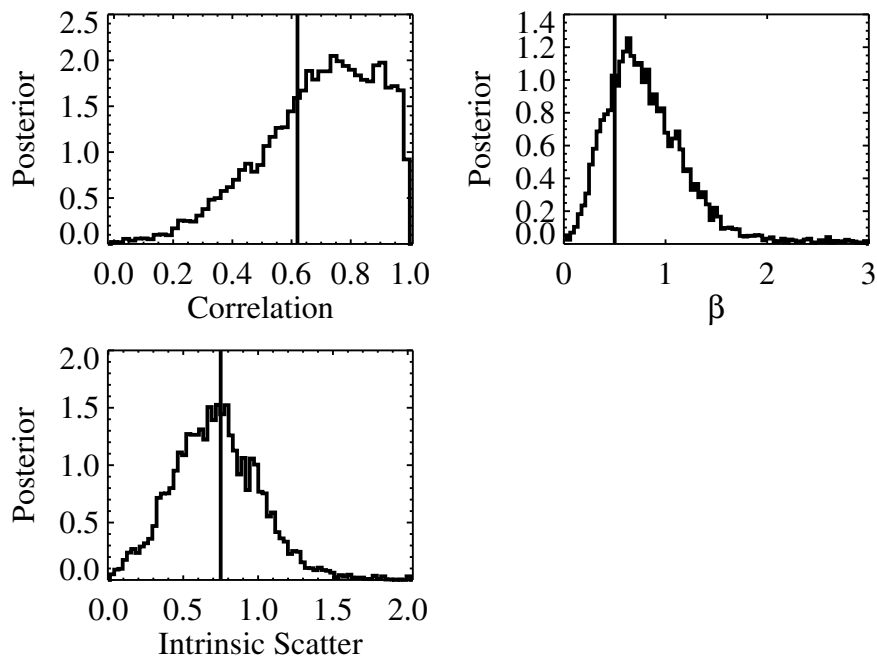
Figure 3.9 Same as Figure 3.6, but for the censored data set shown in Figure 3.8. The true values of the regression parameters are contained within the spread of the posteriors, implying that bounds on the regression parameters inferred from the posterior are trustworthy.

our considerable uncertainty in the slope, but they are sufficient for finding a positive correlation between the two variables, ξ and η. Therefore, despite the high level of censoring and measurement error in this data set, we would still be able to conclude that η increases as ξ increases.

3.9 APPLICATION TO REAL ASTRONOMICAL DATA: DEPENDENCE OF Γ_X ON L_{bol}/L_{Edd} FOR RADIO-QUIET QUASARS

To further illustrate the effectiveness of the method, I apply it to a data set drawn from my work on investigating the X-ray properties of radio-quiet quasars (RQQs). Recent work has suggested a correlation between quasar X-ray spectral slope, $\alpha_X, f_\nu \propto \nu^{-\alpha_X}$, and quasar Eddington ratio, L_{bol}/L_{Edd} (e.g., Porquet et al., 2004; Piconcelli et al., 2005; Shemmer et al., 2006). In this section I apply the regression method to a sample of 39 $z < 0.83$ RQQs and confirm the Γ_X–L_{bol}/L_{Edd} correlation. Because the purpose of this section is to illustrate the use of this regression method on real astronomical data, I defer a more in-depth analysis to a future paper.

Estimation of the Eddington luminosity, $L_{Edd} \propto M_{BH}$, requires an estimate of the black hole mass, M_{BH}. Black hole virial masses may be estimated as $M_{BH} \propto Rv^2$, where R is the broad line region size, and v is the velocity dispersion of the gas emitting the broad emission lines. A correlation has been found between the luminosity of a source and the size of it's broad line region (the R–L relationship, e.g., Kaspi et al., 2005). One can then exploit this relationship, and use the broad line $FWHM$ as an estimate for v, obtaining virial mass estimates $\hat{M}_{BH} \propto L^\theta v^2$ (e.g., Wandel et al., 1999), where the exponent is $\theta \approx 0.5$ (e.g., Vestergaard & Peterson, 2006). Unfortunately, the uncertainty on the broad line estimates of M_{BH} can be considerable, having a standard deviation of $\sigma_m \sim 0.4$

dex (e.g., McLure & Jarvis, 2002; Vestergaard & Peterson, 2006; Kelly & Bechtold, 2007). For ease of comparison with previous work, I estimate M_{BH} using only the Hβ emission line. The logarithm of the virial mass estimates were calculated using the Hβ luminosity and $FWHM$ according to the relationship given by Vestergaard & Peterson (2006).

My sample consists of a subset of the sample of Kelly et al. (2007). These sources have measurements of the X-ray photon index, $\Gamma_X = \alpha_X + 1$, obtained from *Chandra* observations, and measurements of the optical/UV luminosity at $2500\mathring{A}$, denoted as L_{2500}, obtained from SDSS spectra. The Hβ profile was modeled as a sum of Gaussians and extracted from the SDSS spectra according to the procedure described in Kelly & Bechtold (2007). I estimated the Hβ $FWHM$ and luminosity from the line profile fits.

I estimate the bolometric luminosity, L_{bol}, from the luminosity at $2500\mathring{A}$, assuming a constant bolometric correction $L_{bol} = 5.6L_{2500}$ (Elvis et al., 1994). The standard deviation in this bolometric correction reported by Elvis et al. (1994) is 3.1, implying an uncertainty in $\log L_{bol}$ of $\sigma_{bol} \sim 0.25$ dex. Combining this with the ~ 0.4 dex uncertainty on $\log M_{BH}$, the total 'measurement error' on $\log L_{bol}/L_{Edd}$ becomes $\sigma_x \sim 0.47$ dex. The distribution of Γ_X as a function of $\log L_{bol}/L_{Edd}$ is shown in Figure 3.10. As can be seen, the measurement errors on both Γ_X and $\log L_{bol}/L_{Edd}$ are large and make a considerable contribution to the observed scatter in both variables, where $R_y \sim 0.1$ and $R_x \sim 0.8$. Therefore, we expect the measurement errors to have a significant effect on the correlation and regression analysis.

I performed the regression assuming the linear form $\Gamma_X = \alpha + \beta \log L_{bol}/L_{Edd}$, and modellleling the intrinsic distribution of $\log L_{bol}/L_{edd}$ using $K = 2$ Gaussians. Draws from the posterior were obtained using the Gibbs sampler. The marginal
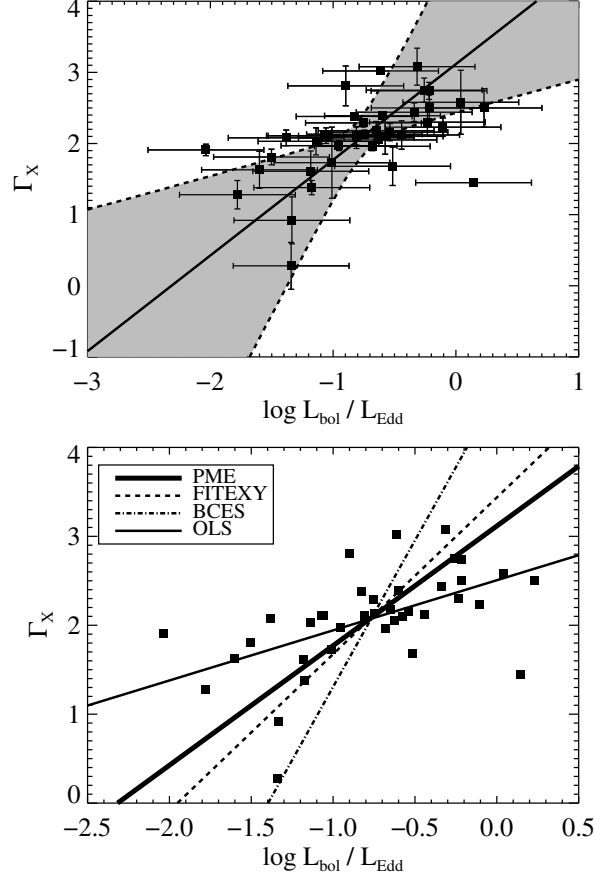
Figure 3.10 The X-ray photon index, Γ_X, as a function of $\log L_{bol}/L_{Edd}$ for 39 $z \lesssim$ 0.8 radio-quiet quasars. In both plots the thick solid line shows the posterior median estimate (PME) of the regression line. In the top plot, the filled region denotes the 95% (2σ) pointwise confidence intervals on the regression line. In the bottom plot, the thin solid line shows the OLS estimate, the dashed line shows the FITEXY estimate, and the dot-dashed line shows the BCES($Y|X$) estimate; the error bars have been omitted for clarity. A significant positive trend is implied by the data.
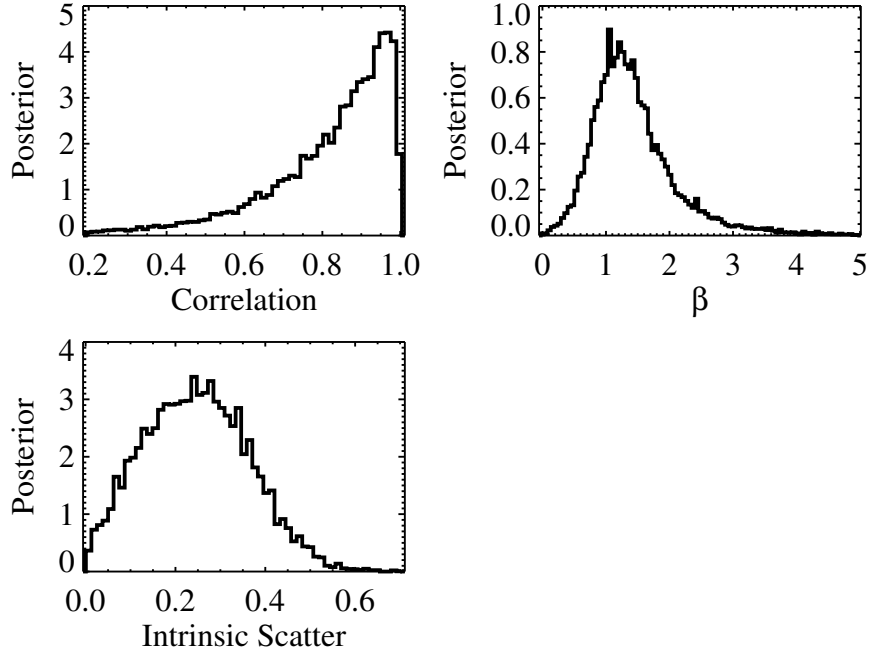
Figure 3.11 Same as Figure 3.6, but for the Γ_X–$\log L_{bol}/L_{Edd}$ regression. Although the uncertainty on the slope and correlation are large, the bounds on them implied by the data are $0 \lesssim \beta \lesssim 3.5$ and $0.2 \lesssim \rho \lesssim 1.0$.

posterior distributions for β, σ, and the correlation between Γ_X and $\log L_{bol}/L_{edd}$, ρ, are shown in Figure 3.11, and the posterior median and 95% (2σ) pointwise intervals on the regression line are shown in Figure 3.10. The posterior median estimate of the parameters are $\hat{\alpha} = 3.12 \pm 0.41$ for the constant, $\hat{\beta} = 1.35 \pm 0.54$ for the slope, $\hat{\sigma} = 0.26 \pm 0.11$ for the intrinsic scatter about the regression line, $\hat{\mu}_\xi = -0.77 \pm 0.10$ for the mean of $\log L_{bol}/L_{Edd}$, and $\hat{\sigma}_\xi = 0.32 \pm 0.12$ dex for the dispersion in $\log L_{bol}/L_{edd}$. Here, I have used a robust estimate of the posterior standard deviation as an 'error bar' on the parameters. These results imply that the observed scatter in $\log L_{bol}/L_{Edd}$ is dominated by measurement error, $\sigma_x/\tau \sim 1.5$, as expected from the large value of R_x.

For comparison, the BCES($Y|X$) estimate of the slope is $\hat{\beta}_{BCES} = 3.29 \pm 3.34$, the FITEXY estimate is $\hat{\beta}_{EXY} = 1.76 \pm 0.49$, and the OLS estimate is $\hat{\beta}_{OLS} = 0.56 \pm 0.14$; the standard error on $\hat{\beta}_{EXY}$ was estimated using bootstrapping. Figure 3.10 also compares the OLS, BCES, and FITEXY best-fit lines with the posterior median estimate. The 95% confidence region on the slope implied by the posterior draws is $0.46 < \beta < 3.44$, whereas the approximate 95% confidence region implied by the BCES, FITEXY, and OLS standard errors are $-3.26 < \beta < 9.84$, $0.80 < \beta < 2.72$, and $0.42 < \beta < 0.70$, respectively. The OLS and FITEXY estimates and the Bayesian approach give 'statistically significant' evidence for a correlation between $\log L_{bol}/L_{Edd}$ and Γ_X; however the BCES estimate is too variable to rule out the null hypothesis of no correlation. As noted before, the large measurement errors on $\log L_{bol}/L_{Edd}$ bias the OLS estimate of β toward shallower values and the FITEXY estimate of β toward steeper values. Because of this bias, confidence regions based on $\hat{\beta}_{OLS}$ and $\hat{\beta}_{EXY}$ are not valid because they are not centered on the true value of β, and thus do not contain the true value with the stated probability (e.g., 95%). On the other hand, confidence regions based on the BCES estimate are likely to be approximately valid; however, in this example the large measurement errors have caused $\hat{\beta}_{BCES}$ to be too variable to give meaningful constraints on the regression slope.

The BCES-type estimate of the intrinsic dispersion was $\hat{\sigma}_{BCES} = 0.32$ and the OLS estimate of the intrinsic dispersion was $\hat{\sigma}_{OLS} = 0.41$, where both were calculated in the same manner as in § 3.8.1. The FITEXY estimate of the intrinsic dispersion was $\hat{\sigma}_{EXY} = 0$, as $\chi^2_{EXY}/(n-2) < 1$. The BCES-type estimate of σ is similar to the Bayesian posterior median estimate, while $\hat{\sigma}_{OLS}$ overestimates the scatter compared to the Bayesian estimate by $\approx 58\%$. In contrast, the FITEXY estimator does not find any evidence for intrinsic scatter in the regression, which

is inconsistent with the posterior distribution of σ.

From the posterior distribution, we can constrain the correlation between Γ_X and $\log L_{bol}/L_{Edd}$ to be $0.328 \lesssim \rho \lesssim 0.998$ with $\approx 95\%$ probability, confirming the positive correlation between Γ_X and Eddington ratio seen previously. The posterior median estimate of the correlation is $\hat{\rho} = 0.87$, compared with an estimate of $\hat{r} = 0.54$ if one naively calculates the correlation directly from the measured data. The large measurement errors significantly attenuate the observed correlation, making the observed correlation between Γ_X and $\log L_{bol}/L_{edd}$ appear weaker than if one does not correct for the measurement errors.

3.10   CONCLUSIONS

In this work I have derived a likelihood function for handling measurement errors in linear regression of astronomical data. Our probability model assumes that the measurement errors are Gaussian with zero mean and known variance, that the intrinsic scatter in the dependent variable about the regression line is Gaussian, and that the intrinsic distribution of the independent variables can be well approximated as a mixture of Gaussians. I extend this model to enable the inclusion of non-detections, and describe how to incorporate the data selection process. A Gibbs sampler is described to enable simulating random draws from the posterior distribution.

I illustrated the effectiveness of structural Gaussian mixture models using simulation. For the specific simulations performed, a maximum-likelihood estimator based on the Gaussian structural model performed better than the OLS, BCES($Y|X$), and FITEXY estimators, especially when the measurement errors were large. In addition, our method also performed well when the measurement errors were large and the detection fraction was small, with the posterior distri-

butions giving reasonable bounds on the regression parameters. These results were in spite of the fact that the intrinsic distribution of the independent variable was not a sum of Gaussians for the simulations, suggesting that approximating the distribution of the independent variable as a mixture of Gaussians does not lead to a significant bias in the results. Finally, I concluded by using the method to fit the radio-quiet quasar X-ray photon index as a function of $\log L_{bol}/L_{Edd}$, using a sample of 39 $z < 0.83$ sources. The posterior distribution for this data set constrained the slope to be $0 \lesssim \beta \lesssim 3.5$ and the linear correlation coefficient to be $0.2 \lesssim \rho \lesssim 1.0$, confirming the correlation between X-ray spectral slope and Eddington ratio seen by other authors.

Although I have focused on linear regression in this work, the approach that I have taken is quite general and can be applied to other applications. In particular, Equations (3.11), (3.40), and (3.42) are derived under general conditions and are not limited to regression. In this work, I assume forms for the respective probability densities that are appropriate for linear regression; however, these equation provide a framework for constructing more general probability models of one's data, as in, for example, nonlinear fitting or estimation of luminosity functions.

CHAPTER 4

A FLEXIBLE METHOD OF ESTIMATING LUMINOSITY FUNCTIONS

4.1 CHAPTER ABSTRACT

We describe a Bayesian approach to estimating luminosity functions. We derive the likelihood function and posterior probability distribution for the luminosity function, given the observed data, and we compare the Bayesian approach with maximum-likelihood by simulating sources from a Schechter function. For our simulations confidence intervals derived from bootstrapping the maximum-likelihood estimate can be too narrow, while confidence intervals derived from the Bayesian approach are valid. We develop our statistical approach for a flexible model where the luminosity function is modelled as a mixture of Gaussian functions. Statistical inference is performed using markov chain monte carlo (MCMC) methods, and we describe a Metropolis-Hastings algorithm to perform the MCMC. The MCMC simulates random draws from the probability distribution of the luminosity function parameters, given the data, and we use a simulated data set to show how these random draws may be used to estimate the probability distribution for the luminosity function. In addition, we show how the MCMC output may be used to estimate the probability distribution of any quantities derived from the luminosity function, such as the peak in the space density of quasars. The Bayesian method we develop has the advantage that it is able to place accurate constraints on the luminosity function even beyond the survey detection limits, and that it provides a natural way of estimating the probability distribution of any quantities derived from the luminosity function, including those that rely on information beyond the survey detection limits.

4.2   CHAPTER INTRODUCTION

The luminosity function (LF) has been an important tool for understanding the evolution of galaxies and quasars, as it provides a census of the galaxy and quasar populations over cosmic time. Quasar luminosity functions have been estimated for optical surveys (e.g., Fan et al., 2001; Wolf et al., 2003; Croom et al., 2004; Richards et al., 2006; Jiang et al., 2006a), X-ray surveys (e.g., Steffen et al., 2003; Ueda et al., 2003; Barger et al., 2005; La Franca et al., 2005), infrared surveys (e.g., Barger et al., 2005; Matute et al., 2006; Babbedge et al., 2006), radio surveys (e.g., Waddington et al., 2001; Willott et al., 2001), and emission lines (Hao et al., 2005). In addition, luminosity functions across different bands have been combined to form an estimate of the bolometric luminosity function (Hopkins et al., 2007). Besides providing an important constraint on models of quasar evolution and supermassive black hole growth (e.g., Wyithe & Loeb, 2003; Hopkins et al., 2006a), studies of the LF have found evidence for 'cosmic downsizing', where the space density of more luminous quasars peaks at higher redshift. Attempts to map the growth of supermassive black holes start from the local supermassive black hole distribution, and employ the argument of Soltan (1982), using the quasar luminosity function as a constraint on the black hole mass distribution. These studies have found evidence that the highest mass black holes grow first (e.g., Yu & Tremaine, 2002; Marconi et al., 2004; Merloni, 2004), suggesting that this cosmic downsizing is the result of an anti-hierarchical growth of supermassive black holes.

Similarly, galaxy luminosity functions have been estimated in the optical (e.g., Blanton et al., 2003; Dahlen et al., 2005; Brown et al., 2007; Marchesini et al., 2007), X-ray (e.g., Kim et al., 2006; Ptak et al., 2007), infrared (e.g., Cirasuolo et al., 2007; Huynh et al., 2007), ultraviolet (e.g., Budavári et al., 2005; Paltani et al., 2007), ra-

dio (e.g., Lin & Mohr, 2007; Mauch & Sadler, 2007), for galaxies in clusters (e.g., Popesso et al., 2006; Harsono & de Propris, 2007), and for galaxies in voids (Hoyle et al., 2005). The galaxy luminosity function probes several aspects of the galaxy population; namely (a) the evolution of stellar populations and star formation histories (e.g., Faber et al., 2007), (b) the local supermassive black hole mass distribution (e.g, Yu & Tremaine, 2002; Marconi et al., 2004) via the Magorrian relationship (Magorrian et al., 1998), (c) the dependence of galaxy properties on evironment (e.g., Croton et al., 2005; Lauer et al., 2007), and (d) places constraints on models of structure formation and galaxy evolution (e.g., Bower et al., 2006; Finlator et al., 2006; Marchesini & van Dokkum, 2007).

Given the importance of the luminosity function as an observational constraint on models of quasar and galaxy evolution, it is essential that a statistically accurate approach be employed when estimating these quantities. However, the existence of complicated selection functions hinders this, and, as a result, a variety of methods have been used to accurately account for the selection function when estimating the LF. These include various binning methods (e.g., Schmidt, 1968; Avni & Bahcall, 1980; Page & Carrera, 2000), maximum-likelihood fitting (e.g., Marshall et al., 1983; Fan et al., 2001), and a powerful semi-parameteric approach (Schafer, 2007). In addition, there have been a variety of methods proposed for estimating the cumulative distribution function of the LF (e.g., Lynden-Bell, 1971; Efron & Petrosian, 1992; Maloney & Petrosian, 1999).

Each of these statistical methods has advantages and disadvantages. Statistical inference based on the binning procedures cannot be extended beyond the support of the selection function, and the cumulative distribution function methods typically assume that luminosity and redshift are statistically independent. Furthermore, one is faced with the arbitrary choice of bin size. The maximum-

likelihood approach typically assumes a restrictive and somewhat *ad hoc* parameteric form, and has not been used to give an estimate of the LF normalization; instead, for example, the LF normalization is often chosen to make the expected number of sources detected in one's survey equal to the actual number of sources detected. In addition, confidence intervals based on the errors derived from the various procedures are typically derived by assuming that the uncertainties on the LF parameters have a Gaussian distribution. While this is valid as the sample size approaches infinity, it is not necessarily a good approximation for finite sample sizes. This is particularly problematic if one is employing the best fit results to extrapolating the luminosity function beyond the bounds of the selection function. It is unclear if the probability distribution of the uncertainty in the estimated luminosity function below the flux limit is even asymptotically normal.

Motivated by these issues, we have developed a Bayesian method for estimating the luminosity function. We derive the likelihood function of the LF by relating the observed data to the true LF, assuming some parameteric form, and derive the posterior probability distribution of the LF parameters, given the observed data. While the likelihood function and posterior are valid for any parameteric form, we focus on a flexible parameteric model where the LF is modeled as a weighted sum of Gaussian functions. This is a type of 'non-parameteric' approach, where the basic idea is that the individual Gaussian functions do not have any physical meaning, but that given enough Gaussian functions one can obtain a suitably accurate approximation to the true LF; a similar approach has been taken by Blanton et al. (2003) for estimating galaxy LFs, and by Kelly et al. (2007) within the context of linear regression with measurement error. Modelling the LF as a mixture of Gaussian functions avoids the problem of choosing a particular parameteric form, especially in the absence of any guidance from

astrophysical theory. The mixture of Gaussians model has been studied from a Bayesian perspective by numerous authors (e.g., Roeder & Wasserman, 1997; Jasra et al., 2005; Dellaportas & Papageorgiou, 2006). In addition, we describe a markov chain monte carlo (MCMC) algorithm for obtaining random draws from the posterior distribution. These random draws allow one to estimate the posterior distribution for the LF, as well as any quantities derived from it. The MCMC method therefore allows a straight-forward method of calculating uncertainties on any quantity derived from the LF, such as the redshift where the space density of quasars or galaxies peaks; this has proven to be a challenge for other statistical methods developed for LF estimation. Because the Bayesian approach is valid for any sample size, one is therefore able to place reliable constraints on the LF and related quantities even below the survey flux limits.

Because of the diversity and mathematical complexity of some parts of this paper, we summarize the main results here. We do this so that the reader who is only interested in specific aspects of this paper can conveniently consult the sections of interest.

- In § 4.3 we derive the general form of the likelihood function for luminosity function estimation. We show that the commonly used likelihood function based on the Poisson distribution is incorrect, and that the correct form of the likelihood function is derived from the binomial distribution. However, because the Poisson distribution is the limit of the binomial distribution as the probability of including a source in a survey approaches zero, the maximum-likelihood estimates derived from the two distribution give nearly identical results so long as a survey's detection probability is small. The reader who is interested in using the correct form of the likelihood function of the LF should consult this section.

- In § 4.4 we describe a Bayesian approach to luminosity function estimation. We build on the likelihood function derived in § 4.3 to derive the probability distribution of the luminosity function, given the observed data (i.e., the posterior distribution). We use a simple example based on a Schechter function to illustrate the Bayesian approach, and compare it with the maximum-likelihood approach. For this example, we find that confidence intervals derived from the posterior distribution are valid, while confidence intervals derived from bootstrapping the maximum-likelihood estimate can be too small. The reader who is interested in a Bayesian approach to luminosity function estimation, and how it compares with maximum-likelihood, should consult this section.

- In § 4.5 we develop a mixture of Gaussian functions model for the luminosity function, deriving the likelihood function and posterior distribution for the model. Under this model, the LF is modelled as a weighted sum of Gaussian functions. This model has the advantage that given a suitably large enough number of Gaussian functions, it is flexible enough to give an accurate estimate of any smooth and continuous LF. This allows the model to adapt to the true LF, thus minimizing the bias that can result when assuming a parameteric form of the LF. This is particularly useful when extrapolating beyond the flux limits of a survey, where bias caused by parameteric misspecification can be a significant concern. The reader who are interested in employing the mixture of Gaussian functions model should consult this section.

- Because of the large number of parameters often associated with luminosity function estimation, Bayesian inference is most easily performed by obtain-

ing random draws of the LF from the posterior distribution. In § 4.6 we describe the Metropolis-Hastings algorithm (MHA) for obtaining random draws of the LF from the posterior distribution. As an example, we describe a MHA for obtaining random draws of the parameters for a Schechter function from the posterior distribution. Then, we describe a more complex MHA for obtaining random draws of the parameters for the mixture of Gaussian functions model. The reader who is interested in the computational aspects of 'fitting' the mixture of Gaussian functions model, or who is interested in the computational aspects of Bayesian inference for the LF, should consult this section. A computer routine for performing the Metropolis-Hastings alogrithm for the mixture of Gaussian functions model is available on request from B. Kelly.

- In § 4.7 we use simulation to illustrate the effectiveness of our Bayesian Gaussian mixture model for luminosity function estimation. We construct a simulated data set similar to the Sloan Digital Sky Survey DR3 Quasar Cataloge (Schneider et al., 2005). We then use our mixture of Gaussian functions model to recover the true LF and show that our mixture model is able to place reliable constraints on the LF. We also illustrate how to use the MHA output to constrain any quantity derived from the LF, and how to use the MHA output to assess the quality of the fit. The reader who is interested in assessing the effectiveness of our statistical approach, or who is interested in using the MHA output for statistical inference on the LF, should consult this section.

We adopt a cosmology based on the the WMAP best-fit parameters ($h = 0.71, \Omega_m = 0.27, \Omega_\Lambda = 0.73$, Spergel et al., 2003)

4.3   THE LIKELIHOOD FUNCTION

4.3.1   Notation

We use the common statistical notation that an estimate of a quantity is denoted by placing a 'hat' above it; e.g., $\hat{\theta}$ is an estimate of the true value of the parameter θ. The parameter θ may be scalar or multivalued. We denote a normal density[1] (i.e., a Gaussian distribution) with mean μ and variance σ^2 as $N(\mu, \sigma^2)$, and we denote as $N_p(\mu, \Sigma)$ a multivariate normal density with p-element mean vector μ and $p \times p$ covariance matrix Σ. If we want to explicitly identify the argument of the Gaussian function, we use the notation $N(x|\mu, \sigma^2)$, which should be understood to be a Gaussian with mean μ and variance σ^2 as a function of x. We will often use the common statistical notation where "$\sim$" means "is drawn from" or "is distributed as". This should not be confused with the common usage of implying "similar to". For example, $x \sim N(\mu, \sigma^2)$ states that x is drawn from a normal density with mean μ and variance σ^2, whereas $x \sim 1$ states that the value of x is similar to one.

In this work, the maximum-likelihood estimate of the luminosity function refers to an estimate of the LF obtained by maximizing the likelihood function of the unbinned data. Therefore, the maximum-likelihood estimate does not refer to an estimate obtained by maximizing the likelihood function of binned data, such as fitting the results obtained from the $1/V_a$ technique.

4.3.2   Derivation of the Luminosity Function Likelihood

The luminosity function, denoted as $\phi(L, z)dL$, is the number of sources per co-moving volume $V(z)$ with luminosities in the range $L, L + dL$. The luminosity

---

[1]We use the terms probability density and probability distribution interchangeably.

function is related to the probability density of (L, z) by

$$p(L, z) = \frac{1}{N}\phi(L, z)\frac{dV}{dz},$$ (4.1)

where N is the total number of sources in the observable universe, and is given by the integral of ϕ over L and $V(z)$. Note that $p(L, z)dLdz$ is the probability of finding a source in the range $L, L+dL$ and $z, z+dz$. Equation (4.1) seperates the LF into its shape, given by $p(L, z)$, and its normalization, given by N. Once we have an estimate of $p(L, z)$, we can easily convert this to an estimate of $\phi(L, z)$ using Equation (4.1). In general, it is easier to work with the probability distribution of L and z, instead of directly with the LF, because $p(L, z)$ is more directly related to the likelihood function.

If we assume a parameteric form for $\phi(L, z)$, with parameters θ, we can derive the likelihood function for the observed data. The likelihood function is the probability of observing one's data, given the assumed model. The presense of flux limits and various other selection effects can make this difficult, as the observed data likelihood function is not simply given by Equation (4.1). In this case, the set of luminosities and redshifts observed by a survey gives a biased estimate of the true underlying distribution, since only those sources with L above the flux limit at a given z are detected. In order to derive the observed data likelihood function, it is necessary to take the survey's selection method into account. This is done by first deriving the joint likelihood function of both the observed and unobserved data, and then integrating out the unobserved data.

Because the data points are independent, the likelihood function for all N sources in the universe is

$$p(L, z|\theta) = \prod_{i=1}^{N} p(L_i, z_i|\theta).$$ (4.2)

In reality, we do not know the luminosities and redshifts for all N sources, nor

do we know the value of N, as our survey only covers a fraction of the sky and is subject to a selection function. As a result, our survey only contains n sources. Because of this, the selection process must also be included in the probability model, and the total number of sources, N, is an additional parameter that needs to be estimated.

We can incorporate the sample selection into the likelihood function by including the random detection of sources. We introduce an N-element indicator vector $\mathbf{I}$ that takes on the values $I_i = 1$ if the ith source is included in our survey and $I_i = 0$ otherwise. Note that $\mathbf{I}$ is a vector of size N containing only ones and zeros. In this case, the selection function is the probability of including a source given L and z, $p(I_i = 1|L_i, z_i)$. The complete data likelihood is then the probability that all objects of interest in the universe (e.g., all quasars) have luminosities $L_1, \ldots, L_N$ and redshifts $z_1, \ldots, z_N$, and that the selection vector I has the values $I_1, \ldots, I_N$, given our assumed luminosity function:

$$p(L, z, \mathbf{I}|\theta, N) = C_n^N \prod_{i \in \mathcal{A}_{obs}} p(I_i = 1|L_i, z_i)p(L_i, z_i|\theta) \prod_{j \in \mathcal{A}_{mis}} p(I_j = 0|L_j, z_j)p(L_j, z_j|\theta).$$

(4.3)

Here, $C_n^N = N!/n!(N-n)!$ is the binomial coefficient, $\mathcal{A}_{obs}$ denotes the set of n included sources, and $\mathcal{A}_{mis}$ denotes the set of $N-n$ missing sources. The number of sources detected in a survey is random, and therefore the binomial coefficient is necessary in normalizing the likelihood function, as it gives the number of possible ways to select a subset of n sources from a set of N total sources.

Because we are interested in the probability of the observed data, given our assumed model, the complete data likelihood function is of little use by itself. However, we can integrate Equation (4.3) over the missing data to obtain the observed data likelihood function. This is because the marginal probability distribution of the observed data is obtained by integrating the joint probability distribution of

the observed and the missing data over the missing data:

$$p(L_{obs}, z_{obs}, \mathbf{I}|\theta, N) = C_n^N \prod_{i \in \mathcal{A}_{obs}} p(I_i = 1|L_i, z_i)p(L_i, z_i|\theta) \tag{4.4}$$

$$\times \prod_{j \in \mathcal{A}_{mis}} \int_0^\infty \int_0^\infty p(I_j = 0|L_j, z_j)p(L_j, z_j|\theta) \; dL_j \; dz_j \tag{4.5}$$

$$\propto C_n^N \left[p(I = 0|\theta)\right]^{N-n} \prod_{i \in \mathcal{A}_{obs}} p(L_i, z_i|\theta), \tag{4.6}$$

where the probability that the survey misses a source, given the parameters θ, is

$$p(I = 0|\theta) = \int \int p(I = 0|L, z)p(L, z|\theta) \; dL \; dz. \tag{4.7}$$

Here, we have introduced the notation that L_{obs} and z_{obs} denote the set of values of L and z for those sources included in one's survey, and we have omitted terms that do not depend on θ or N from Equation (4.6). Equation (4.6) is the observed data likelihood function, given an assumed luminosity function (Eq.[4.1]). Qualitatively, the observed data likelihood function is the probability of observing the set of n luminosities $L_1, \ldots, L_n$ and redshifts $z_1, \ldots, z_n$ given the assumed luminosity function parameterized by θ, multiplied by the probability of not detecting $N - n$ sources given θ, multiplied by the number of ways of selecting a subset of n sources from a set of N total sources. The observed data likelihood function can be used to calculate a maximum likelihood estimate of the luminosity function, or combined with a prior distribution to perform Bayesian inference.

4.3.3 Comparison with the Poisson Likelihood

The observed data likelihood given by Equation (4.6) differs from that commonly used in the luminosity function literature. Instead, a likelihood based on the Poisson distribution is often used. Marshall et al. (1983) give the following equation for the log-likelihood function based on the Poisson distribution:

$$\log p(L_{obs}, z_{obs}|\theta, N) = \sum_{i \in \mathcal{A}_{obs}} \log \phi(L_i, z_i|N, \theta) + \int \int p(I = 1|L, z)\phi(L, z|\theta, N)\frac{dV}{dz} \; dL \; dz. \tag{4.8}$$

Inserting Equation (4.1) for $\phi(L, z|\theta)$, the log-likelihood based on the poisson like-lihood becomes

$$\log p(L_{obs}, z_{obs}|\theta, N) = n \log N + \sum_{i \in \mathcal{A}_{obs}} \log p(L_i, z_i|\theta) - N p(I = 1|\theta), \qquad (4.9)$$

where, $p(I = 1|\theta) = 1 - p(I = 0|\theta)$, and $p(I = 0|\theta)$ is given by Equation (4.7). In contrast, the log-likelihood we have derived based on the binomial distribution is the logarithm of Equation (4.6):

$$\log p(L_{obs}, z_{obs}|\theta, N) = \log N! - \log n! - \log(N-n)! + \sum_{i \in \mathcal{A}_{obs}} \log p(L_i, z_i|\theta) + (N-n) \log p(I = 0|\theta).$$

$$(4.10)$$

The likelihood functions implied by Equations (4.9) and (4.10) are functions of N, and thus the likelihoods may also be maximized with respect to the LF normalization. This is contrary to what is often claimed in the literature, where the LF normalization is typically chosen to make the expected number of sources observed in one's survey equal to the actual number observed.

The binomial likelihood, given by Equation (4.6), contains the term C_n^N, resulting from the fact that the total number of sources included in a survey, n, follows a binomial distribution. For example, suppose one performed a survey over one quarter of the sky with no flux limit. Assuming that sources are uniformly distributed on the sky, the probability of including a source for this survey is simply $1/4$. If there are N total sources in the universe, the total number of sources that one would find within the survey area follows a binomial distribution with N 'trials' and probability of 'success' $p = 1/4$. However, the Poisson likelihood is derived by noting that the number of sources detected in some small bin in (L, z) follows a poisson distribution. Since the sum of a set of poisson distributed random variables also follows a poisson distribution, this implies that the total number of sources detected in one's survey, n, follows a poisson distribution.

However, n actually follows a binomial distribution, and thus the observed data likelihood function is not given by the poisson distribution. The source of this error is largely the result of approximating the number of sources in a bin as following a Poisson distribution, when in reality it follows a binomial distribution.

Although the poisson likelihood function for the LF is incorrect, the previous discussion should not be taken as a claim that previous work based on the poisson likelihood function is incorrect. When the number of sources included in one's sample is much smaller than the total number of sources in the universe, the binomial distribution is well approximated by the poisson distribution. Therefore, if the survey only covers a small fraction of the sky, or if the flux limit is shallow enough such that $n \ll N$, then the poisson likelihood function should provide an accurate approximation to the true binomial likelihood function. When this is true, statistical inference based on the poisson likelihood should only exhibit negligible error, so long as there are enough sources in one's survey to obtain an accurate estimate of the LF normalization. In § 4.4.3 we use simulate to compare results obtained from the two likelihood functions, and to compare the maximum-likelihood approach to the Bayesian approach.

4.4 POSTERIOR DISTRIBUTION FOR THE LF PARAMETERS

We can combine the likelihood function for the LF with a prior probability distribution on the LF parameters to perform Bayesian inference on the LF. The result is the posterior probability distribution of the LF parameters, i.e., the probability distribution of the LF parameters given our observed data. This is in contrast to the maximum likelihood approach, where the maximum likelihood approach seeks to relate the observed value of the MLE to the true parameter value through an estimate of the sampling distribution of the MLE. In Appendix A we give a

more thorough introduction to the difference between the maximum likelihood and Bayesian approaches.

4.4.1 Derivation of the Posterior Probability Distribution

The posterior probability distribution of the model parameters is related to the likelihood function and the prior probability distribution as

$$p(\theta, N | L_{obs}, z_{obs}, \mathbf{I}) \propto p(\theta, N) p(L_{obs}, z_{obs}, \mathbf{I} | \theta, N), \tag{4.11}$$

where $p(\theta, N)$ is the prior on (θ, N), and $p(L_{obs}, z_{obs}, \mathbf{I} | \theta, N)$ is is the observed data likelihood function, given by Equation (4.6). The posterior distribution is the probability distribution of θ and N, given the observed data, L_{obs} and z_{obs}. Because the luminosity function depends on the parameters θ and N, the posterior distribution of θ and N can be used to obtain the probability distribution of $\phi(L, z)$, given our observed set of luminosities and redshifts.

It is of use to decompose the posterior as $p(\theta, N | L_{obs}, z_{obs}) \propto p(N | \theta, L_{obs}, z_{obs}) p(\theta | L_{obs}, z_{obs})$; here we have dropped the explicit conditioning on $\mathbf{I}$. This decomposition seperates the posterior into the conditional posterior of the LF normalization at a given θ, $p(N | L_{obs}, z_{obs}, \theta)$, from the marginal posterior of the LF shape, $p(\theta | L_{obs}, z_{obs})$. In this work we assume that N and θ are independent in their prior distribution, $p(\theta, N) = p(N) p(\theta)$, and that the prior on N is uniform over $\log N$. A uniform prior on $\log N$ corresponds to a prior distribution on N of $p(N) \propto 1/N$, as $p(\log N) d \log N = p(N) dN$. Under this prior, one can show that the marginal posterior probability distribution of θ is

$$p(\theta | L_{obs}, z_{obs}) \propto p(\theta) \left[p(I = 1 | \theta) \right]^{-n} \prod_{i \in \mathcal{A}_{obs}} p(L_i, z_i | \theta), \tag{4.12}$$

where $p(I = 1 | \theta) = 1 - p(I = 0 | \theta)$. We derive Equation (4.12) in Appendix B (see also Gelman et al., 2004). Under the assumption of a uniform prior on θ, Equa-

tion (4.12) is equivalent to Equation (22) in Fan et al. (2001), who use a different derivation to arrive at a similar result.

Under the prior $p(\log N) \propto 1$, the conditional posterior distribution of N at a given θ is a negative binomial distribution with parameters n and $p(I = 1|\theta)$. The negative binomial distribution gives the probability that the total number of sources in the universe is equal to N, given that we have observed n sources in our sample with probability of inclusion $p(I = 1|\theta)$:

$$p(N|n, \theta) = C_{n-1}^{N-1} \left[p(I = 1|\theta)\right]^n \left[p(I = 0|\theta)\right]^{N-n}. \tag{4.13}$$

Here, $p(I = 0|\theta)$ is given by Equation (4.7) and $p(I = 1|\theta) = 1 - p(I = 0|\theta)$. Further description of the negative binomial distribution is given in § C. The complete joint posterior distribution of θ and N is then the product of Equations (4.12) and (4.13), $p(\theta, N|L_{obs}, z_{obs}) \propto p(N|\theta, n)p(\theta|L_{obs}, z_{obs})$.

Because it is common to fit a luminosity function with a large number of parameters, it is computationally intractable to directly calculate the posterior distribution from Equations (4.12) and (4.13). In particular, the number of grid points needed to calculate the posterior will scale exponentially with the number of parameters. Similarly, the number of integrals needed to calculate the marginal posterior probability distribution of a single parameters will also increase exponentially with the number of parameters. Instead, Bayesian inference is most easily performed by simulating random draws of N and θ from their posterior probability distribution. Based on the decomposition $p(\theta, N|L_{obs}, z_{obs}) \propto p(N|n, \theta)p(\theta|L_{obs}, z_{obs})$, we can obtain random draws of (θ, N) from the posterior by first drawing values of θ from Equation (4.12). Then, for each draw of θ, we draw a value of N from the negative binomial distribution. The values of N and θ can then be used to compute the values of luminosity function via Equation (4.1). The values of the LF computed from the random draws of N and θ are then

treated as a random draw from the probability distribution of the LF, given the observed data. These random draws can be used to estimate posterior means and variances, confidence intervals, and histogram estimates of the marginal distributions. Random draws for θ may be obtained via markov chain monte carlo (MCMC) methods, described in § 4.6, and we describe in § C how to obtain random draws from the negative binomial distribution. In § 4.7.2 we give more details on using random draws from the posterior to perform statistical inference on the LF.

4.4.2 Illustration of the Bayesian Approach: Schechter Function

Before moving to more advanced models, we illustrate the Bayesian approach by applying it to a simulated set of luminosities drawn from a Schechter function. We do this to give an example of how to calculate the posterior distribution, how to obtain random draws from the posterior and use these random draws to draw scientific conclusions based on the data, and to compare the Bayesian approach with the maximum-likelihood approach (see § 4.4.3). The Schechter luminosity function is:

$$\phi(L) = \frac{N}{L^*\Gamma(\alpha+1)} \left(\frac{L}{L^*}\right)^\alpha e^{-L/L^*}, \quad \theta = (\alpha, L^*). \tag{4.14}$$

For simplicity, we ignore a z dependence. The Schechter function is equivalent to a Gamma distribution with shape parameter $k = \alpha + 1$, and scale parameter L^*. Note that $k > 0$ and $\alpha > -1$; otherwise the integral of Equation (4.14) may be negative or become infinite. For our simulation, we randomly draw $N = 1000$ galaxy luminosities from Equation (4.14) using a value of $\alpha = 0$ and $L^* = 10^{44}$ erg s^{-1}.

To illustrate how the results depend on the detection limit, we placed two different detection limits on our simulated survey. The first limit was at $L_{min} = 2 \times 10^{43}$ ergs s^{-1}, and the second was at $L_{min} = 2 \times 10^{44}$ ergs s^{-1}. We used a hard

detection limit, where all sources above L_{min} were detected and all sources below L_{min} were not: $p(I = 1|L > L_{min}) = 1$ and $p(I = 1|L < L_{min}) = 0$. Note that the first detection limit lies below L^*, while the second detection limit lies above L^*. We were able to detect $n \sim 818$ sources for $L_{min} = 2 \times 10^{43}$ ergs s^{-1} and $n \sim 135$ sources for $L_{min} = 2 \times 10^{44}$ ergs s^{-1}

The marginal posterior distribution of α and L^* can be calculated by inserting into Equation (4.12) an assumed prior probability distribution, $p(\alpha, L^*)$, and the likelihood function, $p(L_i|\alpha, L^*)$. Because we are ignoring redshift in our example, the likelihood function is simply $p(L_i|\alpha, L^*) = \phi(L)/N$. In this example, we assume a uniform prior on $\log L^*$ and α, and therefore $p(L^*, \alpha) \propto 1/L^*$. From Equations (4.12) and (4.14), the marginal posterior distribution of the parameters is

$$p(\alpha, L^*|L_{obs}) \propto \frac{1}{L^*} \left[p(I = 1|\alpha, L^*)\right]^{-n} \prod_{i=1}^{n} \frac{1}{L^*\Gamma(\alpha + 1)} \left(\frac{L_i}{L^*}\right)^{\alpha} e^{-L_i/L^*}, \qquad (4.15)$$

where the survey detection probability is

$$p(I = 1|\alpha, L^*) = \int_{L_{min}}^{\infty} \frac{1}{L^*\Gamma(\alpha + 1)} \left(\frac{L_i}{L^*}\right)^{\alpha} e^{-L_i/L^*} \, dL. \qquad (4.16)$$

The conditional posterior distribution of N at a given θ is given by inserting in Equation (4.16) into Equation (4.13), and the joint posterior of α, L^*, and N is obtained by multiplying Equation (4.15) by Equation (4.13).

We perform statistical inference on the LF by obtaining random draws from the posterior distribution. In order to calculate the marginal posterior distributions, $p(\alpha|L_{obs}), p(L^*|L_{obs})$, and $p(N|L_{obs})$, we would need to numerically integrate the posterior distribution over the other two parameters. For example, in order to calculate the marginal posterior of α, $p(\alpha|L_{obs}, z_{obs})$, we would need to integrate $p(\alpha, L^*, N|L_{obs})$ over L^* and N on a grid of values for α. While feasible for the simple 3-dimensional problem illustrated here, it is faster to simply obtain

a random draw of α, L^*, and N from the posterior, and then use a histogram to estimate $p(\alpha|L_{obs})$. Further details are given in § 4.7.2 on performing Bayesian inference using random draws from the posterior.

We used the Metropolis-Hastings algorithm described in § 4.6.1 to obtain a random draw of α, L^*, and N from the posterior probability distribution. The result was a set of 10^5 random draws from the posterior probability distribution of α, L^*, and N. In Figure 4.1 we show the estimated posterior distribution of α, L^*, and N for both detection limits. While L^* is fairly well constrained for both detection limits, the uncertainties on α and N are highly sensitive to whether the detection limit lies above or below L^*. In addition, the uncertainties on these parameters are not Gaussian, as is often assumed for the MLE.

The random draws of α, L^*, and N can also be used to place constraints on the LF. This is done by computing Equation (4.14) for each of the random draws of α, L^*, and N, and plotting the regions that contain, say, 90% of the probability. In Figure 4.2 we show the posterior median estimate of the LF, as well as the region containing 90% of the posterior probability. As can be seen, the 90% bounds contain the true value of the LF, and increase or decrease to reflect the amount of data available as a function of L. Furthermore, unlike the traditional MLE, these bounds do not rely on an assumption of Gaussian uncertainties, and therefore the confidence regions are valid for any sample size.

4.4.3   Comparison with Maximum-likelihood: Schechter Function

We also use monte carlo simulation to compare the Bayesian approach to maximum-likelihood for both the binomial and Poisson likelihood functions. We simulated 20 data sets for four types of surveys: (1) A large area shallow survey, (2) a large area medium depth survey, (3) a small area deep survey, and (4) a large area deep survey for rare objects, such as $z \sim 6$ quasars (e.g., Fan et al., 2006). For all four
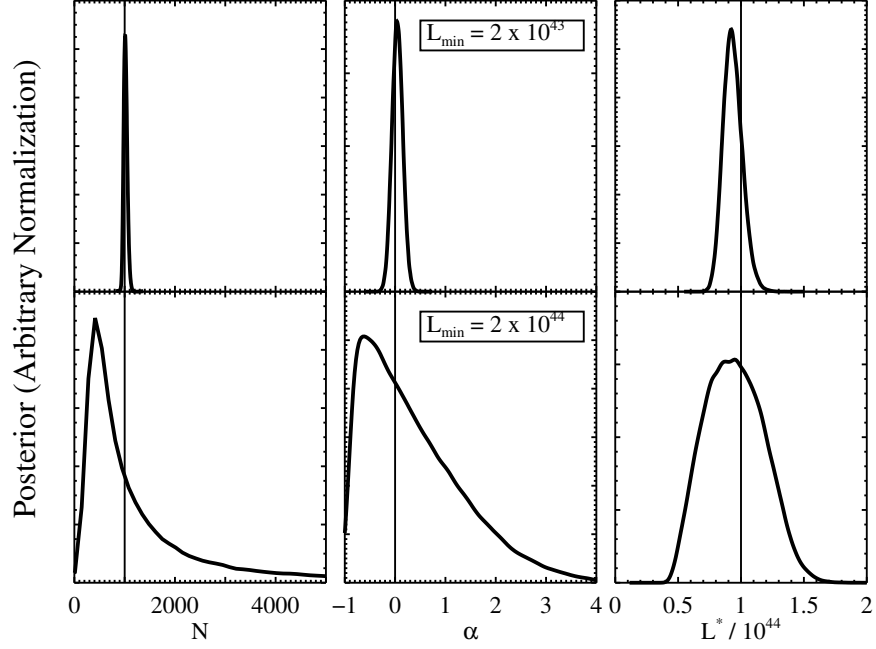
Figure 4.1 Posterior probability distribution of the Schechter luminosity function parameters, $N, \alpha,$ and L^*, for the simulated sample described in § 4.4.2. The top three panels show the posterior when the luminosity limit of the survey is $L > 2 \times 10^{43}$ [erg s^{-1}], and the bottom three panels show the posterior distribution when the luminosity limit of the survey is $L > 2 \times 10^{44}$ [erg s^{-1}]. The vertical lines mark the true values of the parameters, $N = 1000, \alpha = 0,$ and $L^* = 10^{44}$ [erg s^{-1}]. The uncertainty on the parameters increases considerably when $L_{min} > L^*$, reflecting the fact that the bright end of the Schechter LF contains little information on α or N.
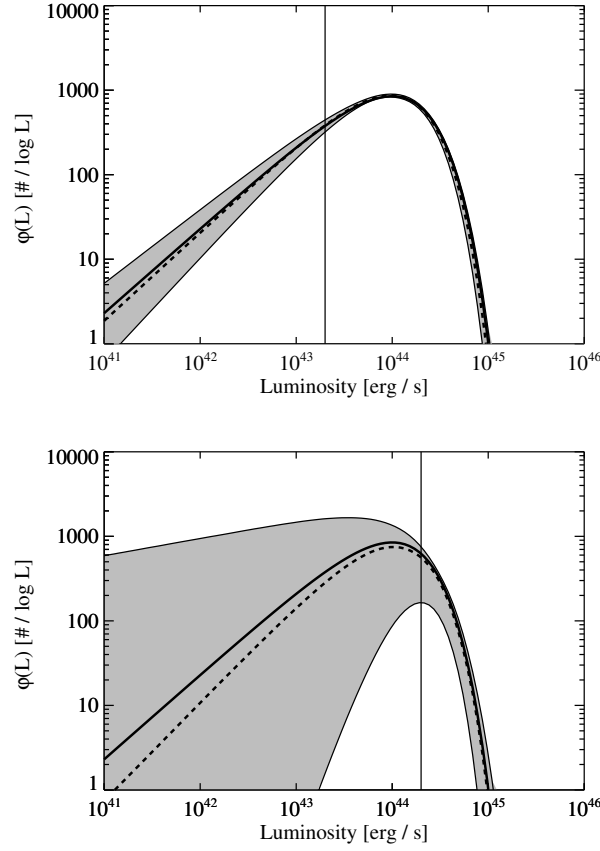
Figure 4.2 True value of the Schechter luminosity function (dashed line), compared with the best fit luminosity function calculated from the posterior median of the Schechter function parameters, N, α, and L^* (solid line), from the simulated sample described in § 4.4.2. The top panel summarizes the posterior probability distributin of the LF when the luminosity limit is $L > 2 \times 10^{43}$ [erg s^{-1}], and the bottom panel summarizes the posterior distribution of the LF when the luminosity limit is $L > 2 \times 10^{44}$ [erg s^{-1}]. In both panels the shaded region contains 90% of the posterior probability, and the vertical line marks the lower luminosity limit of the simulated survey. The uncertainty on the LF below the luminosity limit increases considerably when $L_{min} > L^*$, reflecting the fact that the bright end of the Schechter LF contains little information on α or N, and therefore contains little information on the faint end of the LF.

survey types we simulated quasars from a Schechter luminosity function with parameters the same as in § 4.4.2. For the large area shallow survey we used a total number of sources of $N = 10^5$, an area of $\Omega = 10^4 \deg^2$, and a lower luminosity limit of $L_{min} = 5 \times 10^{44} \text{ erg s}^{-1}$. Only $n \sim 160$ sources are expected to be detected by this survey. For the large area medium depth survey we also used a LF normalization of $N = 10^5$ and area of $\Omega = 10^4 \deg^2$, but instead used a lower luminosity limit of $L_{min} = 5 \times 10^{43} \text{ erg s}^{-1}$. The large area medium depth survey is expected to detect $n \sim 1.5 \times 10^4$ sources. For the small area deep survey we used a survey area of $\Omega = 448 \text{ arcmin}^2$, a LF normalization of $N = 5 \times 10^7$ sources, and a lower luminosity limit of $L_{min} = 10^{43} \text{ erg s}^{-1}$. This survey is expected to detect $n \sim 140$ sources. Finally, for the large area deep rare object survey we used an area of $\Omega = 10^4 \deg^2$, a LF normalization of $N = 75$ sources, and a lower luminosity limit of $L_{min} = 10^{43} \text{ erg s}^{-1}$. Only $n \sim 16$ sources are expected to be detected by the rare object survey.

We fit each of the 20 simulated data sets by maximum-likelihood for both the binomial and Poisson likelihood functions. The 95% confidence intervals on the best-fit parameters were determined using 2000 bootstrap samples. We use the bootstrap to estimate the confidence intervals because the bootstrap does not assume that the errors are Gaussian, and because it is a common technique used in the LF literature. We estimate the 95% confidence intervals directly from the 0.025 and 0.975 percentiles of the bootstrap sample. While bootstrap confidence intervals derived in this manner are known to be biased (e.g., Efron, 1987; Davison & Hinkley, 1997), additional corrections to the bootstrap samples are complicated. In addition, it is common practice to estimate bootstrap confidence intervals in this manner, and it is worth testing their accuracy. For the Bayesian approach, we used the MHA algorithm described in § 4.6.1 to simulate 5×10^4 random draws

from the posterior distribution. The MHA algorithm was faster than fitting the 2000 bootstrap samples using maximum likelihood.

For each of the simulated samples, we counted the number of times that the true values of $N, \alpha,$ and L^* were contained within the estimated 95% confidence interval. Because we estimated values of three parameters for 20 simulated data sets of 4 different types of surveys, we had 240 'trials' with probability of 'success' $p = 0.95$. If the estimated 95% confidence regions corresponded to the true regions, then with $\approx 99\%$ probability between 220 and 236 of the 'trials' would fall within the estimated confidence region. For the binomial likelihood, the true value of a parameter was within the estimated 95% confidence region only 210 times (88%), and for the Poisson likelihood the true value of a parameter was within the estimated 95% confidence region only 202 times (84%). In contrast, the Bayesian approach was able to correctly constrain the true value of a parameter to be within the 95% confidence region 233 times (97%). Therefore, for our simulations confidence regions derived from bootstrapping the maximum-likelihood estimate are too narrow, while the confidence regions derived from the Bayesian method are correct.

Most of the failure in the maximum-likelihood confidence intervals came from the difficulty of the maximum-likelihood approach in constraining the LF normalization, N, for the small area deep survey and for the rare object survey. In particular, for these two surveys the bootstrap 95% confidence intervals for both the binomial and Poisson likelihood function only contained the true value of N $\sim 50\%$ of the time. In addition, the Poisson and binomial likelihood functions gave slightly different results for the larger area medium depth survey. For this survey the 95% confidence intervals for the maximum-likelihood estimate derived from the Poisson distribution were somewhat smaller than those for the

binomial distribution, only correcting including the true values of $N, k,$ and L^* $\sim 85\%$ of the time. This is expected, because the Poisson distribution is the limit of the binomial distribution as the probability of including a source approaches zero; however, the detection probability for the large area medium depth survey is ≈ 0.15.

4.5   MIXTURE OF GAUSSIAN FUNCTIONS MODEL FOR THE LUMINOSITY FUNCTION

In this section we describe a mixture of Gaussian functions model for the luminosity function. The mixture of Gaussians model is a common 'non-parameteric' model that allows flexibility when estimating a distribution, and is often employed when there is uncertainty regarding the specific functional form of the distribution of interest. The basic idea is that one can use a suitably large enough number of Gaussian functions to accurately approximate the true LF, even though the individual Gaussians have no physical meaning. As a result, we avoid the assumption of a more restrictive parameteric form, such as a power-law, which can introduce considerable bias when extrapolating beyond the bounds of the observable data.

In this work we assume the mixture of Gaussian functions for the joint distribution of $\log L$ and $\log z$, as the logarithm of a strictly positive variable tends to more closely follow a normal distribution than does the untransformed variable. Therefore, we expect that a fewer number of Gaussians will be needed to accurately approximate the true LF, thus reducing the number of free parameters. Assuming a mixture of Gaussian functions for the joint distribution of $\log L$ and $\log z$ is equivalent to assuming a mixture of log-normal distributions for the distribution of L and z. The mixture of K Gaussian functions model for the ith data

point is

$$p(\log L_i, \log z_i | \pi, \mu, \Sigma) = \sum_{k=1}^{K} \frac{\pi_k}{2\pi |\Sigma_k|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}_i - \mu_k)^T \Sigma_k^{-1}(\mathbf{x}_i - \mu_k)\right], \ \theta = (\pi, \mu, \Sigma),$$

(4.17)

where $\sum_{k=1}^{K} \pi_k = 1$. Here, $\mathbf{x}_i = (\log L_i, \log z_i)$, μ_k is the 2-element mean (i.e., position) vector for the kth Gaussian, Σ_k is the 2×2 covariance matrix for the kth Gaussian, and $\mathbf{x}^T$ denotes the transpose of $\mathbf{x}$. In addition, we denote $\pi = (\pi_1, \ldots, \pi_K)$, $\mu = (\mu_1, \ldots, \mu_K)$, and $\Sigma = (\Sigma_1, \ldots, \Sigma_K)$. The variance in $\log L$ for Gaussian k is $\sigma_{l,k}^2 = \Sigma_{11,k}$, the variance in $\log z$ for Gaussian k is $\sigma_{z,k}^2 = \Sigma_{22,k}$, and the covariance between $\log L$ and $\log z$ for Gaussian k is $\sigma_{lz,k} = \Sigma_{12,k}$.

Under the mixture model, the LF can be calculated from Equations (4.1) and (4.17). Noting that $p(L, z) = p(\log L, \log z)/(Lz(\ln 10)^2)$, the mixture of Gaussian functions model for the LF is

$$\phi(L, z | \theta, N) = \frac{N}{Lz(\ln 10)^2} \left(\frac{dV}{dz}\right)^{-1} \sum_{k=1}^{K} \frac{\pi_k}{2\pi |\Sigma_k|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mu_k)^T \Sigma_k^{-1}(\mathbf{x} - \mu_k)\right],$$

(4.18)

where, as before, $\mathbf{x} = (\log L, \log z)$. A mixture of Gaussian functions models was also used by Blanton et al. (2003) to estimate the $z = 0.1$ galaxy LF from the Sloan Digital Sky Survey (SDSS). Our mixture of Gaussian functions model differs from that used by Blanton et al. (2003) in that we do not fix the gaussian function centroids to lie on a grid of values, and their individual widths are allowed to vary. This flexibility enables us to use a smaller number of Gaussian functions (typically $\sim 3 - 6$) to accurately fit the LF.

4.5.1 Prior Distribution

In this section we describe the prior distribution that we adopt on the mixture of Gaussian functions parameters. While one may be tempted to assumed a uniform prior on π, μ, and Σ, this will lead to an improper posterior, i.e., the poste-

rior probability density does not integrate to one (Roeder & Wasserman, 1997). Therefore, a uniform prior cannot be used, and we need to develop a more informative prior distribution. Following Roeder & Wasserman (1997), we assume a uniform prior on $\pi_1, \ldots, \pi_K$ under the constraint that $\sum_{k=1}^{K} \pi_k = 1$; formally, this is a $\mathrm{Dirichlet}(1, \ldots, 1)$ prior, where $\mathrm{Dirichlet}(\alpha_1, \ldots, \alpha_K)$ denotes a Dirichlet density with parameters $\alpha_1, \ldots, \alpha_K$. We give further details on the Dirichlet probability distribution in Appendix C.

Although our prior knowledge of the LF is limited, it is reasonable to assume *a priori* that the LF should be unimodal, i.e., that the LF should not exhibit multiple peaks. To reflect this, we construct our prior distribution to place more probability on situations where the individual Gaussian functions are close together in terms of their widths. In addition, we only specify the parameteric form of the prior distribution, but allow the parameters of the prior distribution to vary and to be determined by the data. This allows our prior distribution to be flexible enough to have a minimal effect on the final results beyond conveying our prior consideration that the LF should be unimodal. We introduce our prior to place more probability on unimodal luminosity functions, to ensure that the posterior integrates to one, and to aid in convergence of the MCMC. Figure 4.3 illustrates the general idea that we are attempting to incorporate into our prior distribution. In this figure, we show a situation where the Gaussian functions are close together with respect to their widths, and far apart with respect to their widths. When the distances between the individual Gaussian functions, normalized by their covariance matrices (the measure of their 'width'), is small, the LF is unimodal; however, when the distances between the Gaussian functions are large with respect to to their covariance matrices, the LF exhibits multiple modes. We construct a prior distribution that places less probability on the latter situation.
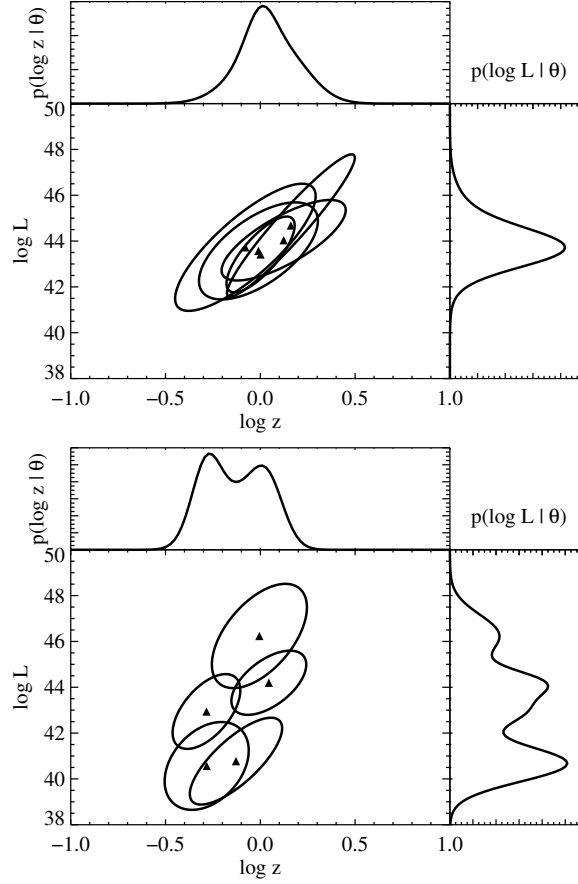
Figure 4.3 An illustration of our prior distribution for the Gaussian function parameters, for $K = 5$ Gaussian functions. Shown are a case when the Gaussian functions used in modelling the luminosity function are close together with respect to their covariances (top), and when the Gaussian functions are far apart with respect to their covariance matrices (bottom). The marginal distributions of $\log z$ are shown above the plots, and the marginal distributions of $\log L$ are shown to the right of the plots. When the Gaussian functions are close, the LF is unimodal, but when the Gaussian functions are far apart, the LF is multimodal. Because our prior distribution is constructed to place more probability on situations when the Gaussian functions are closer together with respect to their individual covariance matrices, it would place more probability on the situation shown in the left plot *a priori*. Our prior therefore reflects our expectation that the LF should not exhibit multiple peaks (modes).

Our prior on the Gaussian mean vectors and covariance matrices is similar to the prior described by Roeder & Wasserman (1997), but generalized to 2-dimensions. For our prior, we assume an independent multivariate Cauchy distribution for each of the Gaussian means, μ, with 2-dimensional mean vector μ_0 and 2×2 scale matrix T. A Cauchy distribution is equivalent to a student's t distributions with 1 degree of freedom, and when used as a function in astronomy and physics it is commonly referred to as a Lorentzian; we describe the Cauchy distribution further in Appendix C. The scale matrix is chosen to be the harmonic mean of the Gaussian function covariance matrices:

$$T = \left(\frac{1}{K} \sum_{k=1}^{K} \Sigma_k^{-1} \right)^{-1}. \tag{4.19}$$

Qualitatively, this prior means that we consider it more likely that the centroids of the individual Gaussian functions should scatter about some mean vector μ_0, where the width of this scatter should be comparable to the typical width of the individual Gaussian functions. The prior mean, μ_0, is left unspecified and is an additional free parameter to be estimated from the data. We choose a Cauchy prior because the Cauchy distribution is heavy tailed, and therefore does not heavily penalize the Gaussian functions for being too far apart. As a result, the Cauchy prior is considered to be robust compared to other choices, such as the multivariate normal distribution.

Because we use a random walk computational technique to explore the parameter space and estimate the posterior distribution, we find it advantageous to impose additional constraints on the Gaussian centroids. Both μ and μ_0 are constrained to the region $\log L_{low} \leq \mu_{l,k} \leq \log L_{high}$ and $\log z_{low} \leq \mu_{z,k} \leq \log z_{high}$, where $\mu_{l,k}$ is the mean in $\log L$ for the kth Gaussian, $\mu_{z,k}$ is the mean in $\log z$ for the kth Gaussian. These constraints are imposed to keep the markov chains (see § 4.6) from 'wandering' into unreasonable regions of the parameter space. The

flux limit sets a lower limit on the luminosity of detected sources as a function of z, and therefore there is nothing in the observed data to 'tell' the random walk that certain values of μ_l are unreasonable. For example, suppose our survey is only able to detect quasars with $L \gtrsim 10^{10} L_\odot$. Because of this, there is nothing in our data, as conveyed through the likelihood function, that says values of, say, $L \sim 10 L_\odot$ are unreasonable, and thus the markov chains can get stuck wandering around values of $\mu_l \sim 1$. However, we know *a priori* that values of $\mu_l \sim 1$ are unphysical, and therefore it is important to incorporate this prior knowledge into the posterior, as it is not reflected in the likelihood function. The values of these limits should be chosen to be physically reasonable. As an example, for the SDSS DR3 quasar LF with luminosities measured at $\lambda L_\lambda (2500\text{Å})$, it might be reasonable to take $L_{low} = 10^{40}$ erg s^{-1}, $L_{high} = 10^{48}$ erg s^{-1}, $z_{low} = 10^{-4}$, and $z_{high} = 7$.

Generalizing the prior of Roeder & Wasserman (1997), we assume independent inverse Wishart priors on the individual Gaussian covariance matrices with $\nu = 1$ degrees of freedom, and common scale matrix A. We give a description of the Wishart and inverse Wishart distributions in § C. This prior states that the individual Σ_k are more likely to be similar rather than different. The common scale matrix, A, is left unspecified so it can adapt to the data. As with μ, we recommend placing upper and lower limits on the allowable values of dispersion in $\log L$ and $\log z$ for each Gaussian..

Mathematically, our prior is

$$
p(\pi, \mu, \Sigma, \mu_0, A) \quad \propto \quad \prod_{k=1}^{K} p(\mu_k | \mu_0, \Sigma) p(\Sigma_k | A) \tag{4.20}
$$

$$
\propto \quad \prod_{k=1}^{K} \text{Cauchy}_2(\mu_k | \mu_0, T) \text{Inv-Wishart}_1(\Sigma_k | A), \tag{4.21}
$$

under the constraints given above. Here, $\text{Cauchy}_2(\mu_k | \mu_0, T)$ denotes a 2-dimensional Cauchy distribution as a function of μ_k, with mean vector μ_0 and scale matrix T.

In addition, Inv-Wishart$_1(\Sigma_k|A)$ denotes an inverse Wishart density as a function of Σ_k, with one degree of freedom and scale matrix A. We have also experimented with using a uniform prior on the parameters, constricted to some range. In general, this did not change our constraints on the LF above the flux limit, but resulted in somewhat wider confidence regions on the LF below the flux limit. This is to be expected, since our adopted Cauchy prior tends to restrict the inferred LF to be unimodal, and therefore limits the number of possible luminosity functions that are considered to be consistent with the data.

4.5.2 Posterior Distribution for Mixture of Gaussians Model

Now that we have formulated the prior distribution, we can calculate the posterior distribution for the mixture of Gaussians model of $\phi(L, z)$. Because we have formulated the mixture model for the LF in terms of $\log L$ and $\log z$, the marginal posterior distribution of θ is

$$p(\theta, \mu_0, A| \log L_{obs}, \log z_{obs}) \propto p(\theta, \mu_0, A)\, [p(I = 1|\theta)]^{-n} \prod_{i \in \mathcal{A}_{obs}} p(\log L_i, \log z_i|\theta), \ \ \theta = (\pi, \mu, \Sigma),$$

$$(4.22)$$

where $p(\theta, \mu_0, A)$ is given by Equation (4.21), $p(\log L_i, \log z_i|\theta)$ is given by Equation (4.17), and

$$p(I = 1|\theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(I = 1| \log L, \log z)p(\log L, \log z|\theta)\, d\log L\, d\log z \quad (4.23)$$

is the probability of including a source, given the model parameters θ. The conditional posterior distribution of N given π, μ, and Σ is given by inserting Equation (4.23) into (4.13). The complete joint posterior distribution is then

$$p(\theta, N, \mu_0, A| \log L_{obs}, \log z_{obs}) \propto p(N|\theta, n)p(\theta, \mu_0, A| \log L_{obs}, \log z_{obs}). \quad (4.24)$$

4.6   USING MARKOV CHAIN MONTE CARLO TO ESTIMATE THE POSTERIOR DISTRIBUTION OF THE LUMINOSITY FUNCTION

For our statistical model, μ_0 has 2 free parameters, A has 3 free parameters, and each of the K Gaussian components has 6 free parameters. Because the values of π are constrained to sum to one, there are only $6K - 1$ free parameters for the Gaussian mixture model. The number of free parameters in our statistical model is therefore $6K + 4$. The large number of parameters precludes calculation of the posterior on a grid of $\pi, \mu, \Sigma, \mu_0, A$, and N. Furthermore, the multiple integrals needed for marginalizing the posterior, and thus summarizing it, are numerically intractable. Because of this, we employ Markov Chain Monte Carlo (MCMC) to obtain a set of random draws from the posterior distribution. For the case of a random walk, a Markov chain is a random walk where the probability distribution of the current location only depends on the previous location. To obtain random numbers generated from the posterior distribution, one constructs a Markov chain that performs a random walk through the parameter space, where the Markov chain is constructued to eventually converge to the posterior distribution. Once convergence is reached, the values of the markov chain are saved at each iteration, and the values of these locations can be treated as a random draw from the posterior distribution. These draws may then be used to estimate the posterior distribution of $\phi(L, z)$, and thus an estimate of the LF and its uncertainty can be obtained.

In this work we use the Metropolis-Hastings algorithm (MHA, Metropolis & Ulam, 1949; Metropolis et al., 1953; Hastings, 1970) to perform the MCMC. We describe the particular MHA we employ for Bayesian inference on the LF; however for a more general and complete description of the MHA, we refer the reader to Chib & Greenberg (1995) or Gelman et al. (2004). We use the MHA to

obtain a set of random draws from the marginal posterior distribution of θ, given by Equation (4.22). Then, given these random draws of θ, random draws for N may be obtained directly from the negative binomial distribution (see Eq.[4.13] and Appendix C).

The basic idea behind the MHA is illustrated in Figure 4.4 for the special case of a symmetric jumping distribution. First, one starts with an initial guess for θ. Then, at each iteration a propsed value of θ is randomly drawn from some 'jumping' distribution. For example, this jumping distribution could be a normal density with some fixed covariance matrix, centered at the current value of θ. Then, if the proposed value of θ improves the posterior, it is stored as the new value of θ. Otherwise, it is stored as the new value with probability equal to the ratio of the values of the posterior distribution at the proposed and current value of θ. If the proposed value of θ is rejected, then the value of θ does not change, and the current value of θ is stored as the 'new' value of θ. The process is repeated until convergence. If the jumping distribution is not symmetric, then a correction needs to be made to the acceptance rule in order to account for asymetry in the jumping distribution. A jumping distribution is symmetric when the probability of jumping from a current value θ to a new value θ^* is the same as jumping from θ^* to θ. For example, the normal distribution is symmetric, while the log-normal distribution is not.

4.6.1 Metropolis-Hastings Algorithm for Schechter Luminosity Function

Before describing our MHA algorithm for the mixture of Gaussian functions model, we describe a simpler MHA algorithm for the Schechter function model given by Equation (4.14) in § 4.4.2. We do this to illustrate the MHA using a more familiar luminosity function. An MHA for obtaining random draws of α, L^*, and N from Equation (4.15) is:
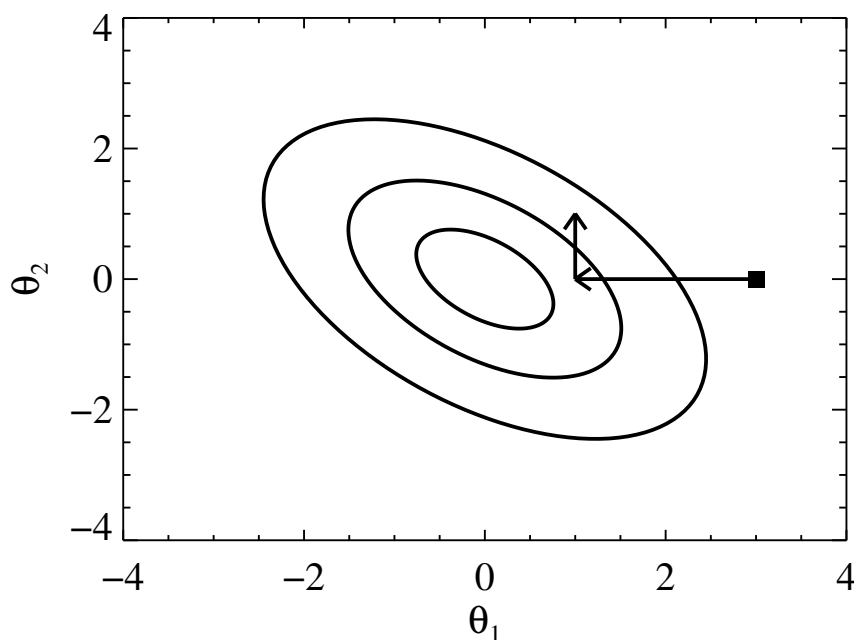
Figure 4.4 Schematic diagram illustrating the random walk Metropolis-Hastings algorithm. The posterior probability distribution is illustrated by the contours, and the random walk is initially at the position marked with a square. A new proposed value of θ_1 is randomly drawn, marked by the arrow pointing to the left. Because the proposed value of θ_1 is at a location with higher posterior probability, the new value of θ_1 is saved, and the random walk 'jumps' to the position marked by the arrow. Then, a new proposal for θ_2 is randomly drawn, marked by the arrow pointing upward. Because this proposed value of θ_2 is at a location with lower posterior probability, it is only accepted with probability equal to the ratio of the values of the posterior at the proposed position and the current position. If the proposed value is kept, then the new value of θ_2 is saved, otherwise the current value of θ_2 is saved. Next, a proposed value of θ_1 is randomly drawn, and the process repeats, creating a random walk through the parameter space. Because the amount of time that the random walk spends in any given bin in θ_1 and θ_2 is proportional to the posterior probability distribution, after the random walk has converged, the values of θ_1 and θ_2 from the random walk may be treated as a random draw from the posterior distribution.

1. Start with an initial value of α and L^*, denoted as $\tilde{\alpha}$ and $\tilde{L}^*$. A good initial value is the maximum-likelihood estimate.

2. Draw a proposal value of $\log L^*$ from a normal distribution centered on the current value of $\log L^*$, $\log \tilde{L}^*$. The variance in the jumping distribution of $\log L^*$ should be fixed at the beginning of the MHA. A larger jumping variance will lead to jumps that travel greater distances, but will then lead to lower MHA acceptance rates. The value of the jumping variance should be tuned to give acceptance rates ~ 0.4. We use a normal jumping distribution to vary $\log L^*$ because $\log L^*$ is defined on $(-\infty, \infty)$, while L^* is only defined on $(0, \infty)$. While we could use a jumping distribution to directly vary L^*, it is not always easy to simulate random variables directly from distributions that are only defined for $L^* > 0$.

   Denoting the proposal value of L^* as $\hat{L}^*$, calculate the ratio

   $$r_{L^*} = \frac{\hat{L}^* p(\tilde{\alpha}, \hat{L}^* | L_{obs})}{\tilde{L}^* p(\tilde{\alpha}, \tilde{L}^* | L_{obs})} \tag{4.25}$$

   Here, $p(\alpha, L^* | L_{obs})$ is the posterior distribution for the Schechter function, given by Equation (4.15). If $r_{L^*} > 1$, then keep the proposal and set $\tilde{L}^* = \hat{L}^*$. If $r_{L^*} < 1$, then draw a random number u uniformly distributed between 0 and 1. If $u < r_{L^*}$, then keep the proposal and set $\tilde{L}^* = \hat{L}^*$. Otherwise, the proposal is rejected and the value of $\tilde{L}^*$ is unchanged. The factor of $\hat{L}^*/\tilde{L}^*$ is necessary in Equation (4.25) in order to correct for the asymmetry in the log-normal jumping distribution.

3. Draw a proposal value of $\log k = \log(\alpha + 1)$ from a normal distribution centered at the current value of $\log k$, $\log \tilde{k} = \log(\tilde{\alpha} + 1)$. Similar to the MHA step for L^*, we use a normal jumping distribution to vary $\log k$ because $\log k$ is defined on $(-\infty, \infty)$, while α is only defined on $(-1, \infty)$.

Denoting the proposal value of k as $\hat{k}$, the proposal value of $\hat{\alpha}$ is $\hat{\alpha} = \hat{k} - 1$. Using the values of $\tilde{\alpha}$ and $\hat{\alpha}$, calculate the ratio

$$r_\alpha = \frac{\hat{k} p(\hat{\alpha}, \tilde{L}^* | L_{obs})}{\tilde{k} p(\tilde{\alpha}, \tilde{L}^* | L_{obs})} \tag{4.26}$$

If $r_\alpha > 1$, then keep the proposal and set $\tilde{\alpha} = \hat{\alpha}$. If $r_\alpha < 1$, then draw a random number u uniformly distributed between 0 and 1. If $u < r_\alpha$, then keep the proposal and set $\tilde{\alpha} = \hat{\alpha}$. Otherwise, the proposal is rejected and the value of $\tilde{\alpha}$ is unchanged. As with the MHA step for L^*, the factor of $\hat{k}/\tilde{k}$ is necessary in Equation (4.26) in order to correct for the asymmetry in the log-normal jumping distribution.

4. Repeat steps (2)–(3) until the MHA algorithm converges. Techniques for monitoring convergence are described in Gelman et al. (2004). After convergence, use Equation (4.13) to directly simulate random draws of the LF normalization, N, for each simulated value of α and L^* obtained from the above random walk. Equation (4.13) has the form of a negative binomial distribution, and a method for simulated random variables from the negative binomial distribution is described in Appendix C.

4.6.2 Metropolis-Hastings Algorithm for the Mixture of Gaussian Functions Luminosity Function

Our MHA for the mixture of Gaussian functions model is a more complex version of that used for the Schechter function model. As before, we denote the current value of a parameter by placing a ˜ over its symbol, and we denote the proposed value by placing a ˆ over its symbol. For example, if one were updating π, then $\tilde{\pi}$ denotes the current value of π in the random walk, and $\hat{\pi}$ denotes the proposed value of π. We will only update one parameter at a time, so, if we are drawing a

proposal for π, the current value of θ is denoted as $\tilde{\theta} = (\tilde{\pi}, \tilde{\mu}, \tilde{\Sigma})$, and the proposed value of θ is denoted as $\hat{\theta} = (\hat{\pi}, \tilde{\mu}, \tilde{\Sigma})$.

Our MHA for the mixture of Gaussian functions model is:

1. Start with initial guesses for $\pi, \mu, \Sigma, \mu_0, A$, and T.

2. Draw a proposal value for π from a $\text{Dirichlet}(\tilde{g}_1, \ldots, \tilde{g}_K)$ density, where $\tilde{g}_k = c_\pi n \tilde{\pi}_k + 1$, n is the number of sources in the survey, and c_π is a fixed positive constant that controls how far the 'jumps' in π go. Because c_π controls the variance of the Dirichlet density, a smaller value of c_π produces values of $\hat{\pi}$ that are further from $\tilde{\pi}$. The value of c_π should be chosen so that about 15–40% of the MHA proposals are accepted.

   After drawing a proposal for π, calculate the value of the posterior distribution at the new value of θ, $\hat{\theta}$, and at the old value of θ, $\tilde{\theta}$. Then, use these values to calculate the ratio

$$r_\pi = \frac{\text{Dirichlet}(\tilde{\pi}|\hat{g})}{\text{Dirichlet}(\hat{\pi}|\tilde{g})} \frac{p(\hat{\theta}|L_{obs}, z_{obs})}{p(\tilde{\theta}|L_{obs}, z_{obs})}, \tag{4.27}$$

   where $\hat{g} = c_\pi n \hat{\pi}_1 + 1, \ldots, c_\pi n \hat{\pi}_K + 1$. The ratio of Dirichlet densities in Equation (4.27) corrects the MHA acceptance rule for the asymmetry in the Dirichlet jumping distribution. If $r_\pi \geq 1$ then keep the proposed value of π: $\tilde{\pi} = \hat{\pi}$. Otherwise keep the proposal with probability r_π. This is done by drawing a uniformly distributed random variable between 0 and 1, denoted by u. If $u < r_\pi$, then set $\tilde{\pi} = \hat{\pi}$. If $u > r_\pi$ then keep the current value of π.

   Methods for simulating from the Dirichlet distribution, as well as the functional form of the Dirichlet distribution, are given in § C.

3. For each Gaussian function, draw a proposal for μ_k by drawing $\hat{\mu}_k \sim N_2(\tilde{\mu}_k, V_k)$, where V_k is some set covariance matrix. Because the jumping density is symmetric, the MHA acceptance ratio is just given by the ratio of the posterior distributions at the proposed and current value of μ_k: $r_\mu = p(\hat{\theta}|L_{obs}, z_{obs})/p(\tilde{\theta}|L_{obs}, z_{obs})$. If $r_\mu \geq 1$ then set $\tilde{\mu}_k = \hat{\mu}_k$, otherwise set $\tilde{\mu}_k = \hat{\mu}_k$ with probability r_μ. The MHA update should be performed seperately for each Gaussian. The covariance matrix of the jumping kernel, V_k, should be chosen such that $\sim 30\%$ of the MHA jumps are accepted.

Since we have constructed the prior distribution with the constraint $\log L_{low} \leq \mu_{l,k} \leq \log L_{high}$ and $\log z_{min} \leq \mu_{z,k} \leq \log z_{high}$ for all k, any values of $\hat{\mu}_k$ that fall outside of this range should automatically be rejected.

4. For each Gaussian, draw a proposal for Σ_k by drawing $\hat{\Sigma}_k \sim \text{Wishart}_{\nu_k}(\tilde{\Sigma}_k/\nu_k)$, where ν_k is some set degrees of freedom. Larger values of ν_k will produce values of $\hat{\Sigma}_k$ that are more similar to $\tilde{\Sigma}_k$. The MHA acceptance ratio is

$$r_\Sigma = \left(\frac{|\tilde{\Sigma}_k|}{|\hat{\Sigma}_k|}\right)^{\nu_k - 3/2} \exp\left\{-\frac{\nu_k}{2}tr\left[(\hat{\Sigma}_k)^{-1}\tilde{\Sigma}_k - \tilde{\Sigma}_k^{-1}\hat{\Sigma}_k\right]\right\} \frac{p(\hat{\theta}|x_{obs})}{p(\tilde{\theta}|x_{obs})}, \qquad (4.28)$$

where $tr(\cdot)$ denotes the trace of a matrix. If $r_\Sigma \geq 1$ then set $\tilde{\Sigma}_k = \hat{\Sigma}_k$, otherwise set $\tilde{\Sigma}_k = \hat{\Sigma}_k$ with probability r_Σ. The MHA update should be performed seperately for each Gaussian. The degrees of freedom of the jumping kernel, ν_k, should be chosen such that $\sim 15\text{--}40\%$ of the MHA jumps are accepted.

If there are any bounds on Σ_k incorporated into the prior distribution, then values of Σ_k that fall outside of this range should automatically be rejected. Methods for simulating from the Wishart distribution, as well as the functional form of the Wishart distribution, are given in § C.

5. Draw a proposal for the prior parameter μ_0 as $\hat{\mu}_0 \sim N_2(\tilde{\mu}_0, V_0)$. The acceptance ratio only depends on the prior distribution and is

$$r_0 = \left[\prod_{k=1}^{K} \frac{\mathrm{Cauchy}_2(\mu_k|\hat{\mu}_0, T)}{\mathrm{Cauchy}_2(\mu_k|\tilde{\mu}_0, T)} \right] \left[\frac{\int_{\log L_{low}}^{\log L_{high}} \int_{\log z_{low}}^{\log z_{high}} \mathrm{Cauchy}_2(\mu_k|\tilde{\mu}_0, T)\, d\mu_k}{\int_{\log L_{low}}^{\log L_{high}} \int_{\log z_{low}}^{\log z_{high}} \mathrm{Cauchy}_2(\mu_k|\hat{\mu}_0, T)\, d\mu_k} \right]^{K}.$$
(4.29)

Here, T is given by Equation (4.19) and the integrals are needed because of the prior constraints on μ. If $r_0 \geq 1$ then set $\tilde{\mu}_0 = \hat{\mu}_0$, otherwise set $\tilde{\mu}_0 = \hat{\mu}_0$ with probability r_0. We have found a good choice for V_0 to be the sample covariance matrix of $\tilde{\mu}$.

6. Finally, update the value of A, the common scale matrix. Because we can approximately calculate the conditional distribution of A, given Σ, we can directly simulate from $p(A|\Sigma)$. Directly simulating from the conditional distributions is referred to as a Gibbs sampler. We perform a Gibbs update to draw a new value of $\tilde{A}$:

$$\hat{A} \sim \mathrm{Wishart}_{\nu_A}(S) \tag{4.30}$$

$$\nu_A = K + 3 \tag{4.31}$$

$$S = \left(\sum_{k=1}^{K} \tilde{\Sigma}_k^{-1} \right)^{-1}. \tag{4.32}$$

For the Gibbs sampler update, we do not need to calculate an acceptance ratio, and every value of $\hat{A}$ is accepted: $\tilde{A} = \hat{A}$. If there are any prior bounds set on Σ, then this is technically only an approximate Gibbs update, as it ignores the constraint on Σ. A true MHA update would account for the constraint on Σ by renormalizing the conditional distribution appropriately; however, this involves a triple integral that is expensive to compute. If there are prior bounds on Σ, then Equation (4.30) is approximately correct, and ignoring the normalization in $p(A|\Sigma)$ does not did not have any effect on our results.

Steps 2–6 are repeated until the MCMC converges, where one saves the values of $\tilde{\theta}$ at each iteration. After convergence, the MCMC is stopped, and the values of $\tilde{\theta}$ may be treated as a random draw from the marginal posterior distribution of θ, $p(\theta | \log L_{obs}, \log z_{obs})$. Techniques for monitering convergence of the Markov Chains are described in Gelman et al. (2004). If one wishes to assume a uniform prior on μ and Σ, constrained within some set range, instead of the prior we suggest in § 4.5.1, then only steps 2–4 need to be performed. Given the values of θ obtained from the MCMC, one can then draw values of N from the negative binomial density (cf. Eq.[4.13]). In § C we describe how to simulate random variables from a negative binomial distribution. The speed of our MHA algorithm depends on the sample size and the programming language. As a rough guide, on a modern computer our MHA can take a couple of hours to converge for sample sizes of ~ 1000, and our MHA can take as long as a day or two to converge for sample sizes $\sim 10^4$.

When performing the MCMC it is necessary to perform a 'burn-in' stage, after which the markov chains have approximately converged to the posterior distribution. The values of θ from the MCMC during the burn-in stage are discarded, and thus only the values of θ obtained after the burn-in stage are used in the analysis. We have found it useful to perform $\sim 10^4$ iterations of burn-in, although this probably represents a conservative number. In addition, the parameters for the MHA jumping distributions should be tuned during the burn-in stage. In particular, the parameters $\Sigma_\alpha, \sigma_{\sigma_l}^2, c_\pi, \Sigma_{\mu,k}$, and ν_k should be varied within the burn-in stage to make the MHA more efficient and have an acceptance rate of ~ 0.15–0.4 (Gelman, Roberts, & Gilks, 1995). These jumping distribution parameters cannot be changed after the burn-in stage. Jasra et al. (2005) and Neal (1996) described additional complications and considerations developing MHAs for mixture mod-

els.

Some post processing of the Markov chains is necessary. This is because some chains can get 'stuck' wandering in regions far below flux limit, likely in the presence of a local maximum in the posterior. While such chains will eventually converge and mix with the other chains, they do not always do so within the finite number of iterations used when running the random walk MHA. In general, we have found that divergent chains are easy to spot. Because the divergent chains usually get stuck in regions far below the flux limit, they correspond to luminosity functions with implied extremely low detection probabilities, i.e., $p(I = 1|\theta) \ll 1$. As a result, the random draws of N from the posterior for these chains tend to have values that are too high and far removed from the rest of the posterior distribution of N. The divergent chains are therefore easily found and removed by inspecting a histogram of $\log N$. In fact, we have found that the divergent chains often become too large for the long integer format used in our computer routines, and therefore are returned as negative numbers. Because negative values of N are unphysical, it is easy to simply remove such chains from the analysis.

Having obtained random draws of N and θ from $p(\theta, N | \log L_{obs}, \log z_{obs})$, one can then use these values to calculate an estimate of $\phi(L, z)$, and its corresponding uncertainty. This is done by inserting the MCMC values of θ and N directly into Equation (4.18). The posterior distribution of $\phi(L, z)$ can be estimated for any value of L and z by plotting a histogram of the values of $\phi(L, z)$ obtained from the MCMC values of θ and N. In § 4.7, we illustrate in more detail how to use the MHA results to perform statistical inference on the LF.

4.7   APPLICATION TO SIMULATED DATA

As an illustration of the effectiveness of our method, we applied it to a simulated data set. We construct a simulated sample, and then recover the luminosity function based on our mixture of Gaussian functions model. We assume the effective survey area and selection function reported for the DR3 quasar sample (Richards et al., 2006).

4.7.1   Construction of the Simulated Sample

We first drew a random value of N_Ω quasars from a binomial distribution with probability of success $\Omega/4\pi = 0.0393$ and number of trials $N = 3 \times 10^5$. Here, $\Omega = 1622 \deg^2$ is the effective sky area for our simulated survey, and we chose the total number of quasars to be $N = 3 \times 10^5$ in order to ultimately produce a value of $n \sim 1300$ observed sources, after accounting for the SDSS selection function. This first step of drawing from a binomial distribution simulates a subset of $N_\Omega \sim 1.2 \times 10^4$ sources from N total sources randomly falling within an area Ω on the sky. For simplicity, in this simulation we ignore the effect of obscuration on the observed quasar population. While our choice of $N = 3 \times 10^5$ produces a much smaller sample than the actual sample of $n \sim 1.5 \times 10^4$ quasars from the SDSS DR3 luminosity function work (Richards et al., 2006), we chose to work with this smaller sample to illustrate the effectiveness of our method on more moderate sample sizes.

For each of these $N_\Omega \sim 1.2 \times 10^4$ sources, we simulated values of L and z. We first simulated values of $\log z$ from a marginal distribution of the form

$$f(\log z) = \frac{4\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\exp(a\zeta^*)}{(1 + \exp(\zeta^*))^{a+b}}, \tag{4.33}$$

where $\zeta^* = 4(\log z - 0.4)$. The parameters $a = 1.25$ and $b = 2.5$ were chosen to give an observed redshift distribution similar to that seen for SDSS DR3 quasars

(e.g., Richards et al., 2006). Values of $\log z$ are easily drawn from Equation (4.33) by first drawing $x^* \sim \text{Beta}(a, b)$, and then setting $\log z = \text{logit}(x^*)/4 + 0.4$; here, $\text{Beta}(a, b)$ is a beta probability density, and $\text{logit}(x) = \ln(x/(1-x))$ is the logit function.

For each simulated value of z, we simulated a value of L using a similar functional form. The conditional distribution of $\log L$ given z is

$$f(\log L|z) = \frac{\Gamma(\alpha(z) + \beta(z))}{\Gamma(\alpha(z))\Gamma(\beta(z))} \frac{(L/L^*(z))^{\alpha(z)/\ln 10}}{[1 + (L/L^*(z))^{1/\ln 10}]^{\alpha(z)+\beta(z)}} \tag{4.34}$$

$$\alpha(z) = 6 + \log z \tag{4.35}$$

$$\beta(z) = 9 + 2 \log z \tag{4.36}$$

$$L^*(z) = 10^{45} z^2, \tag{4.37}$$

where $L^*(z)$ approximately marks the location of the peak in $f(\log L|z)$, $t(z)$ is the age of the universe in Gyr at redshift z, $\alpha(z)$ is the slope of $\log f(\log L|z)$ for $L \lesssim L^*(z)$, and $\beta(z)$ is the slope of $\log f(\log L|z)$ for $L \gtrsim L^*(z)$. In this simulated 'universe', both the peak and logarithmic slopes of the LF evolve. The form of the luminosity function assumed by Equation (4.34) is similar to the double power-law form commonly used in the quasar LF literature, but has a more gradual transition between the two limiting slopes.

After using Equations (4.33) and (4.34) to generate random values of L and z, we simulated the effects at a selection function. We randomly kept each source for $z < 4.5$, where the probability of including a source given its luminosity and redshift was taken to be the SDSS DR3 Quasar selection function, as reported by Richards et al. (2006). After running our simulated sample through the selection function, we were left with a sample of $n \sim 1300$ sources. Therefore, our simulated survey is only able to detect $\sim 0.4\%$ of the $N = 3 \times 10^5$ total quasars in our simulated 'universe'. The distributions of L and z are shown in Figure 4.5

for both the detected sources and the full sample. As can be seen, the majority of sources are missed by our simulated survey.

The joint probability distribution of L and z is $f(L, z) = f(L|z)f(z)$, and therefore Equations (4.33) and (4.34) imply that the true LF for our simulated sample is

$$\phi_0(L, z) = \frac{N}{zL(\ln 10)^2} \left(\frac{dV}{dz}\right)^{-1} f(\log L|z)f(\log z) \tag{4.38}$$

Figure 4.6 shows $\phi_0(L, z)$ at several redshifts. Also shown in Figure 4.6 is the best fit for a mixture of $K = 4$ gaussian functions. Despite the fact that $\phi_0(L, z)$ has a rather complicated parameteric form, a mixture of four gaussian functions is sufficient to achieve an excellent approximation to $\phi_0(L, z)$; in fact, the mixture of four gaussian functions approximation is indistinguishable from the true LF.

4.7.2   Performing Statistical Inference on the LF with the MCMC Output

We performed the MHA algorithm described in § 4.6 to obtain random draws from the posterior probability distribution for our this simulated sample, assuming the Gaussian mixture model described in § 4.5. We performed 10^4 iterations of burn-in, and then ran the markov chains for 3×10^4 more iterations. We ran 20 chains simultaneously in order to monitor convergence (e.g., see Gelman et al., 2004) and explore possible multimodality in the posterior. We saved the values of θ for the Markov chains after the initial 10^4 burn-in iterations, and, after removing divergent chains with $N < 0$ we were left with $\sim 8 \times 10^4$ random draws from the posterior distribution, $p(\theta, N|L_{obs}, z_{obs})$.

The output from the MCMC can be used to perform statistical inference on the LF. Denote the $T \sim 8 \times 10^4$ random draws of θ and N obtained via the MHA as $\theta^1, \ldots, \theta^T$ and $N^1, \ldots, N^T$, respectively. The individual values of (θ^t, N^t) can then be used to construct histograms as estimates of the posterior distribution for each parameter. For each random draw of θ and N, we can also calculate a random
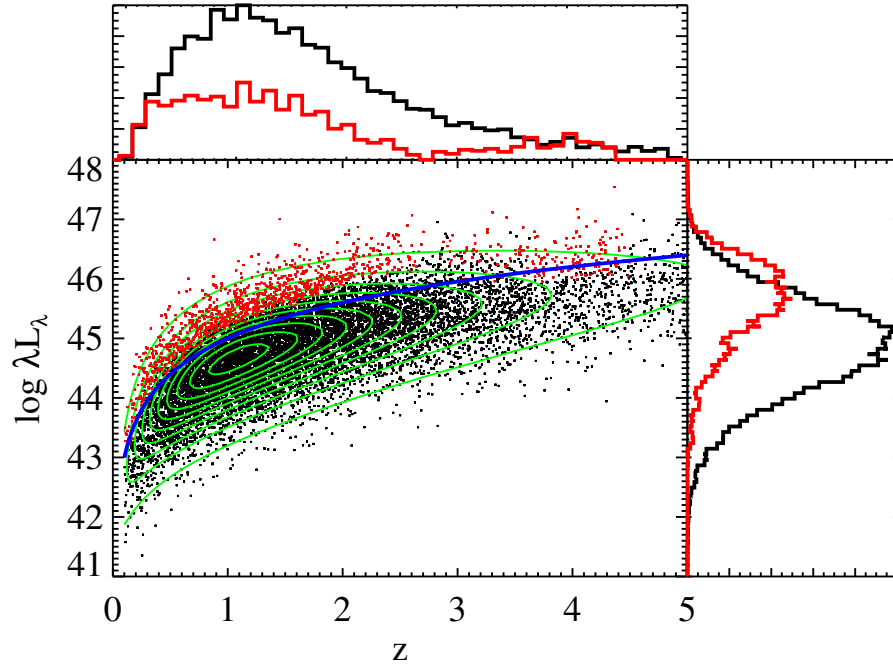
Figure 4.5 The distribution of L and z for the simulated sample described in
§ 4.7.1. Red dots denote sources included in the sample, black dots denote sources
not included in the sample, the blue line denotes L^* as a function of z, and the
green contours display the 2-d luminosity function. Also shown are histograms
of the marginal distributions of $\log L$ and z, for all simulated objects (black his-
togram) and only the detected ones (red histogram). For clarity, the histogram of
the detected sources has been forced to peak at a value equal to half of the peak
of the histogram of all objects.
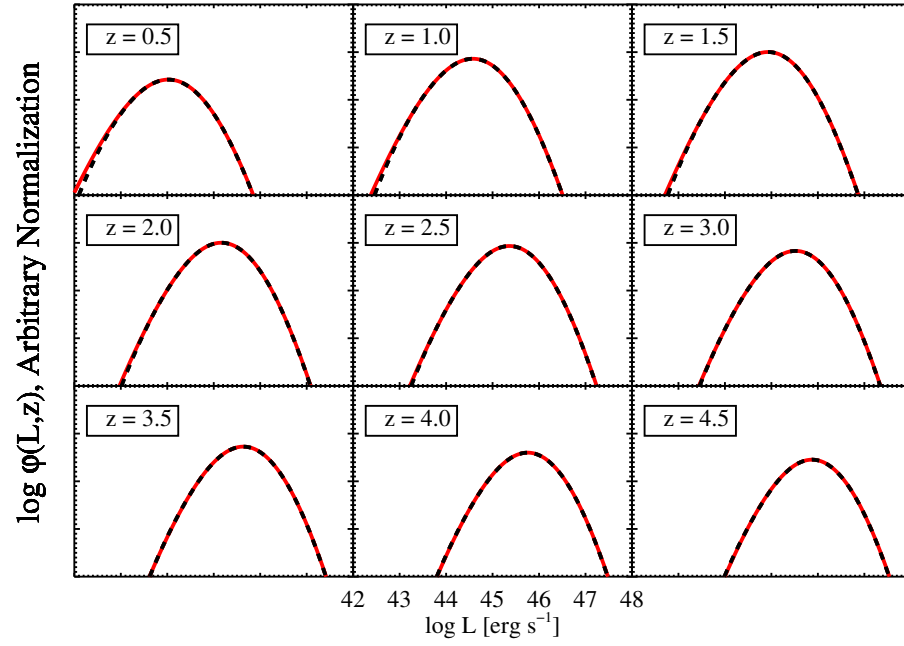
Figure 4.6 The true LF (solid red line) at several values of z, and the best $K = 4$ Gaussian function fit (dashed black line). In this case, approximating the LF with four 2-dimensional Gaussian functions provides an excellent fit.

.

draw of $\phi(L, z)$ from its posterior distribution. In particular, the tth random draw of the LF under the mixture of normals model, denoted as $\phi^t(L, z)$, is calculated by inserting θ^t and N^t into Equation (4.18). The T values of $\phi^t(L, z)$ can then be used to estimate the posterior distribution of $\phi(L, z)$ for any given value of L and z. Furthermore, random draws from the posterior for quantities that are computed directly from the LF, such as the location of its peak as a function of z, are obtained simply by computing the quantity of interest from each of the T values of $\phi^t(L, z)$.

In Figures 4.7 and 4.8 we show $\phi(\log L, z)$ at several different redshifts, on both a linear scale and a logarithmic scale. In general, we find it easier to work with $\phi(\log L, z) = \ln 10 L \phi(L, z)$, as $\phi(\log L, z)$ can span several orders of magnitude in L. Figures 4.7 and 4.8 show the true value of the LF, $\phi_0(\log L, z)$, the best-fit estimate of $\phi(\log L, z)$ based on the mixture of Gaussian functions model, and the regions containing 90% of the posterior probability. Here, as well as throughout this work, we will consider the posterior median of any quantity to be the 'best-fit' for that quantity. In addition, in this work we will report errors at the 90% level, and therefore the regions containing 90% of the posterior probability can be loosely interpreted as asymmetric error bars of length $\approx 1.65\sigma$. The region containing 90% of the probability for $\phi(\log L, z)$ is easily estimated from the MCMC output by finding the values of t_1 and t_2 such that 90% of the values of $\phi^1(\log L, z), \ldots, \phi^T(\log L, z)$ have $\phi^{t_1}(\log L, z) < \phi^t(\log L, z) < \phi^{t_2}(\log L, z)$. As can be seen, the true value of $\phi(\log L, z)$ is contained within the 90% probability region for all almost values of L, even those below the survey detection limit.

Figure 4.9 compares the true integrated $z < 6$ number distribution of $\log L$, $n(\log L, z < 6)$, with the mixture of Gaussian functions estimate. The quantity $n(\log L, z < 6) d \log L$ gives the number of quasars at $z < 6$ with black hole masses
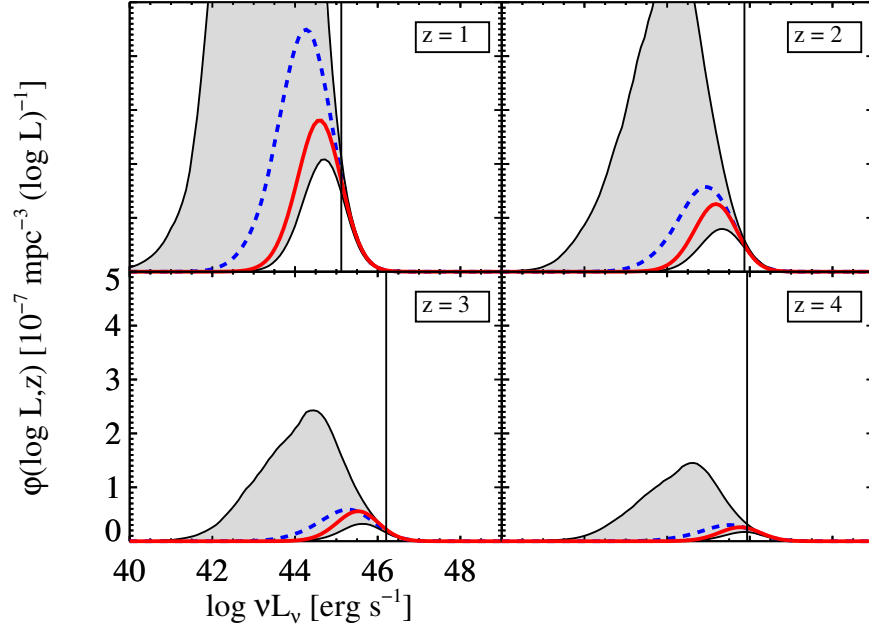
Figure 4.7 The true LF (solid red line) at several redshifts for the simulated sample described in § 4.7.1. The axis labels are the same for all panels, but for clarity we only label the bottom left panel. Also shown is the posterior median estimate of the LF based on the mixture of Gaussian functions model (dashed blue line), the region containing 90% of the posterior probability (shaded region). The bayesian mixture of Gaussian functions model is able to accurately constrain the LF, even below the survey detection limit.
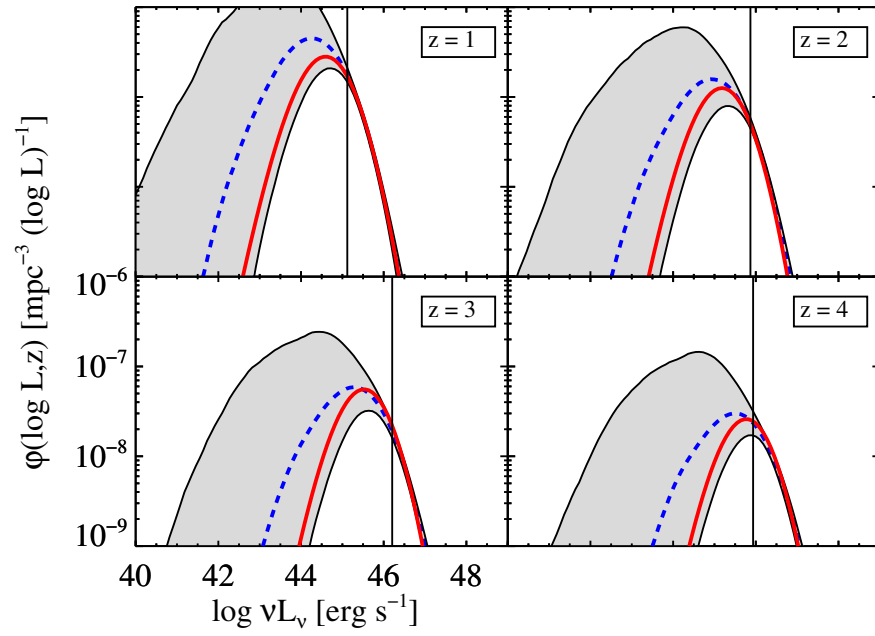
Figure 4.8 Same as Figure 4.7, but shown with a logarithmic stretch.

between $\log L$ and $\log L + d \log L$. It is calculated as

$$n(\log L, z < 6) = \int_0^6 \phi(\log L, z) \left(\frac{dV}{dz}\right) dz, \tag{4.39}$$

which, for the mixture of normals model, is

$$n(\log L, z < z_0) = N \sum_{k=1}^K \pi_k N(\log L | \mu_{l,k}, \sigma_{l,k}^2) \Phi \left[\frac{\log z_0 - E(\log z | L, k)}{\sqrt{Var(\log z | L, k)}}\right] \tag{4.40}$$

$$E(\log z | L, k) = \mu_{z,k} + \frac{\sigma_{lz,k}}{\sigma_{z,k}^2} (logL - \mu_{l,k}) \tag{4.41}$$

$$Var(\log z | L, k) = \sigma_{z,k}^2 - \frac{\sigma_{lz,k}^2}{\sigma_{z,k}^2}. \tag{4.42}$$

Here, $\Phi(\cdot)$ is the cumulative distribution function for the standard normal density. Similar to Figures 4.7 and 4.8, the true value of $n(\log L, z < 6)$ is contained within the 90% probability region for all values of L, even those below the survey detection limit.

In addition, in Figure 4.9 we show the comoving number density of broad line AGN as a function of redshift, $n(z)$. This is obtained by integrating $\phi(L, z)$ over all possible values of L. For the mixture of normals model, this becomes

$$n(z) = N \left(\frac{dV}{dz}\right)^{-1} \sum_{k=1}^K \pi_k p(z | k), \tag{4.43}$$

where the marginal distribution of $z | k$ is

$$p(z | k) = \frac{1}{z \ln 10 \sqrt{2\pi\sigma_{z,k}^2}} \exp \left\{-\frac{1}{2}\left(\frac{\log z - \mu_{z,k}}{\sigma_{z,k}}\right)^2\right\}. \tag{4.44}$$

As before, the true value of $n(z)$ is contained within the 90% probability region, despite the fact that the integration extends over *all* L, even those below the detection limit. The wider confidence regions reflect additional uncertainty in $n(z)$ resulting from integration over those L below the detection limit. In particular, the term dV/dz becomes small at low redshift, making the estimate of $n(z)$ more unstable as $z \to 0$, and thus inflating the uncertainties at low z.
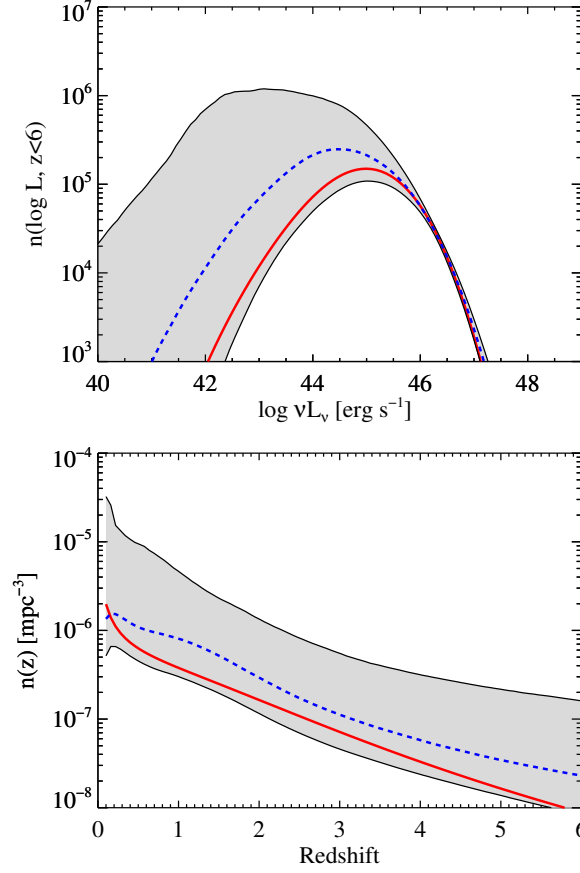
Figure 4.9 The integrated $z < 6$ quasar number density (number per $\log L$ interval, top) and the quasar comoving quasar number density as a function of z (number per mpc^3, bottom) for the simulated sample described in § 4.7.1. As with Figure 4.7, the solid red line denotes the true value for the simulation, the dashed blue line denotes the posterior median for the mixture of Gaussian functions model, and the shaded region contain 90% of the posterior probability. The posterior median provides a good fit to the true values, and the uncertainties derived from the MCMC algorithm based on the Gaussian mixture model are able to accurately constrain the true values of these quantities, despite the flux limit.

Two other potentially useful quantities are the comoving luminosity density for quasars, $\rho_L(z)$, and its derivative. The comoving quasar luminosity density is given by $\rho_L(z) = \int_0^\infty L\phi(L,z)\, dL$. For the mixture of Gaussian functions model it may be shown that

$$
\begin{aligned}
\rho_L(z) &= N\left(\frac{dV}{dz}\right)^{-1} \sum_{k=1}^{K} \pi_k p(z|k) \\
&\times \exp\left\{\ln 10 E(\log L|z,k) + \frac{(\ln 10)^2}{2} Var(\log L|z,k)\right\} \quad (4.45) \\
E(\log L|z,k) &= \mu_{l,k} + \frac{\sigma_{lz,k}}{\sigma_{z,k}^2}(\log z - \mu_{z,k}) \quad (4.46) \\
Var(\log L|z,k) &= \sigma_{l,k}^2 - \frac{\sigma_{lz,k}^2}{\sigma_{z,k}^2}, \quad (4.47)
\end{aligned}
$$

where $p(z|k)$ is given by Equation (4.44). We calculate the derivative of $\rho_L(z)$ numerically. Figure 4.10 compares the true values of $\rho_L(z)$ and its derivative with the posterior distribution for $\rho_L(z)$ inferred from the mixture model, both as a function of z and the age of the universe at redshift z, $t(z)$. Comparison with Figure 4.9 reveals that the comoving quasar luminosity density, $\rho_L(z)$, is a better constrained quantity than the comoving quasar number density, $n(z)$. Furthermore, $n(z)$ appears to peak much later than $\rho_L(z)$. In addition, we can correctly infer that the comoving quasar luminosity density reaches it point of fastest growth at $t(z) \sim 2$ Gyr, and its point of fastest decline at $t(z) \sim 5$ Gyr.

Figure 4.11 quantifies the suggestion that $n(z)$ peaks later than $\rho_L(z)$ by displaying the posterior distribution for the location of the respective peaks in $n(z)$ and $\rho_L(z)$. We can still constrain the peak in $n(z)$ to be at $z \lesssim 0.5$. In contrast, the location of the peak in $\rho_L(z)$ is constrained to occur earlier at $1 \lesssim z \lesssim 3$. This is a consequence of the fact that while there were more quasars per comoving volume element in our simulated universe at $z \lesssim 0.5$, their luminosities were much higher at higher redshift. This evolution in characteristic L is quantified in Figure
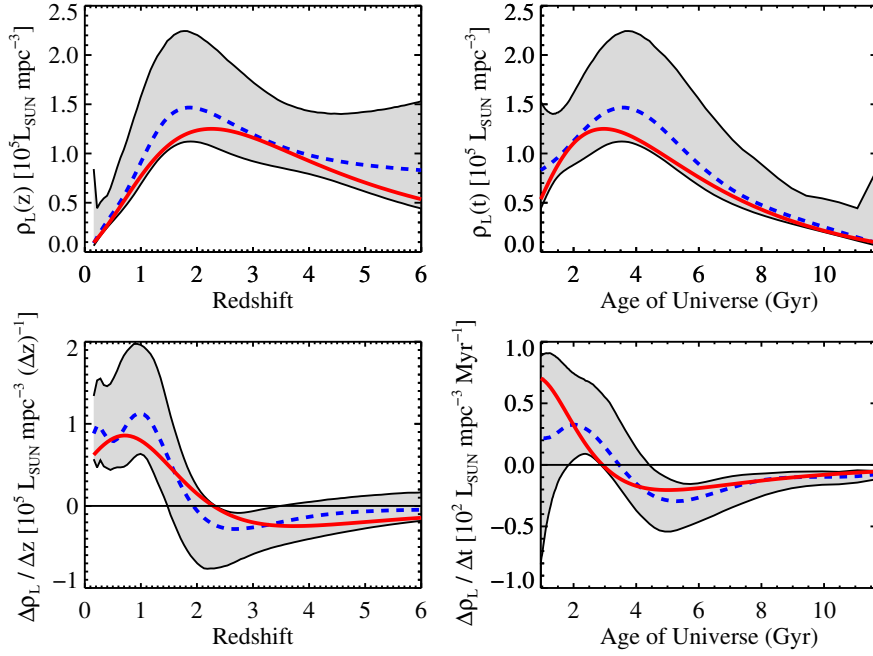
Figure 4.10 Comoving quasar luminosity density (top two panels) and its deriva-
tive (bottom two panels), shown as a function of redshift (left two panels) and
cosmic age (right two panels) for the simulated sample described in § 4.7.1. The
plotting symbols are the same as in Figure 4.9. As in the previous figures, the
Gaussian mixture model is able to provide an accurate fit to the true values of
$\rho_L(z)$, and the bayesian MCMC approach is able to provide accurate constraints
on $\rho_L(z)$ and $d\rho_L/dz$, despite the fact that the integral used for calculating these
quanties extends below the survey detection limit.

4.12, which summarizes the posterior distribution for the location of the peak in $\phi(\log L, z)$ as a function of redshift and $t(z)$. As can be seen, the location of the peak in the LF shows a clear trend of increasing 'characteristic' L with increasing z, although there is considerable uncertainty on the actual value of the location of the peak.

4.7.3 Using the MCMC Output to Evaluate the LF Fit

Throughout this section we have been analyzing the MCMC results by comparing to the true LF. However, in practice we do not have access to the true LF, and thus a method is needed for assessing the quality of the fit. The statistical model may be checked using a technique known as posterior predictive checking (e.g., Rubin, 1981, 1984; Gelman, Meng, & Stern, 1998). Here, the basic idea is to use each of the MCMC outputs to simulate a new random observed data set. The distributions of the simulated observed data sets are then compared to the true observed data in order to assess whether the statistical model gives an accurate representation of the observed data. It is important to construct simulated data sets for each of the MCMC draws in order to incorporate our uncertainty in the model parameters.

For each value of N^t and θ^t obtained from the MCMC output, a simulated data set of (l_{obs}^t, z_{obs}^t) may be obtained through a similar procedure to that described in § 4.7.1. First, one draws a value of N_Ω^t from a binomial distribution with N^t trials and probability of 'success' $p = \Omega/4\pi$. Then, one draws N_Ω^t values of L^t and z^t from $p(L, z|\theta^t)$.

For our model, $p(\log L, \log z|\theta^t)$ is a mixture of normal densities, and one needs to employ a two-step process in order to simulate a random value from $p(\log L, \log z|\theta^t)$. First, one needs to randomly assign the ith data point to one of the Gaussian distributions. Since π_k gives the probability that a data point will be drawn from the
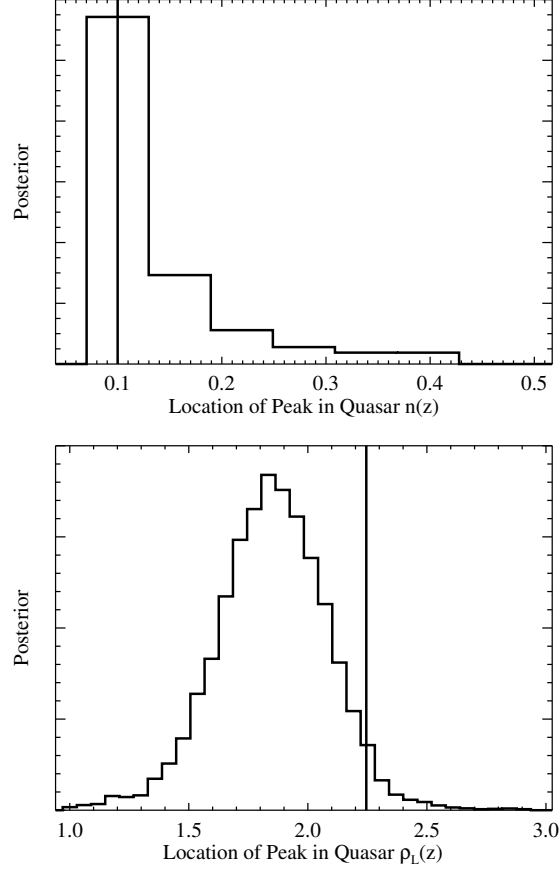
Figure 4.11 Posterior distribution for the redshift location of the peak in the co-moving number density of quasars (top) and the peak in the comoving quasar luminosity density (bottom) for the simulated sample described in § 4.7.1. For clarity we only show the posterior distribution for the peak in $n(z)$ at $z > 0.5$, since values of the peak at $z < 0.5$ arise because the term $(dV/dz)^{-1}$ becomes very large at low z. The vertical lines denote the true values. The posterior distribution inferred from the MCMC output is able to accurately constrain the true values of the argumentative maximum in $n(z)$ and $\rho_L(z)$.
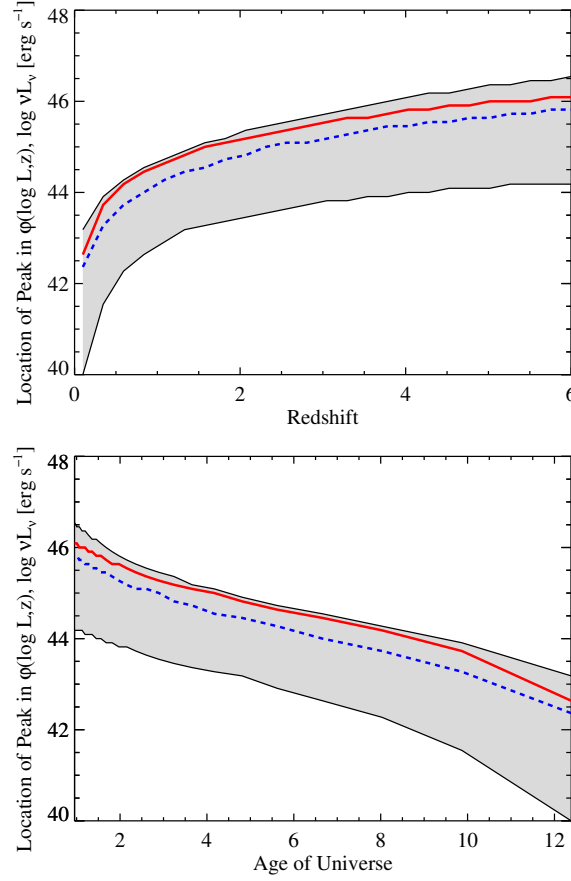
Figure 4.12 Location of the peak in the LF as a function of z (top) and cosmic age (bottom) for the simulated sample described in § 4.7.1. The plot symbols are the same is in Figure 4.9. In general the posterior median of the Gaussian mixture model provides a good estimate of the true peak locations, although the uncertainty is high due to the survey flux limit. However, it is clear from these plots that the location of the peak in $\phi(L, z)$ evolves.

k^{th} Gaussian distribution, one first needs to simulate a random vector $\mathbf{G}_i^t$ from a multinomial distribution with one trial and probability of success for the kth class π_k^t; i.e., first draw $\mathbf{G}_i^t \sim \text{Multinom}(1, \pi_1^t, \ldots, \pi_K^t)$. The vector $\mathbf{G}_i^t$ gives the class membership for the ith data point, where $G_{ik}^t = 1$ if the ith data point comes from the kth Gaussian, and $G_{ij}^t = 0$ if $j \neq k$. Then, given $G_{ik}^t = 1$, one then simulates a value of $(\log L_i^t, \log z_i^t)$ from a 2-dimensional Gaussian distribution with mean μ_k^t and covariance matrix Σ_k^t. This is repeated for all N_Ω^t sources, leaving one with a random sample $(\log L^t, \log z^t) \sim p(\log L, \log z | \theta^t)$.

A random draw from $\text{Multinom}(1, \pi_1, \ldots, \pi_K)$, may be obtained as a sequence of binomial random draws. First, draw $n_1' \sim \text{Binomial}(1, \pi_1)$. If $n_1' = 1$, then assign the data point to the first Gaussian distribution, i.e., set $G_{i1} = 1$. If $n_1' = 0$, then draw $n_2' \sim \text{Binomial}(1, \pi_2 / \sum_{k=2}^{K} \pi_k)$. If $n_2' = 1$, then assign the data point to the second Gaussian distribution, i.e., set $G_{i2} = 1$. If $n_2' = 0$, then the process is repeated for the remaining Gaussian distribution as follows. For $j = 3, \ldots, K-1$, sequentially draw $n_j' \sim \text{Binomial}(1, \pi_j / \sum_{k=j}^{K} \pi_k)$. If at any time $n_j' = 1$, then stop the process and assign the data point to the jth Gaussian distribution. Otherwise, if none of the $n_j' = 1$, then assign the data point to the Kth Gaussian distribution.

Once one obtains a random draw of (L^t, z^t), randomly 'observe' these sources, where the probability of including a source given L_i^t and z_i^t is given by the selection function. This will leave one with a simulated observed data set, (L_{obs}^t, z_{obs}^t). This process is repeated for all T values of N^t and θ^t obtained from the MCMC output, leaving one with T simulated data sets of (L_{obs}^t, z_{obs}^t). One can then compare the distribution of the simulated data sets with the true values of L_{obs}, and z_{obs} to test the statistical model for any inconsistencies.

In Figure 4.13 we show histograms for the true observed distributions of z and $\log L$. These histograms are compared with the posterior median of the dis-

tributions based on the mixture of Gaussian functions model, as well as error bars containing 90% of the simulated values. Also shown is a plot comparing the true values of the maximum of L_{obs} as a function of z with those based on L_{obs}^t and z_{obs}^t. As can be seen, the distributions of the observed data assuming the mixture of Gaussian functions model are consistent with the true distributions of the observed data, and therefore there is no reason to reject the mixture model as providing a poor fit.

4.8   SUMMARY

We have derived the observed data likelihood function which relates the quasar LF to the observed distribution of redshifts, luminosities. This likelihood function is then used in a Bayesian approach to estimating the LF, where the LF is approximated as a mixture of Gaussian functions. Because much of this work was mathematically technical, we summarize the important points here.

- Equation 4.6 gives the likelihood function for an assumed parameteric luminosity function. This likelihood function differs from the poisson likelihood commonly used in the LF literature because it correctly models the sample size as a binomial random variable, whereas the poisson likelihood approximates the sample size as a poisson random variable. In practice, the difference in the maximum-likelihood estimates obtained from the two likelihood functions do not seem to be significantly different so long as the probability of including a source in a survey is small.

- The product of Equations (4.12) and (4.13) is the joint posterior probability distribution of the LF, given the observed data. These equations may be used to perform Bayesian inference on the LF, after assuming a prior distribution on the LF parameters. Bayesian inference is often most eas-
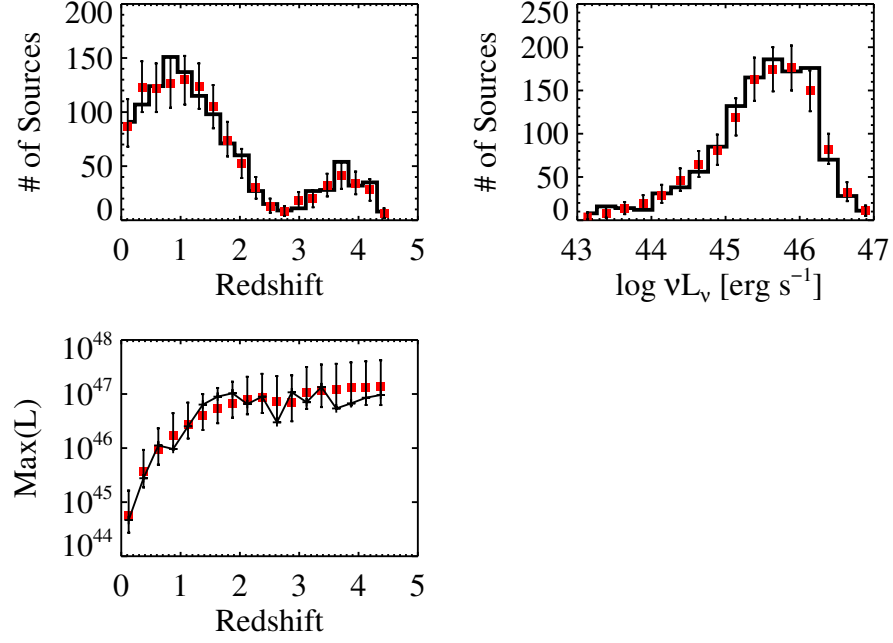
Figure 4.13 Posterior predictive check for the Gaussian mixture model (see § 4.7.3). The histograms show the actual distributions of L_{obs} and z_{obs}, the red squares denote the posterior medians for the number of sources in each respective bin, and the error bars contain the inner 90% of the histogram values for the samples simulated from the posterior. Also shown is a plot of the maximum observed luminosity as a function of z for the simulated samples, the red squares mark the value from the actual sample used in the fit, and the error bars contain 90% of the values simulated from the posterior distribution. The mixture of Gaussian functions model is able to provide an accurate prediction of the observed distribution of luminosity, redshift, and line widths, and thus there is not any evidence to reject it as providing a poor fit.

ily performed by simulating random variables drawn from the posterior probability distribution. These random draws may be used to estimate the posterior distribution for the LF, as well as to estimate the posterior distribution for any quantities calculated from the LF. The posterior distribution provides statistically accurate uncertainties on the LF and related quantities, even when the sample size is small and one is including information below the survey detection limits. In contrast, confidence intervals derived from bootstrapping the maximum-likelihood estimate can be too small.

- We describe a flexible model for the LF, where the LF is modelled as a mixture of Gaussian functions. Equation (4.17) describes the probability distribution of $\log L$ and $\log z$ under the mixture of Gaussian functions model, and (4.18) describes the LF under the mixture of Gaussian functions model.

  Equation (4.21) gives our prior distribution for the Gaussian function parameters. The marginal posterior distribution of the mixture model parameters is given by Equation (4.22), the conditional posterior distribution of N at a given θ is given by Equation (4.13), and the complete joint posterior distribution is the product of Equations (4.22) and (4.13).

- We describe in § 4.6 a Metropolis-Hastings algorithm for obtaining random draws from the posterior distribution for the LF assuming a Schechter function or mixture of Gaussian functions model. In § 4.7, we use a simulated sample, modelled after the SDSS DR3 quasar catalogue, to illustrate the effectiveness of our statistical method, as well as to give an example on how to use the Metropolis-Hastings output to perform statistical inference on the LF and assess the LF fit.

So long as the mixture of Gaussian functions model is an accurate approximation to the true LF over all luminosities, the uncertainties on the LF, assuming the Gaussian mixture model, are trustworthy because they are calculated directly from the probability distribution of the LF, given the observed data. Statistical inference on the LF below the flux limit can become significantly biased if one assumes an incorrect and restrictive parameteric form, as extrapolation errors can become large. In this case, the derived posterior may not contain the true LF because the wrong parameteric model was assumed; this type of error is known as model mispecification. For example, consider a case when the true LF is a Schechter function, and one is only able to detect sources brighter than L^*. If one were to assume a power-law for the LF, extrapolation below the flux limit would be signficantly biased. In contrast, the mixture of Gaussian functions model, while incorrect, is flexible enough to accurately approximate the true Schechter function form, thus minimizing extrapolation bias due to model mispecification. Of course, in this example, the most accurate results would be obtained by fitting a Schechter function to the data, since it is the correct parameteric form. Therefore, the mixture of Gaussian functions model will not perform as well as assuming the correct parameteric model, or at least as well as an alternative parameteric model that better approximates the true LF.

Although we have focused on the mixture of Gaussian functions model, the likelihood and posterior distribution are applicable for any parameteric form, as illustrated in § 4.4.2. The observed data likelihood function for the LF is given by Equation (4.6), and the posterior distribution is given by the product of Equations (4.12) and (4.13). Then, one can use Equation (4.1) to 'plug-in' any parameteric form of the LF into the appropriate likelihood function and posterior distribution, as was done in Equation (4.15) for a Schechter function. In addition, the

metropolis-hastings algorithm is a general method for obtaining random draws from the posterior, and can be developed for any parameteric form of the LF. This therefore allows one to perform Bayesian inference for any variety of parameteric models of the LF, and one is not merely limited to the mixture of Gaussian functions model or Schechter function considered in this work.

CHAPTER 5

DETERMINING QUASAR BLACK HOLE MASS FUNCTIONS FROM THEIR BROAD
EMISSION LINES: APPLICATION TO THE BRIGHT QUASAR SURVEY

5.1   CHAPTER ABSTRACT

We describe a Bayesian approach to estimating quasar black hole mass functions (BHMF) when using the broad emission lines to estimate black hole mass. We show how using the broad line mass estimates in combination with statistical techniques developed for luminosity function estimation (e.g., the $1/V_a$ correction) leads to biased results. We derive the likelihood function for the BHMF based on the broad line mass estimates, and derive the posterior distribution for the BHMF, given the observed data. We develop our statistical approach for a flexible model where the BHMF is modelled as a mixture of Gaussian functions. Statistical inference is performed using markov chain monte carlo (MCMC) methods, and we describe a metropolis-hasting algorithm to perform the MCMC. The MCMC simulates random draws from the probability distribution of the BHMF parameters, given the data, and we use a simulated data set to show how these random draws may be used to estimate the probability distribution for the BHMF. In addition, we show how the MCMC output may be used to estimate the probability distribution of any quantities derived from the BHMF, such as the peak in the space density of quasars. Our method has the advantage that it is able to place accurate constraints on the BHMF even beyond the survey detection limits, accounts for measurement errors and the intrinsic uncertainty in broad line mass estimates, and provides a natural way of estimating the probability distribution of any quantities derived from the BHMF. We conclude by using our method to estimate the local active BHMF using the $z < 0.5$ Bright Quasar Survey sources.

At $z \sim 0.2$, the quasar BHMF falls off approximately as a power law with slope ~ 2 for $M_{BH} \gtrsim 10^8 M_\odot$. We find that at a given M_{BH}, $z < 0.5$ broad line quasars have a typical Eddington ratio of ~ 0.4 and a dispersion in Eddington ratio of $\lesssim 0.5$ dex.

5.2 CHAPTER INTRODUCTION

It is widely accepted that the extraordinary activity associated with quasars[1] involves accretion onto a supermassive black hole (SMBH). The correlation between SMBH mass and both host galaxy luminosity (e.g., Kormendy & Richstone, 1995; Magorrian et al., 1998; McLure & Dunlop, 2001; Marconi & Hunt, 2003) and stellar velocity dispersion (M_{BH}–σ relationship, e.g., Gebhardt et al., 2000a; Merritt & Ferrarese, 2001; Tremaine et al., 2002), together with the fact that quasars have been observed to reside in early-type galaxies (McLure et al., 1999; Kukula et al., 2001; McLeod & McLeod, 2001; Nolan et al., 2001; Percival et al., 2001; Dunlop et al., 2003), implies that the evolution of spheroidal galaxies and quasars is intricately tied together (e.g., Silk & Rees, 1998; Haehnelt & Kauffmann, 2000; Merritt & Poon, 2004; Di Matteo et al., 2005; Hopkins et al., 2006a). Therefore, investigating the evolution of active super-massive black holes (SMBHs) is an important task of modern astronomy, giving insight into the importance of AGN activity on the formation of structure in the universe. Determination of the comoving number density, energy density, and mass density of active black holes is a powerful probe of the quasar-galaxy connection and the evolution of active black holes.

Recently, advances in reverberation mapping (e.g., Peterson et al., 2004) have made it possible to estimate the masses of black holes for broad line AGN. A correlation has been found between the size of the region emitting the broad lines

---

[1]Throughout this work we will use the terms quasar and AGN to refer generically to broad line AGNs. No luminosity difference between the two is assumed.

and the luminosity of the AGN (Kaspi et al., 2005; Bentz et al., 2006), allowing one to use the source luminosity to estimate the distance between the broad line region (BLR) and the central black hole. In addition, one can estimate the velocity dispersion of the BLR gas from the broad emission line width. One then combines the BLR size estimate with the velocity estimate to obtain a virial black hole mass (e.g., Wandel et al., 1999; McLure & Dunlop, 2002; Vestergaard & Peterson, 2006). Estimates of M_{BH} obtained from the broad emission lines have been used to estimate the distribution of quasar black hole masses at a variety of redshifts (e.g., McLure & Dunlop, 2004; Vestergaard, 2004; Kollmeier et al., 2006; Wang et al., 2006; Greene & Ho, 2007; Vestergaard et al., 2008).

Given the importance of the BHMF as an observational constraint on models of quasar evolution, it is essential that a statistically accurate approach be employed when estimating the BHMF. However, the existence of complicated selection functions hinders this. A variety of methods have been used to accurately account for the selection function when estimating the quasar luminosity function. These include various binning methods (e.g., Schmidt, 1968; Avni & Bahcall, 1980; Page & Carrera, 2000), maximum-likelihood fitting (e.g., Marshall et al., 1983; Fan et al., 2001), a semi-parameteric approach (Schafer, 2007), and a Bayesian approach we develop in a companion paper (Kelly, Fan, & Vestergaard, 2008, hereafter KFV08). In addition, there have been a variety of methods proposed for estimating the cumulative distribution function of the luminosity function (e.g., Lynden-Bell, 1971; Efron & Petrosian, 1992; Maloney & Petrosian, 1999). While these techniques have been effective for estimating luminosity functions, estimating the BHMF from the broad line mass estimates is a more difficult problem, and currently there does not exist a statistically correct method of estimating the BHMF.

If we could directly measure black hole mass for quasars, and if the selection function only depended on M_{BH} and z, then we could simply employ the formalism described in KFV08 after replacing L with M_{BH}. However, surveys are selected based on luminosity and redshift, not on M_{BH}. At any given luminosity there exists a scatter in black hole mass, and thus one cannot simply employ the luminosity selection function 'as-is' to correct for the flux limit. In other words, completeness in flux is not the same thing as completeness in M_{BH}, and the use of a flux selection results in a softer selection function for M_{BH}. Moreover, we cannot directly observe M_{BH} for large samples of quasars, but rather derive an estimate of M_{BH} from their broad emission lines. The intrinsic uncertainty on M_{BH} derived from the broad emission lines is ~ 0.4 dex (Vestergaard & Peterson, 2006), and the uncertainty on M_{BH} broadens the inferred distribution of M_{BH} (e.g., Kelly & Bechtold, 2007; Shen et al., 2007). As a result, even if there is no flux limit, the BHMF inferred directly from the broad line mass estimates will be systematically underestimated near the peak and overestimated in the tails. In order to ensure an accurate estimate of the BHMF it is important to correct for the uncertainty in the estimates of M_{BH}.

Motivated by these issues, we have developed a Bayesian method for estimating the BHMF. In KFV08 we derived the likelihood function and posterior probability distribution for luminosity function estimation, and we described a mixture of Gaussian functions model for the luminosity function. In this work, we extend our statistical method and derive the likelihood function of the BHMF by relating the observed data to the true BHMF, and derive the posterior probability distribution of the BHMF parameters, given the observed data. While the likelihood function and posterior are valid for any parameteric form, we focus on a flexible parameteric model where the BHMF is modeled as a sum of Gaus-

sian functions. This is a type of 'non-parameteric' approach, where the basic idea is that the individual Gaussian functions do not have any physical meaning, but that given enough Gaussian functions one can obtain a suitably accurate approximation to the true BHMF. Modelling the BHMF as a mixture of normals avoids the problem of choosing a particular parameteric form, especially in the absence of any guidance from astrophysical theory. In addition, we describe a markov chain monte carlo (MCMC) algorithm for obtaining random draws from the posterior distribution. These random draws allow one to estimate the posterior distribution for the BHMF, as well as any quantities derived from it. The MCMC method therefore allows a straight-forward method of calculating errors on any quantity derived from the BHMF. Because the Bayesian approach is valid for any sample size, one is able to place reliable constraints on the BHMF and related quantities, even where the survey becomes incomplete.

Because of the diversity and mathematical complexity of some parts of this paper, we summarize the main results here. We do this so that the reader who is only interested in specific aspects of this paper can conveniently consult the sections of interest.

- In § 5.3.2 we derive the general form of the likelihood function for black hole mass function estimation based on quasar broad emission lines. Because we can not directly observe M_{BH} for a large sample of quasars, the likelihood function gives the probability of observing a set of redshifts, luminosities, and line widths, given an assumed BHMF. In § 5.3.3 we derive the black hole mass selection function, and discuss how the differences between the M_{BH} selection function and the luminosity selection function affect estimating the BHMF. The reader who is interested in the likelihood function of the broad line quasar BHMF, or issues regarding correcting for incompleteness

in M_{BH}, should consult this section.

- In § 5.4 we describe a Bayesian approach to black hole mass function estimation. We build on the likelihood function derived in § 5.3.2 to derive the probability distribution of the BHMF, given the observed data (i.e., the posterior distribution). The reader who is interested in a Bayesian approach to BHMF estimation should consult this section.

- In § 5.5 we develop a mixture of Gaussian functions model for the black hole mass function, deriving the likelihood function and posterior distribution for this model. Under this model, the BHMF is modelled as a weighted sum of Gaussian functions. This model has the advantage that, given a suitably large enough number of Gaussian functions, it is flexible enough to give an accurate estimate of any smooth and continuous BHMF. This allows the model to adapt to the true BHMF, thus minimizing the bias that can result when assuming a parameteric form for the BHMF. In addition, we also describe our statistical model for the distribution of luminosities at a given M_{BH}, and the distribution of line widths at a given L and M_{BH}. These two distribution are necessary in order to link the BHMF to the observed set of luminosities and line widths. The reader who are interested in employing our mixture of Gaussian functions model should consult this section.

- Because of the large number of parameters associated with black hole mass function estimation, Bayesian inference is most easily performed by obtaining random draws of the BHMF from the posterior distribution. In § 5.6 we describe a Metropolis-Hastings algorithm (MHA) for obtaining random draws of the BHMF from the posterior distribution, assuming our mixture of Gaussian functions model. The reader who is interested in the computa-

tional aspects of 'fitting' the mixture of Gaussian functions model, or who is interested in the computational aspects of Bayesian inference for the BHMF, should consult this section.

- In § 5.7 we use simulation to illustrate the effectiveness of our Bayesian Gaussian mixture model for black hole mass function estimation. We construct a simulated data set similar to the Sloan Digital Sky Survey DR3 Quasar Cataloge (Schneider et al., 2005). We then use our mixture of Gaussian functions model to recover the true BHMF and show that our mixture model is able to place reliable constraints on the BHMF over all values of M_{BH}. In constrast, we show that estimating the BHMF by binning up the broad line mass estimates, and applying a simple $1/V_a$ correction, systematically biases the inferred BHMF toward larger M_{BH}. We also illustrate how to use the MHA output to constrain any quantity derived from the BHMF, and how to use the MHA output to assess the quality of the fit. Finally, we discuss difficulties associated with inferring the distribution of Eddington ratios. The reader who is interested in assessing the effectiveness of our statistical approach, or who is interested in using the MHA output for statistical inference on the BHMF, should consult this section.

- In § 5.8 we use our statistical method to estimate the $z < 0.5$ BHMF from the Bright Quasar Survey sources. We also attempt to infer the mean and dispersion in the $z < 0.5$ distribution of Eddington ratios. The reader who is interested in the scientific results regarding our estimated $z < 0.5$ BHMF should consult this section.

We adopt a cosmology based on the the WMAP best-fit parameters ($h = 0.71, \Omega_m = 0.27, \Omega_\Lambda = 0.73$, Spergel et al., 2003)

5.3 THE LIKELIHOOD FUNCTION

5.3.1 NOTATION

We use the common statistical notation that an estimate of a quantity is denoted by placing a 'hat' above it; e.g., $\hat{\theta}$ is an estimate of the true value of the parameter θ. We denote a normal density with mean μ and variance σ^2 as $N(\mu, \sigma^2)$, and we denote as $N_p(\mu, \Sigma)$ a multivariate normal density with p-element mean vector μ and $p \times p$ covariance matrix Σ. If we want to explicitly identify the argument of the Gaussian function, we use the notation $N(x|\mu, \sigma^2)$, which should be understood to be a Gaussian function with mean μ and variance σ^2 as a function of x. We will often use the common statistical notation where "$\sim$" means "is drawn from" or "is distributed as". This should not be confused with the common usage of implying "similar to". For example, $x \sim N(\mu, \sigma^2)$ states that x is drawn from a normal density with mean μ and variance σ^2, whereas $x \sim 1$ states that the value of x is similar to one.

5.3.2 Likelihood Function for the BHMF Estimated from AGN Broad Emission Lines

The black hole mass function, denoted as $\phi(M_{BH}, z)dM_{BH}$, is the number of sources per comoving volume $V(z)$ with black hole masses in the range $M_{BH}, M_{BH} + dM_{BH}$. The black hole mass function is related to the probability density of (M_{BH}, z) by

$$p(M_{BH}, z) = \frac{1}{N}\phi(M_{BH}, z)\frac{dV}{dz}, \tag{5.1}$$

where N is the total number of sources in the universe, and is given by the integral of ϕ over M_{BH} and $V(z)$. If we assume a parameteric form for $\phi(M_{BH}, z)$, with parameters θ, we can derive the likelihood function for the observed data. The likelihood function is the probability of observing one's data, given the as-

sumed model. The presense of selection effects and intrinsic uncertainty in the broad line mass estimates can make this difficult, as the observed data likelihood function is not simply given by Equation (5.1). However, we can account for these difficulties by first deriving the likelihood function for the complete set of data, and then integrating over the missing data to obtain the observed data likelihood function.

For broad line AGNs, we can relate the distribution of M_{BH} and z to the joint distribution of $L_\lambda, \mathbf{v}$, and z. Here, $\mathbf{v} = (v_{H\beta}, v_{MgII}, v_{CIV})$, where $v_{H\beta} = v_{H\beta}$ is the the velocity dispersion for the Hβ broad line emitting gas, and similarly for v_{MgII} and v_{CIV}. These three lines are commonly used in estimating M_{BH} from single-epoch spectra of broad line AGN (e.g., McLure & Dunlop, 2002; Kaspi et al., 2005; Vestergaard & Peterson, 2006), where the velocity dispersion is typically estimated from the $FWHM$ of the emission line. The distribution of L_λ and $\mathbf{v}$ are then related to the BHMF via the R–L relationship and the virial theorem.

The BHMF for broad line AGN can be inferred from the distribution of $L_\lambda, \mathbf{v}$, and z, and thus it is necessary to formulate the observed data likelihood function in terms of $(L_\lambda, \mathbf{v}, z)$. While it is possible to formulate the likelihood function in terms of the broad line mass estimates, denoted as $\hat{M}_{BL} \propto L_\lambda^{1/2} V^2$, the broad line mass estimates are simply linear combinations of $\log L_\lambda$ and $\log \mathbf{v}$, and thus statistical inference does not depend on whether we formulate the likelihood function in terms of L_λ and $\mathbf{v}$ or $\hat{M}_{BL}$. We find it mathematically simpler and more intuitive to infer the BHMF directly from the distribution of $L_\lambda, \mathbf{v}$, and z, as opposed to inferring it from the distribution of $L_\lambda, \hat{M}_{BL}$, and z.

Following the discussion in KFV08, we derive the likelihood function for the set of observed luminosities, redshifts, and emission line widhts. We introduce an indicator variable I denoting whether a source is included in the survey or not:

if $I_i = 1$ then a source is included, otherwise, $I_i = 0$. The variable I is considered to be part of the observed data in the sense that we 'observe' whether a source is detected or not. The survey selection function is the probability of including the ith source in one's survey, $p(I_i = 1|\mathbf{v}_i, L_{\lambda,i}, z_i)$. Here, we have assumed that the probability of including a source in one's sample only depends on luminosity, redshift, and emission line width, and is therefore conditionally independent of M_{BH}. This is the case, in general, since one can only select a survey based on quantities that are directly observable. Including the additional 'data' I, the observed data likelihood function for broad line AGN is:

$$p(\mathbf{v}_{obs}, L_{obs}, z_{obs}, I|\theta, N) \propto \tag{5.2}$$

$$C_n^N \prod_{i \in \mathcal{A}_{obs}} \int p(\mathbf{v}_i, L_{\lambda,i}, M_{BH,i}, z_i|\theta) \, dM_{BH,i} \tag{5.3}$$

$$\times \prod_{j \in \mathcal{A}_{mis}} \int \int \int \int p(I = 0|\mathbf{v}_j, L_{\lambda,j}, z_j) p(\mathbf{v}_j, L_{\lambda,j}, M_{BH,j}, z_j|\theta) \tag{5.4}$$

$$\times d\mathbf{v}_j \, dL_{\lambda,j} dM_{BH,j} \, dz_j \tag{5.5}$$

$$\propto \quad C_n^N \left[p(I = 0|\theta) \right]^{N-n} \prod_{i \in \mathcal{A}_{obs}} p(\mathbf{v}_i, L_{\lambda,i}, z_i|\theta), \tag{5.6}$$

where $\mathcal{A}_{obs}$ denotes the set of sources included in one's survey, $\mathcal{A}_{mis}$ denotes the set of sources not included in one's survey, and on the last line we have omitted terms that do not depend on N or θ. Here,

$$p(\mathbf{v}_i, L_{\lambda,i}, z_i|\theta) = \int_0^\infty p(\mathbf{v}_i, L_{\lambda,i}, z_i, M_{BH,i}|\theta) \, dM_{BH,i} \tag{5.7}$$

is the probability of observing values of $\mathbf{v}_i$, $L_{\lambda,i}$, and z_i for the ith source, given θ, and

$$p(I = 0|\theta) = \int_0^\infty \int_0^\infty \int_0^\infty p(I = 0|\mathbf{v}, L_\lambda, z) p(\mathbf{v}, L_\lambda, z|\theta) \, d\mathbf{v} \, dL_\lambda \, dz \tag{5.8}$$

is the probability that the survey misses a source, given θ; note that $p(I = 0|\theta) = 1 - p(I = 1|\theta)$. Qualitatively, the observed data likelihood function for the BHMF

is the probability of observing a set of n emission line widths $\mathbf{v}_1, \ldots, \mathbf{v}_n$, luminosities $L_{\lambda,1}, \ldots, L_{\lambda,n}$, and redshifts $z_1, \ldots, z_n$ given the assumed BHMF model parameterized by θ, multiplied by the probability of not detecting $N - n$ sources given θ, multiplied by the number of ways to select a subset of n sources from a set of N total sources. Equation (5.5) can be maximized to calculate a maximum likelihood estimate of the black hole mass function when using broad line estimates of M_{BH}, or combined with a prior distribution to perform Bayesian inference.

It is often preferred to write the BHMF observed data likelihood function by factoring the joint distribution of $\mathbf{v}, L_\lambda, M_{BH}$, and z into conditional distributions. This has the advantage of being easier to interpret and work with, especially when attempting to connect the distribution of line widths and luminosities to the distribution of black hole mass. The joint distribution can be factored as (Kelly & Bechtold, 2007)

$$p(\mathbf{v}, L_\lambda, M_{BH}, z) = p(\mathbf{v}|L_\lambda, M_{BH}, z)p(L_\lambda|M_{BH}, z)p(M_{BH}, z). \qquad (5.9)$$

Here, $p(\mathbf{v}|L_\lambda, M_{BH}, z)$ is the distribution of emission line widths at a given L_λ, M_{BH}, and z, $p(L_\lambda|M_{BH}, z)$ is the distribution of luminosities at a given M_{BH} and z, and $p(M_{BH}, z)$ is the probability distribution of black hole mass and redshift, related to the BHMF via Equation (5.1). When using broad line estimates of M_{BH}, it is assumed that $p(\mathbf{v}|L_\lambda, M_{BH}, z)$ is set by the virial theorem, where the distance between the central black hole and the broad line-emitting gas depends on L_λ via the R–L relationship. In this work we assume that the R–L relationship does not depend on z (e.g., Vestergaard, 2004), and thus $p(\mathbf{v}|L_\lambda, M_{BH}, z) = p(\mathbf{v}|L_\lambda, M_{BH})$.

Under the factorization given by Equation (5.8), the observed data likelihood function (Eq. [5.5]) becomes

$$p(\mathbf{v}_{obs}, L_{obs}, z_{obs}, I|\theta, N) \propto$$

$$C_n^N \left[p(I = 0|\theta)\right]^{N-n} \tag{5.10}$$

$$\times \prod_{i \in \mathcal{A}_{obs}} \int_0^\infty p(\mathbf{v}_i|L_{\lambda,i}, M_{BH,i}, \theta) p(L_{\lambda,i}|M_{BH,i}, z, \theta) p(M_{BH,i}, z|\theta) \; dM_{BH,i} \tag{5.11}$$

The BHMF likelihood function, given by Equation (5.5) or (5.9), is entirely general, and it is necessary to assume parametric forms in order to make use of it. In § 5.5 we describe a parametric form based on a mixture of Gaussians model, and explicitly calculate Equation (5.5) for the mixture model.

5.3.3   Selection Function

The selection probability, $p(I = 1|\mathbf{v}, L_\lambda, z)$, depends on both the luminosity and redshift through the usual flux dependence, but can also depend on the emission line width. In particular, an upper limit on $\mathbf{v}$ may occur if there is a width above which emission lines become difficult to distinguish from the continuum and iron emission. In this case, if all emission lines in one's spectrum are wider than the maximum line width than one is not able to obtain a reliable estimate of the line width for any emission line, and therefore the source is not used to estimate $\phi(M_{BH}, z)$. A lower limit on the line width may be imposed in order to prevent the inclusion of narrow line AGN, for which broad line mass estimates are not valid. In this case the inclusion criterion might be that at least one emission line is broader than, say, $FWHM = 2000 \; \mathrm{km \; s^{-1}}$. In addition to the limits on line width that may be imposed, there is an upper and lower limit on z due to redshifting of emission lines out of the observable spectral range. For example, if one uses optical spectra than the range of useable spectra is $0 < z \lesssim 4.5$, as the C IV line redshifts into the near-infrared for $z \gtrsim 4.5$.

Denote the upper and lower limit of $\mathbf{v}$ as v_{min} and v_{max}, and the upper and lower limit of z as z_{min} and z_{max}. Furthermore, denote the usual survey selection function in terms of L_λ and z as $s(L_\lambda, z)$, where $s(L_\lambda, z)$ is the probability that a

source is included in the survey before any cuts on line width are imposed; $s(l, z)$ would typically correspond to the selection function used in luminosity function estimation. Note that in this work $s(L_\lambda, z)$ gives the probability that any source in the universe is included in the survey, given its luminosity and redshift, and thus $s(L_\lambda, z) \leq \Omega/4\pi$, where $\Omega/4\pi$ is the fraction of the sky covered by the survey. Then, $p(I_i = 1|\mathbf{v}_i, L_{\lambda,i}, z_i) = s(L_{\lambda,i}, z_i)$ if $z_{min} \leq z_i \leq z_{max}$ and at least one emission line has $v_{min} \leq v_i \leq v_{max}$; otherwise, $p(I_i = 1|\mathbf{v}_i, L_{\lambda,i}, z_i) = 0$. In this case, the probability that a source is included in the survey (see Eq.[5.7]) is

$$p(I = 1|\theta) =$$

$$\int_0^\infty \int_{z_{min}}^{z_{max}} s(L_\lambda, z) \tag{5.12}$$

$$\times \int_{v_{min}}^{v_{max}} \int_0^\infty p(\mathbf{v}|L_\lambda, M_{BH}, \theta) p(L_\lambda|M_{BH}, z, \theta) p(M_{BH}, z|\theta) \, dM_{BH} \, d\mathbf{v} \, dz \, dL_\lambda \tag{5.13}$$

where the inner two integrals are over $\mathbf{v}$ and M_{BH}, and the outer two integrals are over L_λ and z. One can then plug Equation (5.10) into Equation (5.5) to get the likelihood function.

It is informative to express the selection function in terms of black hole mass and redshift. The selection function as a function of black hole mass and redshift is the probability of including a source, given its M_{BH} and z, and is calculated as

$$p(I = 1|M_{BH}, z) = \int_0^\infty s(L_\lambda, z) p(L_\lambda|M_{BH}, z) \int_{v_{min}}^{v_{max}} p(\mathbf{v}|L_\lambda, M_{BH}) \, dL_\lambda \, d\mathbf{v}. \tag{5.14}$$

At any given value of M_{BH} a range of luminosities and emission line widths are possible, and thus sources with low black hole mass can be detected if they are bright enough and have line widths $v_{min} < v < v_{max}$. Conversely, sources with high black hole masses can be missed by the survey if their luminosity is below the flux limit at that redshift, or if their line width falls outside of the detectable range. This has the effect of smoothing the survey's selection function, and thus

the black hole mass selection function is a broadened form of the flux selection function.

As an example, consider the case when the selection function is simply a flux limit. In this case, the selection function is

$$
s(l, z) = \begin{cases} 1 \text{ if } 4\pi f_{min} D_L^2(z) < L_\lambda < 4\pi f_{max} D_L^2(z) \\ 0 \text{ otherwise} \end{cases}, \qquad (5.15)
$$

where f_{min} is the survey's lower flux limit, f_{max} is the survey's upper flux limit, and $D_L(z)$ is the luminosity distance to redshift z. For simplicity, in this example we assume that there is no additional cut on emission line width, i.e., $v_{min} = 0$ and $v_{max} = \infty$. In this case, the black hole mass selection function, $p(I = 1|M_{BH}, z)$, is the convolution of the luminosity selection function with the distribution of L_λ at a given M_{BH}. If the distribution of $\log L_\lambda$ at a given M_{BH} is a Gaussian function with mean $\alpha_0 + \alpha_m \log M_{BH}$ and dispersion σ_l, then the black hole mass selection function is

$$
\begin{aligned}
p(I = 1|M_{BH}, z) &= \Phi\left(\frac{\log L_{max}(z) - \alpha_0 - \alpha_m \log M_{BH}}{\sigma_l}\right) \\
&\quad - \Phi\left(\frac{\log L_{min}(z) - \alpha_0 - \alpha_m \log M_{BH}}{\sigma_l}\right). \qquad (5.16)
\end{aligned}
$$

Here, $L_{max}(z) = 4\pi f_{max} D_L^2(z)$, $L_{min}(z) = 4\pi f_{min} D_L^2(z)$, and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal density.

In Figure 5.1 we show the black hole mass selection function, $p(I = 1|M_{BH}, z)$, given by Equation (5.13) at $z = 1$. Here, we have used the SDSS quasar sample flux limit, $19.1 > i > 15$, $\alpha_0 = 37, \alpha_m = 1$, and $\sigma_l = 0.6$ dex. Because the black hole mass selection function is the convolution of the luminosity selection function with the distribution of L_λ at a given M_{BH}, the black hole mass selection function is positive over a wider range in M_{BH}, as compared to the range in
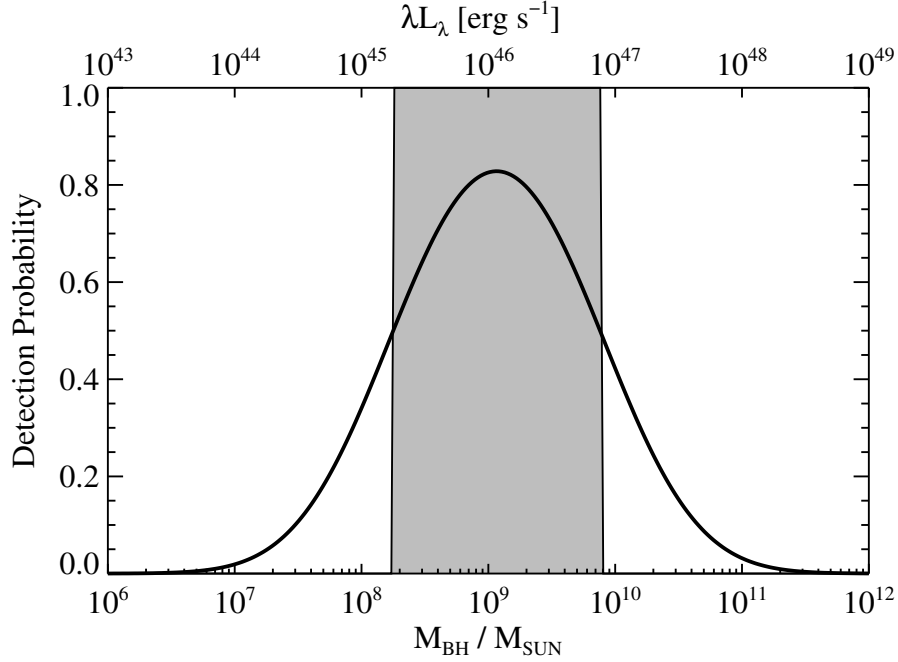
Figure 5.1 Comparison of the selection function for luminosity (shaded region) and black hole mass (curve) for a simple upper and lower flux limit. The selection is complete in luminosity within the flux limits, but is never 'complete' in M_{BH}. The intrinsic physical scatter in luminosity at a given black hole mass creates a more complicated selection function for M_{BH}, since at any given M_{BH} and z only those quasars with luminosities within the flux limits are detected.

L_λ for which $s(L_\lambda, z)$ is positive. However, because $p(I = 1 | M_{BH}, z)$ spreads the selection probability over a wider range in M_{BH}, all bins in M_{BH} are incomplete.

The difference in selection functions for black hole mass and luminosity results in an important distinction between the estimation of the black hole mass functions and the estimation of luminosity functions. First, one cannot correct the binned BHMF for the survey flux limits by simply applying the $1/V_a$ correction. This is a common technique used for estimating binned luminosity functions,

where the number density in a (L_λ, z) bin is corrected using the survey volume in which a source with luminosity L_λ could have been detected and still remained in the redshift bin. In the case of the BHMF, a survey volume in which the black hole could have been detected ceases to have any meaning, as black holes can be detected over many different survey volumes, albeit with varying probability. Alternatively, the $1/V_a$ correction can be thought of as dividing the number of sources in a bin in (L_λ, z) by the detection probability as a function of L_λ and z. Therefore, applying a $1/V_a$ correction to a bin in (M_{BH}, z) is essentially the same as dividing the number of sources in a bin in (M_{BH}, z) by the detection probability as a function of L_λ and z. For the simple example shown in Figure 5.1, those quasars in a given bin in (M_{BH}, z) that happen to have luminosities $L_{min}(z) < L_\lambda < L_{max}(z)$ will receive no correction, since $s(L_\lambda, z) = 1$. However, those quasars which have luminosities outside of the detectable range will not be detected. The end result is a systematic underestimate of the binned BHMF.

The number of sources in a given bin in the M_{BH}–z plane can be estimated by dividing the observed number of black holes in each bin by the black hole mass selection function, $p(I = 1|M_{BH}, z)$. Similarly, one can use a $1/V_a$-type correction by calculating an 'effective' $1/V_a$, found by integrating dV/dz over the black hole mass selection function. This approach has been adopted previously within the context of binned luminosity functions (e.g., Warren et al., 1994; Fan et al., 2001). However, it is essential that the black hole mass selection function be used and *not* the luminosity selection function. Unfortunately, this implies that one must assume a form for $p(L_\lambda|M_{BH}, z)$.

5.4   POSTERIOR DISTRIBUTION FOR THE BHMF PARAMETERS

The posterior probability distribution of the model parameters is

$$p(\theta, N | \mathbf{v}_{obs}, L_{obs}, z_{obs}, I) \propto p(\theta, N) p(\mathbf{v}_{obs}, L_{obs}, z_{obs}, I | \theta, N), \qquad (5.17)$$

where $p(\theta, N)$ is the prior on (θ, N), and $p(\mathbf{v}_{obs}, L_{obs}, z_{obs}, I | \theta, N)$ is the likelihood function, given by Equation (5.5). The posterior distribution gives the probability that θ and N have a given value, given the observed data $(\mathbf{v}_{obs}, L_{obs}, z_{obs})$. Therefore, the posterior distribution of θ and N can be used to obtain the probability that $\phi(M_{BH}, z)$ has any given value, given that we have observed some set of emission line widths, luminosities, and redshifts.

It is of use to decompose the posterior as $p(N, \theta | x_{obs}) \propto p(N | \theta, x_{obs}) p(\theta | x_{obs})$, where we have abbreviated the observed data as $x_{obs} = (\mathbf{v}_{obs}, L_{obs}, z_{obs})$. This decomposition seperates the posterior into the conditional posterior of the BHMF normalization, $p(N | x_{obs}, \theta)$, from the marginal posterior of the BHMF shape, $p(\theta | x_{obs})$. In this work we take N and θ to be independent in their prior distribution, $p(N, \theta) = p(N) p(\theta)$, and that the prior on N is uniform over $\log N$. In this case, one case show (e.g., Gelman et al., 2004, KFV08) that the marginal posterior distribution of θ is

$$p(\theta | \mathbf{v}_{obs}, L_{obs}, z_{obs}) \propto p(\theta) \left[p(I = 1 | \theta) \right]^{-n} \prod_{i \in \mathcal{A}_{obs}} p(\mathbf{v}_i, L_{\lambda,i}, z_i | \theta), \qquad (5.18)$$

where $p(I = 1 | \theta) = 1 - p(I = 0 | \theta)$.

Under the prior $p(\log N) \propto 1$, the conditional posterior of $N | \theta, x_{obs}$ is a negative binomial distribution with parameters n and $p(I = 1 | \theta)$. The negative binomial distribution gives the probability that the total number of sources is equal to N, given that there have been n detections with probability of detection $p(I = 1 | \theta)$:

$$p(N | n, \theta) = C_{n-1}^{N-1} \left[p(I = 1 | \theta) \right]^n \left[p(I = 0 | \theta) \right]^{N-n}. \qquad (5.19)$$

Because of the large number of parameters in the model, Bayesian inference is most easily performed by randomly drawing values of N and θ from their posterior. Based on the decomposition $p(\theta, N|x_{obs}) \propto p(N|n, \theta)p(\theta|x_{obs})$, we can obtain random draws of (N, θ) by first drawing values of θ from Equation (5.15). Then, for each draw of θ, we draw a value of N from the negative binomial distribution. Random draws for θ may be obtained via markov chain monte carlo (MCMC) methods, described in § 5.6, and random draws from the negative binomial distribution are easily obtained using standard methods (e.g., Gelman et al., 2004, KFV08).

5.5   THE STATISTICAL MODEL

In order to compute the likelihood function for the observed set of luminosities, redshifts, and broad emission line widths (see Eq.[5.9]), it is necessary to relate the BHMF to the distribution of L_λ and $\mathbf{v}$. To do this, Equation (5.8) implies that we need three terms. The first term is an assumed BHMF, $p(M_{BH}, z) = N^{-1}(dV/dz)^{-1}\phi(M_{BH}, z)$. The second term is an assumed distribution of luminosities at a given black hole mass and redshift, $p(L_\lambda|M_{BH}, z)$. The third term is an assumed distribution of broad emission line widths at a given luminosity and black hole mass, $p(\mathbf{v}|L_\lambda, M_{BH})$. Once we have a parameteric form for each of these three distributions, we can calculate the observed data likelihood directly from Equation (5.9). In this section we describe parameteric forms for each of these distributions based on a mixture of Gaussian functions model.

5.5.1   Mixture of Gaussian Functions Model for the BHMF

The mixture of Gaussian functions model is a common 'non-parameteric' model that allows flexibility when estimating the BHMF. The basic idea is that one can use a suitably large enough number of Gaussian functions to accurately approx-

imate the true BHMF, even though the individual Gaussian functions have no physical meaning. Furthermore, the Gaussian mixture model is also conjugate to the distributions $p(L_\lambda|m)$ and $p(\mathbf{v}|L_\lambda, m)$ assumed in §§5.5.2 and 5.5.3, thus enabling us to calculate some of the integrals in Equation (5.15) analytically.

In KFV08 we described a mixture of Gaussian functions model for a luminosity function. The mixture of Gaussian functions model of the BHMF is identical to that for the luminosity function, after replacing L with M_{BH}. Our mixture of Guassian functions model, including our adopted prior, is described in KFV08; for completeness we briefly review it here.

The mixture of K Gaussian functions model for the BHMF is

$$p(\log M_{BH}, \log z|\pi, \mu, \Sigma) = \sum_{k=1}^{K} \frac{\pi_k}{2\pi|\Sigma_k|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mu_k)^T \Sigma_k^{-1}(\mathbf{y} - \mu_k)\right], \quad (5.20)$$

where $\sum_{k=1}^{K} \pi_k = 1$. Here, $\mathbf{y} = (\log M_{BH}, \log z)$, μ_k is the 2-element mean vector for the kth Gaussian, Σ_k is the 2×2 covariance matrix for the kth Gaussian, and x^T denotes the transpose of x. In addition, we denote $\pi = (\pi_1, \ldots, \pi_K), \mu = (\mu_1, \ldots, \mu_K)$, and $\Sigma = (\Sigma_1, \ldots, \Sigma_K)$. The variance in m for Gaussian k is $\sigma_{m,k}^2 = \Sigma_{11,k}$, the variance in $\log z$ for Gaussian k is $\sigma_{z,k}^2 = \Sigma_{22,k}$, and the covariance between $\log M_{BH}$ and $\log z$ for Gaussian k is $\sigma_{mz,k} = \Sigma_{12,k}$. Note that Equation(5.17) is equivalent to assuming that $p(M_{BH}, z)$ is a mixture of log-normal densities. Under the mixture model, the BHMF can be calculated from Equations (5.1) and (5.17). Noting that $p(M_{BH}, z) = p(\log M_{BH}, \log z)/(M_{BH}z(\ln 10)^2)$, the mixture of normals model for the BHMF is

$$\phi(M_{BH}, z|\theta, N) = \frac{N}{M_{BH}z(\ln 10)^2}\left(\frac{dV}{dz}\right)^{-1} \sum_{k=1}^{K} \frac{\pi_k}{2\pi|\Sigma_k|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mu_k)^T \Sigma_k^{-1}(\mathbf{y} - \mu_k)\right],$$
$$(5.21)$$

where, as before, $\mathbf{y} = (\log M_{BH}, \log z)$.

5.5.2 The Distribution of L_λ at a Given M_{BH}

We model the distribution of luminosities at a given M_{BH} as a log-normal density, where the average $\log L_\lambda$ at a given M_{BH} depends linearly on $\log M_{BH}$:

$$p(\log L_\lambda | M_{BH}, \alpha) = \frac{1}{\sqrt{2\pi\sigma_l^2}} \exp\left[-\frac{1}{2}\left(\frac{\log L_\lambda - \alpha_0 - \alpha_m \log M_{BH}}{\sigma_l}\right)^2\right]. \qquad (5.22)$$

Here, the unknown parameters are $\alpha = (\alpha_0, \alpha_m, \sigma_l^2)$. This is equivalent to assuming a simple linear regression of $\log L_\lambda$ on $\log M_{BH}$, where α_0 is the constant, α_m is the slope, and σ_l is the standard deviation of the random Gaussian scatter about the regression line. We assume a uniform prior on these parameters, i.e., $p(\alpha_0, \alpha_m, \sigma_l) \propto 1$.

The form of the M_{BH}–L_λ relationship given by Equation (5.19) is motivated by noting that L_λ can be related to M_{BH} as

$$\lambda L_\lambda = 1.3 \times 10^{38}\frac{\Gamma_{Edd}}{C_\lambda}\frac{M_{BH}}{M_\odot} \ \ [\text{erg s}^{-1}], \qquad (5.23)$$

where $\Gamma_{Edd} \equiv L_{bol}/L_{Edd}$ is the Eddington ratio, and C_λ is the bolometric correction to λL_λ. Equation (5.20) implies that the scatter in luminosities at a given black hole mass is caused by the scatter in Eddington ratios and bolometric corrections at a given black hole mass. The distribution of the scatter in $\log L_\lambda$ at a given M_{BH} is the convolution of the distribution of $\log \Gamma_{Edd}$ at a given M_{BH}, with the distribution of $\log C_\lambda$ at a given M_{BH}. The parameter σ_l is thus an estimate of the dispersion in $\log(\Gamma_{Edd}/C_\lambda)$ at a given M_{BH}.

If both Γ_{Edd} and C_λ are statistically independent of M_{BH}, then we would expect that on average $L_\lambda \propto M_{BH}$, i.e., $\alpha_m = 1$. However, if Γ_{Edd} or C_λ are correlated with M_{BH}, then $\alpha_m \neq 1$. Currently, it is unknown whether M_{BH} and Γ_{Edd} are correlated. However, it is likely that quasar SEDs depend on both Γ_{Edd} and M_{BH}, and therefore the bolometric correction will also depend on Γ_{Edd} and M_{BH}. Indeed, recently some authors have found evidence that the bolometric correction

depends on Eddington ratio (Vasudevan & Fabian, 2007) and black hole mass (Kelly et al., 2008). Therefore, it is likely that $\alpha_m \neq 1$, and we therefore leave it as a free parameter. In addition, comparison of Equation (5.19) with Equation (5.20) implies that the average value of Γ_{Edd}/C_λ is related to α_0 according to $E(\log \Gamma_{Edd}/C_\lambda) = \alpha_0 - 38.11$, where $E(x)$ denotes the expectation value of x. Therefore, one can use α_0 to estimate the typical broad line quasar Eddington ratio, assuming a typical bolometric correction.

Currently, there is little known about the distribution of luminosities at a given black hole mass, so for simplicity we assume the simple linear form given by Equation (5.19). Furthermore, the assumption of Gaussian scatter in $\log L$ at a given M_{BH} is consistent with the L–M_{BH} relationship for those AGN with reverberation mapping data (Kelly & Bechtold, 2007). More sophisticated models could include a non-linear dependence on $\log M_{BH}$, an additional redshift dependence, or non-Gaussian scatter. Unfortunately, this introduces additional complexity into the model. Furthermore, an additional redshift dependence in Equation (5.19) implies that the distribution of Γ_{Edd} or C_λ at a given M_{BH} evolves. However, currently most investigations have not found any evidence for significant evolution in Γ_{Edd} (e.g., Vestergaard, 2004; Kollmeier et al., 2006), and it is unclear if the quasar SED evolves at a given M_{BH}. Therefore, there is currently no compelling evidence to justify inclusion of a redshift dependence in Equation (5.19). In addition, we note that it is impossible to use $p(L|M_{BH})$ to infer the distribution of Eddington ratios without making an assumption about the distribution of C_{bol}, as Equation (5.20) shows that Γ_{Edd} and C_{bol} are degenerate. While estimating the distribution of Γ_{Edd} is of significant interest, it is beyond the scope of this work to develop a robust technique to do so, as our goal is to estimate the black hole mass function.

Because of the large number of parameters, large uncertainty in the broad line black hole mass estimates, and flux limit, estimating the BHMF is already a difficult statistical problem. As such, our approach is to initially assume the simple form given by Equation (5.19) in order to keep the degrees of freedom low, and to check if this assumption is consistent with our data (see § 5.7.3). If it is found that the observed data are inconsistent with this statistical model (e.g., see § 5.7.3) then Equation (5.19) should be modified.

5.5.3 The Distribution of v at a given L and M_{BH}

Following Kelly & Bechtold (2007), we can derive the distribution of emission line widths at a given luminosity and black hole mass. Given an AGN luminosity, L_λ^{BL}, the BLR distance R is assumed to be set by the luminosity according to the R–L relationship, $R \propto L_\lambda^{\beta_l}$, with some additional log-normal scatter:

$$p(\log R | L_\lambda^{BL}) = \frac{1}{\sqrt{2\pi\sigma_r^2}} \exp\left[-\frac{1}{2} \left(\frac{\log R - r_0 - \beta_l \log L_\lambda^{BL}}{\sigma_r} \right)^2 \right]. \qquad (5.24)$$

Here, r_0 is a constant, σ_r is the dispersion in $\log R$ at a given luminosity, and L_λ^{BL} is the AGN continuum luminosity at some reference wavelength appropriate for the broad emission line of interest. Note that the reference wavelength for L_λ^{BL} is not necessarily the same wavelength as for L_λ used in § 5.5.2. In particular, the wavelength for L_λ used in the M_{BH}–L_λ relationship should be chosen to adequately account for the selection function, while the reference wavelength for L_λ^{BL} should be appropriate for describing the R–L relationship. Since AGN continua are well described by a power-law, $f_\nu \propto \nu^{-\alpha}$, it should be easy to calculate L_λ at different values so long as the spectral index, α, is known. The intrinsic scatter in R at a given L_λ^{BL} is likely due to variations in quasar SED, reddening, non-instantaneous response of the BLR to continuum variations, etc.

Assuming that the BLR gas is gravitationally bound, the velocity dispersion

of the broad line-emitting gas is related to R and M_{BH} as $M_{BH} = fRv^2/G$. Here, G is the gravitational constant, and f is a factor that converts the virial product, RM_{BH}/G, to a mass. We do not directly measure v, but instead estimate it by the $FWHM$ or dispersion of the broad emission line in a single-epoch spectra. As a result, the measured line width will scatter about the actual value of v, where this scatter may be due in part to variations in line profile shape and the existence of stationary components in the single-epoch line profile. In our statistical model we assume that this scatter is log-normal with a dispersion of σ_v. In addition, the value of f depends on the measure of line width used. Onken et al. (2004) estimated f by comparing black hole masses derived form reverberation mapping with those derived from the M_{BH}–σ relationship, and find that on average $f = 1.4 \pm 0.4$ when using the $FWHM$. This value is consistent with a value of $f = 0.75$ expected from a spherical BLR geometry (e.g., Netzer, 1990b).

Under our model, the distribution of emission line widths at a given BLR size and black hole mass is

$$\log v | R, M_{BH} = \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp\left\{ -\frac{1}{2} \left[\frac{\log v - v_0 - 1/2(\log f + \log R - \log M_{BH})}{\sigma_v} \right]^2 \right\}.$$
(5.25)

where v_0 is a constant. For convenience, here and throughout this paper we denote the estimate of the BLR gas velocity dispersion as v, i.e., v is either the $FWHM$ or dispersion of the emission line. The term v in Equation (5.22) should not be confused with the actual velocity dispersion of the BLR gas, but is an estimate of it based on a measure of the width of the broad emission line. From Equation (5.22) it is apparent that the term f shifts the distribution of $\log v$ by a constant amount, which has the effect of shifting the inferred BHMF by a constant amount in $\log M_{BH}$. Throughout the rest of this work we assume the value of $f = 1.4$ found by Onken et al. (2004).

The distribution of v at a given L and M_{BH} is obtained from Equations (5.21) and (5.22) by averaging the distribution of v at a given R and M_{BH} over the distribution of R at a given L_λ^{BL}:

$$
\begin{aligned}
p(\log v | L_\lambda^{BL}, M_{BH}, \beta) &= \int_{-\infty}^{\infty} p(\log v | \log R, M_{BH}, \beta) p(\log R | L_\lambda^{BL}, \beta)\, d \log R \qquad (5.26) \\
&= \frac{1}{\sqrt{2\pi \sigma_{BL}^2}} \\
&\times \exp\left\{ -\frac{1}{2} \left[\frac{\log v - \beta_0 - 1/2(\beta_l \log L_\lambda^{BL} - \log M_{BH})}{\sigma_{BL}} \right]^2 \right\}. \qquad (5.27)
\end{aligned}
$$

where β_0 is a constant, $\sigma_{BL}^2 = \sigma_v^2 + \sigma_r^2/4$, and $\beta \equiv (\beta_0, \beta_l, \sigma_{BL})$. Note that in Equation (5.24) we have absorbed $\log f$ into the constant term, β_0. We introduce β into Equation (5.24) to make it explicit that Equation (5.24) depends on β. The term σ_{BL} is the dispersion in emission line widths at a given luminosity and black hole mass, and can be related to the intrinsic uncertainty in the broad line estimates of M_{BH}. The usual broad line mass estimates of AGN can be obtained by inverting the mean of Equation (5.24) as $\log \hat{M}_{BL} = \beta_l \log L_\lambda^{BL} + 2 \log v - 2\beta_0$, or equivalently $\hat{M}_{BL} \propto L_{\lambda,BL}^{\beta_l} v^2$. The intrinsic uncertainty on the broad line mass estimates is set by a combination of the intrinsic scatter in the R–L relationship and the uncertainty in using the single-epoch line width as an estimate of the broad line gas velocity dispersion: $\sigma_{\hat{M}_{BL}} = 2\sigma_{BL}$. Equation (5.24) describes the statistical scatter in the broad line mass estimates, and does not account for any additional systematic errors (e.g., Krolik, 2001; Collin et al., 2006).

It is typically the case that one employs multiple emission lines to estimate M_{BH}, producing black hole mass estimates across a broad range of redshifts and luminosities. In our work, we use the Hβ, Mg II, and C IV emission lines. In order to facilitate the use of different emission lines in the BHMF estimation, we introduce an indicator variable denoted by δ. Here, $\delta_{\mathrm{H}\beta} = 1$ if the Hβ line width is available, and $\delta_{\mathrm{H}\beta} = 0$ if the Hβ line widths is not available; δ_{MgII} and δ_{CIV} are

defined in an equivalent manner. For example, if one is using optical spectra, then at $z = 0.4$ only the Hβ emission line is available, and therefore $\delta_{\mathrm{H}\beta} = 1, \delta_{MgII} = 0$, and $\delta_{CIV} = 0$.

Assuming that the line width distributions for each line are independent at a given luminosity and black hole mass, then the observed distribution of line widths is the product of Equation (5.24) for each individual emission line:

$$
\begin{aligned}
p(\log \mathbf{v}|L, M_{BH}, z, \beta) &= \left[N(\log v_{\mathrm{H}\beta}|\bar{v}_{\mathrm{H}\beta}, \sigma_{\mathrm{H}\beta}^2)\right]^{\delta_{\mathrm{H}\beta}} \left[N(\log v_{\mathrm{MgII}}|\bar{v}_{\mathrm{MgII}}, \sigma_{\mathrm{MgII}}^2)\right]^{\delta_{\mathrm{MgII}}} \\
&\quad \times \left[N(\log v_{\mathrm{CIV}}|\bar{v}_{\mathrm{CIV}}, \sigma_{\mathrm{CIV}}^2)\right]^{\delta_{\mathrm{CIV}}} \tag{5.28}
\end{aligned}
$$

Here, the average line width for Hβ is $\bar{v}_{\mathrm{H}\beta} = \beta_0^{\mathrm{H}\beta} - (1/2)\beta_l^{\mathrm{H}\beta} \log L_\lambda^{\mathrm{H}\beta} + (1/2)\log M_{BH}$, and likewise for Mg II and C IV. Here, $L_\lambda^{\mathrm{H}\beta}$ denotes the value of L_λ that is used to calibrate the broad line mass estimates for Hβ, typically $L_\lambda(5100\text{Å})$.

Vestergaard & Peterson (2006) give equations for calculating broad line mass estimates from Hβ and C IV, derived from the most recent reverberation mapping data (Peterson et al., 2004; Kaspi et al., 2005), and Vestergaard et al. (2008, in progress) give an equation for calculating a broad line mass estimate from Mg II. These mass scaling relationships are:

$$
\log \hat{M}_{\mathrm{H}\beta} = -21.09 + 0.50 \log \lambda L_\lambda(5100\text{Å}) + 2 \log FWHM_{\mathrm{H}\beta} \tag{5.29}
$$

$$
\log \hat{M}_{\mathrm{MgII}} = -21.21 + 0.50 \log \lambda L_\lambda(2100\text{Å}) + 2 \log FWHM_{\mathrm{MgII}} \tag{5.30}
$$

$$
\log \hat{M}_{\mathrm{CIV}} = -22.66 + 0.53 \log \lambda L_\lambda(1350\text{Å}) + 2 \log FWHM_{\mathrm{CIV}} \tag{5.31}
$$

For the equations listed above we have used the $FWHM$ of the emission line as an estimate of the velocity dispersion, i.e., $v = FWHM$. Because $\log \hat{M}_{BL} = \beta_l \log \lambda L_\lambda^{BL} + 2 \log v - 2\beta_0$, it follows that $\beta_0^{\mathrm{H}\beta} = 10.55, \beta_0^{\mathrm{MgII}} = 10.61, \beta_0^{CIV} = 11.33$, and $\beta_l \approx 0.5$ for all three emission lines. In addition, Vestergaard & Peterson (2006) find the statistical scatter about M_{BH} in the broad line mass estimates to

be 0.43 dex and 0.36 dex for Hβ and C IV, respectively. Therefore, since $\sigma_{BL} = \sigma_{\hat{M}_{BL}}/2$, if follows that $\sigma_{H\beta} \approx 0.22$ and $\sigma_{CIV} \approx 0.18$ dex. Likewise, Vestergaard et al. (2008, in progress) find the intrinsic uncertainty in the broad line mass estimate for Mg II to be ~ 0.4 dex, and therefore $\sigma_{MgII} \approx 0.2$ dex.

Broad line mass estimates are now fairly well understood, and we derive our prior distribution for β from the scaling results of Vestergaard & Peterson (2006) and Vestergaard et al. (2008, in progress). We fix $\beta_l = 0.5, 0.5,$ and 0.53 for Hβ, Mg II, and C IV, respectively. However, in order to account for the uncertainty in these scaling relationships, we consider β_0 and σ_{BL} to be free parameters in our model. We cannot estimate the normalization and scatter in the broad line mass estimates solely from the distribution of $\mathbf{v}$, L, and z, since β_0 and σ_{BL} are degenerate with the other parameters. Therefore, it is necessary to place constraints on β_0 and σ_{BL} through a prior distribution. This allows us to constrain β_0 and σ_{BL} while still incorporating their uncertainty. The parameters for the prior distribution of β_0 and σ_{BL} are based on the uncertainty in the scaling relationships of Vestergaard & Peterson (2006) and Vestergaard et al. (2008, in progress). Our prior for β_0 are independent Gaussian distributions with means equal to 10.55, 10.61, and 11.33 for Hβ, Mg II, and C IV, respectively, and standard deviations equal to 0.1. To allow greater flexibility in our model, we chose the prior standard deviation of 0.1 to be wider than the formal uncertainty on the scaling factors of ≈ 0.02 reported by Vestergaard & Peterson (2006). For each emission line, our prior for σ_{BL} is a scaled inverse-χ^2 distribution with $\nu = 25$ degrees of freedom and scale parameter equal to 0.2 dex. We chose $\nu = 25$ degrees of freedom because approximately 25 AGN were used to derive the scaling relationships in Vestergaard & Peterson (2006). The values of β_0 were constrained to be within ± 0.3 (i.e., $\pm 3\sigma$) of the values reported by Vestergaard & Peterson (2006) and Vestergaard et al.

(2008, in progress), and the values of σ_{BL} were constrained to be within the inverval containing 99% of the probability for the scaled inverse-χ^2 distribution. By placing these constraints on β_0 and σ_{BL}, we ensure that their values remain consistent with the results derived from reverberation mapping.

5.5.4 Likelihood function for Mixture of Gaussians Model

Now that we have formulated the conditional distributions, we can calculate the likelihood function for the mixture of Gaussians model of $\phi(M_{BH}, z)$. Comparison with Equation (5.15) suggests that we need two terms: $p(\mathbf{v}_i, L_{\lambda,i}, z_i | \theta)$ and $p(I = 1 | \theta)$. The first term is the joint distribution of line widths, luminosities, and redshifts:

$$p(\mathbf{v}_i, L_{\lambda,i}, z_i | \theta) = \int p(\mathbf{v}_i | L_{\lambda,i}, M_{BH,i}, \beta) p(L_{\lambda,i} | M_{BH,i}, \alpha) p(M_{BH,i}, z_i | \pi, \mu, \Sigma) \, dM_{BH,i},$$

$$(5.32)$$

where $\theta = (\alpha, \beta, \pi, \mu, \Sigma)$.

The integral in Equation (5.29) can be done analytically by plugging Equations (5.17), (5.19), and (5.25) into Equation (5.29). However, the result depends on the number of emission lines used for the ith source. Expressing the likelihood function for a single emission line in terms of logarithms, $p(\log v_i, \log L_{\lambda,i}, \log z_i | \theta)$ is a mixture of K 3-dimensional Gaussian functions:

$$p(\log v_i, \log L_{\lambda,i}, \log z_i | \theta) = \sum_{k=1}^{K} \frac{\pi_k}{\sqrt{8\pi^3 |V_k|}} \exp\left\{-\frac{1}{2}(\mathbf{x}_i - \xi_k)^T V_k^{-1}(\mathbf{x}_i - \xi_k)\right\} \quad (5.33)$$

$$\mathbf{x}_i = (\log v_i, \log L_{\lambda,i}, \log z_i) \quad (5.34)$$

$$\xi_k = (\bar{v}_k, \bar{l}_k, \mu_{z,k}) \quad (5.35)$$

$$\bar{l}_k = \alpha_0 + \alpha_m \mu_{m,k} \quad (5.36)$$

$$\bar{v}_k = \beta_0 - \frac{1}{2}\beta_l \bar{l}_{BL,k} + \frac{1}{2}\mu_{m,k} \quad (5.37)$$

$$\bar{l}_{BL,k} = \bar{l}_k + (1d + \alpha_\lambda) \log\left(\frac{\lambda_{BL}}{\lambda_{ML}}\right) \quad (5.38)$$

$$Var(\log v|k) = \sigma_{BL}^2 + \frac{1}{4}\left[\beta_l^2 Var(\log l|k) + (1-\alpha_m)\sigma_{m,k}^2\right] \quad (5.39)$$

$$Var(\log l|k) = \sigma_l^2 + \alpha_m^2\sigma_{m,k}^2 \quad (5.40)$$

$$Cov(\log v, \log l|k) = \frac{1}{2}\alpha_m\sigma_{m,k}^2 - \frac{1}{2}\beta_l Var(\log l|k) \quad (5.41)$$

$$Cov(\log v, \log z|k) = \left(\beta_l\alpha_m + \frac{1}{2}\right)\sigma_{mz,k}, \quad (5.42)$$

where the covariance matrix of $(\log v_i, \log L_{\lambda,i}, \log z_i)$ for the kth Gaussian is

$$V_k = \begin{pmatrix} Var(\log v|k) & Cov(\log v, \log l|k) & Cov(\log v, \log z|k) \\ Cov(\log v, \log l|k) & Var(\log l|k) & \alpha_m\sigma_{mz,k} \\ Cov(\log v, \log z|k) & \alpha_m\sigma_{mz,k} & \sigma_{z,k}^2 \end{pmatrix} \quad (5.43)$$

Here, ξ_k is the mean vector of $(\log v_i, \log L_{\lambda,i}, \log z_i)$ for the kth Gaussian. In addition, $\bar{l}_k$ is the mean $\log L_\lambda$ for Gaussian k, $\bar{v}_k$ is the mean v for Gaussian k, $\bar{l}_{BL}$ is the mean $\log L_\lambda^{BL}$ for Gaussian k, $Var(\log v|k)$ is the variance in $\log v$ for Gaussian k, $Var(\log L_\lambda|k)$ is the variance in $\log L_\lambda$ for Gaussian k, $Cov(\log v, \log L_\lambda|k)$ is the covariance between $\log v$ and $\log L_\lambda$ for Gaussian k, and $Cov(\log v, \log z|k)$ is the covariance between $\log v$ and $\log z$ for Gaussian k; note that $\alpha_m\sigma_{mz,k}$ is the covariance between $\log L_\lambda$ and z for Gaussian k. The mean $\log L_\lambda^{BL}$ for Gaussian k is calculated from $\bar{l}_k$ assuming a power-law continuum of the form $L_\lambda^{BL} = L_\lambda(\lambda_{BL}/\lambda_{ML})^{\alpha_\lambda}$, where λ_{BL} is the wavelength used in the R–L_λ^{BL} relationship for the emission line of interest, and λ_{ML} is the wavelength that the M_{BH}–L_λ is formulated in. For example, $\lambda_{BL} = 5100\text{Å}$ for the Hβ-based mass scaling relationship of Vestergaard & Peterson (2006), and λ_{ML} may be, say, equal to 2500Å. Note that we are assuming that α_λ is known.

In Equation (5.30) it should be understood that v_i, β_0, β_l, and σ_{BL}^2 correspond to the particular emission line being used. For example, if one is using the C IV line width for the ith source, then $v_i = v_{CIV,i}, \beta_0 = \beta_0^{CIV}, \beta_l = \beta_l^{CIV}$, and $\sigma_{BL}^2 = \sigma_{CIV}^2$.

If there are two emission line widths available for the ith AGN, then $p(\mathbf{v}_i, L_{\lambda,i}, z_i|\theta)$ is a mixture of K 4-dimensional Gaussian functions:

$$p(\log \mathbf{v}_i, \log L_{\lambda,i}, z_i|\theta) = \sum_{k=1}^{K} \frac{\pi_k}{\sqrt{16\pi^4|V_k|}} \exp\left\{-\frac{1}{2}(\mathbf{x}_i - \xi_k)^T V_k^{-1}(\mathbf{x}_i - \xi_k)\right\} \tag{5.44}$$

$$\mathbf{x}_i = (\log v_{1,i}, \log v_{2,i}, \log L_{\lambda,i}, \log z_i) \tag{5.45}$$

$$\xi_k = (\bar{v}_{1,k}, \bar{v}_{2,k}, \bar{l}_k, \mu_{z,k}) \tag{5.46}$$

$$Cov(\log v_1, \log v_2|k) = \frac{1}{4}\left(\beta_{l,1}\beta_{l,2}Var(\log L_\lambda|k) + \sigma_m^2\right) \tag{5.47}$$

Here, $Cov(\log v_1, \log v_2|k)$ denotes the covariance between the logarithms of the two line widths, v_1 and v_2, for the kth gaussian. The 4×4 covariance matrix of $(\log \mathbf{v}_i, \log L_{\lambda,i}, \log z_i)$ is

$$V_k = \begin{pmatrix} Var(\log v_1|k) & Cov(\log v_1, \log v_2|k) & Cov(\log v_1, \log L_\lambda|k) & Cov(\log v_1, \log z|k) \\ Cov(\log v_1, \log v_2|k) & Var(\log v_2|k) & Cov(\log v_2, \log L_\lambda|k) & Cov(\log v_2, \log z|k) \\ Cov(\log v_1, \log L_\lambda|k) & Cov(\log v_2, \log L_\lambda|k) & Var(\log L_\lambda|k) & \alpha_m\sigma_{mz,k} \\ Cov(\log v_1, \log z|k) & Cov(\log v_2, \log z|k) & \alpha_m\sigma_{mz,k} & \sigma_{z,k}^2 \end{pmatrix}.$$

$$\tag{5.48}$$

The other terms are given by Equations (5.34)–(5.40), where it should be understood that β_0, β_l, and σ_{BL}^2 correspond to the values appropriate for each emission line. For example, at $z \sim 0.6$ both Hβ and Mg II are observable in the optical spectral region, and thus it is possible to have line widths for both emission lines. In this case, $v_{1,i}$ is the logarithm of the Hβ width for the ith source, $v_{2,i}$ is the logarithm of the Mg II width for the ith source, $\beta_{l,1}$ corresponds to β_l for the Hβ line, and $\beta_{l,2}$ corresponds to β_l for the Mg II line. The labeling of the Hβ line width as v_1 is irrelevant, and the same result would be obtained if we had labeled the Hβ line width as v_2.

It should be noted that in Equation (5.41) we have made the assumption that if *at least* one emission line has $v_{min} < v < v_{max}$, then v is estimated for all emission

lines in the observable spectral range at that redshift. If this is not the case, then Equation (5.41) must be integrated over $v_{1,i}$ or $v_{2,i}$ if either of $v_{1,i}$ or $v_{2,i}$ fall outside of (v_{min}, v_{max}).

The term $p(I = 1|\theta)$ is the probability that a source is included in one's sample for a given set of model parameters θ. Under the mixture of Gaussians model, Equation (5.10) can be simplified, allowing more efficient calculation. However, as above, the actual functional form of $p(I = 0|\theta)$ depends on the number of emission lines used in broad line mass estimation. If only one emission line is used, then Equation (5.10) becomes

$$p(I = 1|\theta) = \int_{-\infty}^{\infty} \int_{z_{min}}^{z_{max}} \frac{s(L_\lambda, z)}{z \ln 10}$$
$$\times \sum_{k=1}^{K} \pi_k f_{\mathbf{v}}(L_\lambda, z, \theta, k) N_2(\mathbf{y}_{lz}|\bar{\mathbf{y}}_{lz,k}, V_{lz,k}) \, dz \, dL_\lambda \tag{5.49}$$
$$\mathbf{y}_{lz} = (\log L_\lambda, \log z) \tag{5.50}$$
$$\bar{\mathbf{y}}_{lz,k} = (\bar{l}_k, \mu_{z,k}) \tag{5.51}$$
$$V_{lz,k} = \begin{pmatrix} Var(\log L_\lambda|k) & \alpha_m \sigma_{mz,k} \\ \alpha_m \sigma_{mz,k} & \sigma_{z,k}^2 \end{pmatrix}. \tag{5.52}$$

The term $f_{\mathbf{v}}(L_\lambda, z, \theta, k)$ is the probability that a source has at least one line width between v_{min} and v_{max} for the kth Gaussian function, given its luminosity and redshift. For redshifts where only one emission line is used, $f_{\mathbf{v}}(L_\lambda, z, \theta, k) = Pr(v_{min} < v < v_{max}|L_\lambda, z, \theta, k)$, where

$$Pr(v_{min} < v < v_{max}|L_\lambda, z, \theta, k) = \Phi\left(\frac{\log v_{max} - E(\log v|L_\lambda, z, k)}{\sqrt{Var(\log v|L_\lambda, z, k)}}\right)$$
$$- \Phi\left(\frac{\log v_{min} - E(\log v|L_\lambda, z, k)}{\sqrt{Var(\log v|L_\lambda, z, k)}}\right) \tag{5.53}$$
$$E(\log v|l, z, k) = \bar{v}_k + \mathbf{c}_k^T V_{lz,k}^{-1} (\mathbf{y}_{lz} - \bar{\mathbf{y}}_{lz,k}) \tag{5.54}$$
$$Var(\log v|L_\lambda, z, k) = Var(\log v|k) - \mathbf{c}_k^T V_{lz,k}^{-1} \mathbf{c}_k^T \tag{5.55}$$
$$\mathbf{c}_k = [Cov(\log v, \log L_\lambda|k), Cov(\log v, \log z|k)] \tag{5.56}$$

Here, $\Phi(\cdot)$ is the cumulative distribution function of the standard normal density, $E(\log v | L_\lambda, z, k)$ is the mean of $\log v$ for the kth Gaussian at a given L_λ and z, $Var(\log v | L_\lambda, z, k)$ is the variance in $\log v$ for the kth Gaussian at a given L_λ and z, and $\mathbf{c}_k$ is a 2-dimensional vector containing the covariances between $\log v$ and both $\log L_\lambda$ and $\log z$. The standard normal cumulative distribution function can be efficiently computed using a look-up table, and therefore only two integrals need to be calculated numerically in Equation (5.46).

If one is using multiple emission lines for estimating $\phi(M_{BH}, z)$, then $f_{\mathbf{v}}(L_\lambda, z, \theta, k)$ must be modified to account for this. Equation (5.50) gives the probability that an emission line has a line width $v_{min} < v < v_{max}$, under the assumption that only one emission line is used at any given redshift. However, if there are redshifts where two emission lines are used, then $f_{\mathbf{v}}(L_\lambda, z, \theta, k)$ must be modified, as in these cases we need the probability that *at least* one emission line has $v_{min} < v < v_{max}$. At redshifts where two emission lines are used, $f_{\mathbf{v}}(L_\lambda, z, \theta, k)$ becomes the probability that either $v_{min} < v_1 < v_{max}$ or $v_{min} < v_2 < v_{max}$:

$$
\begin{aligned}
f_{\mathbf{v}}(L_\lambda, z, \theta, k) \ = \ & Pr(v_{min} < v_1 < v_{max} | L_\lambda, z, \theta, k) \\
+ \ & Pr(v_{min} < v_2 < v_{max} | L_\lambda, z, \theta, k) \\
- \ & Pr(v_{min} < v_1 < v_{max} | L_\lambda, z, \theta, k) \\
\times \ & Pr(v_{min} < v_2 < v_{max} | L_\lambda, z, \theta, k),
\end{aligned}
\tag{5.57}
$$

where $Pr(v_{min} < v_j < v_{max} | L_\lambda, z, \theta, k)$ are given by Equation (5.50) for $j = 1, 2$, respectively.

As an example, at $z \sim 0.2$ only the Hβ line is available in the optical spectral region, and thus, at this redshift, an optical survey can only employ the Hβ line for estimating the BHMF. In this case, $p(\log v_i, \log L_{\lambda,i}, \log z_i | \theta)$ is given by Equation (5.30), and $f_{\mathbf{v}}(L_\lambda, z, \theta, k)$ is given by Equation (5.50). However, at $z \sim 0.6$,

both Hβ and Mg II are observable in the optical spectral region, and thus both may be employed for estimating the BHMF. At this redshift, $p(\log \mathbf{v}_i, \log L_{\lambda,i}, \log z_i|\theta)$ is given by Equation (5.41), and $f_{\mathbf{v}}(L_\lambda, z, \theta, k)$ is given by Equation (5.55), where $\mathbf{v} = (v_1, v_2)$, v_1 is the Hβ line width, and v_2 is the Mg II line width (or vice versa). If only one emission line is available at any particular redshift, either because of limited spectral range or because of a choice on the part of the researcher to ignore certain emission lines, then only Equations (5.30) and (5.50) need be used.

The functional forms of $p(\mathbf{v}_i, L_{\lambda,i}, z_i|\theta)$ and $p(I = 1|\theta)$ given above can be plugged into Equation (5.5) to obtain the likelihood function for the mixture of normals model. A maximum-likelihood estimate of $\phi(M_{BH}, z)$ can be obtained by first maximizing Equation (5.5) with respect to N and $\theta = (\alpha_0, \alpha_m, \sigma_l^2, \beta_0, \beta_l, \sigma_{BL}^2, \pi, \mu, \Sigma)$. Then, using the maximum-likelihood estimate of (N, π, μ, Σ), the maximum-likelihood estimate of $\phi(M_{BH}, z)$ is calculated by using Equation (5.17) in Equation (5.1). Unfortunately, for $K > 1$ gaussians, maximizing the likelihood for the Gaussian mixture model is a notoriously difficult optimization problem. The maximization is probably most efficiently performed using the Expectation-Maximization (EM, Dempster, Laird, & Rubin, 1977) algorithm, or employing a stochastic search routine. Since we focus on Bayesian inference, a derivation of the EM algorithm for the BHMF is beyond the scope of this work.

The posterior distribution of θ and N can be calculated using the forms given above for $p(\log \mathbf{v}_i, \log L_{\lambda,i}, z_i|\theta)$ and $p(I = 1|\theta)$. In this case, one plugs the equations for $p(\log \mathbf{v}_i, \log L_{\lambda,i}, \log z_i|\theta)$ and $p(I = 1|\theta)$ for the Gaussian mixture model into Equations (5.15) and (5.16). The prior distribution, $p(\theta)$, is given by Equation (21) in KFV08.

5.5.5 Accounting for Measurement Error

The preceding discussion has assumed that $\mathbf{v}_i$ and $L_{\lambda,i}$ are known. However, in general, both quantities are measured with error. The effect of measurement error is to broaden the observed distributions of $\mathbf{v}_i$ and $L_{\lambda,i}$. Because the Bayesian approach attempts to define the set of BHMFs that are consistent with the observed distribution of $\mathbf{v}_i, L_{\lambda,i}$, and z_i, where 'consistency' is measured by the posterior probability distribution, measurement error can affect statistical inference on the BHMF. If the variance of the measurement errors on $\mathbf{v}_i$ and $L_{\lambda,i}$ are small compared to the intrinsic physical variance in these quantities, then measurement error does not have a significant effect on the results. In general, the measurement errors on $L_{\lambda,i}$ will likely be small compared to the real physical scatter in AGN luminosities, so we neglect them. This may not always be the case for the emission line widths, and in this section we modify the likelihood function for the mixture of Gaussian functions model to include measurement errors in $\mathbf{v}_i$. The general method of handling measurement errors within a Bayesian or likelihood function approach is described in Kelly (2007). For the sake of brevity, we omit the derivations and simply report the modifications to the likelihood function.

If one is only employing one emission line at a given redshift, then Equation (5.29) can be factored as

$$p(\log v_i, \log L_{\lambda,i}, \log z_i | \theta) = p(\log v_i | L_{\lambda,i}, z_i, \theta) p(\log L_{\lambda,i}, \log z_i | \theta). \qquad (5.58)$$

Under the mixture of Gaussian functions model, the joint distribution of luminosity and redshift is obtained from Equations (5.30)–(5.36) by simply omitting the terms that depend on v_i:

$$p(\log L_{\lambda,i}, \log z_i | \theta) = \sum_{k=1}^{K} \frac{\pi_k}{\sqrt{4\pi^2 |V_{lz,k}|}} \exp\left\{ -\frac{1}{2} (\mathbf{y}_{lz,i} - \bar{\mathbf{y}}_{lz,k})^T V_{lz,k}^{-1} (\mathbf{y}_{lz,i} - \bar{\mathbf{y}}_{lz,k}) \right\}.$$

$$(5.59)$$

Here, $\mathbf{y}_{lz,i} = (\log L_{\lambda,i}, \log z_i)$, $\bar{\mathbf{y}}_{lz,k}$ is given by Equation (5.48) and $V_{lz,k}$ is given by Equation (5.49). The distribution of the measured $\log v_i$ at $L_{\lambda,i}$ and z_i is

$$
\begin{aligned}
p(\log v_i | L_{\lambda,i}, z_i, \theta) &= \sum_{k=1}^{K} \frac{\pi_k}{\sqrt{2\pi[Var(\log v | L_{\lambda,i}, z_i, k) + \sigma_{v,i}^2]}} \\
&\quad \times \exp\left\{ -\frac{1}{2} \frac{(\log v_i - E(\log v | L_{\lambda,i}, z_i, k))^2}{Var(\log v | L_{\lambda,i}, z_i, k) + \sigma_{v,i}^2} \right\}.
\end{aligned}
\tag{5.60}
$$

Here, $\sigma_{v,i}^2$ is the variance of the measurement error on v_i, $E(\log v | L_{\lambda,i}, z_i, k)$ is given by Equation (5.51), and $Var(\log v | L_{\lambda,i}, z_i, k)$ is given by Equation (5.52). From Equation (5.58) the effect of measurement error on the line width becomes apparent: the distribution of line widths at a given luminosity and redshift is broadened by an amount dependent on the magnitude of the line width measurement error. If $\sigma_{v,i}^2 \ll Var(\log v | L_{\lambda,i}, z_i, k)$ then Equation (5.56) reduces to Equation (5.29). Otherwise, if measurement error on v_i is a concern, Equations (5.56)–(5.58) should be used for Equation (5.29) instead of Equation (5.30).

If one is employing two emission lines at a given redshift, then Equation (5.29) becomes

$$
p(\log \mathbf{v}_i, \log L_{\lambda,i}, \log z_i | \theta) = p(\log v_{1,i} | L_{\lambda,i}, z_i, \theta) p(\log v_{2,i} | L_{\lambda,i}, z_i, \theta) p(\log L_{\lambda,i}, \log z_i | \theta).
\tag{5.61}
$$

In this case, $p(\log v_{j,i} | L_{\lambda,i}, z_i, \theta), j = 1, 2$, must be calculated seperately for each emission line from Equation (5.58).

5.6 POSTERIOR DISTRIBUTION OF THE BHMF VIA MARKOV CHAIN MONTE CARLO

The number of free parameters in our statistical model is $6K + 8$, where K is the number of Gaussians used to approximate $\phi(\log M_{BH}, \log z)$. Because of the large number of free parameters, summarizing the posterior is most efficiently done by using Markov Chain Monte Carlo techniques to simulate random draws of θ and

N from the posterior distribution. In this work we use the Metropolis-Hastings algorithm (MHA, Metropolis & Ulam, 1949; Metropolis et al., 1953; Hastings, 1970) to perform the MCMC. We use the MHA to obtain a set of random draws from the marginal posterior distribution of θ, given by Equation (5.15). Then, given the values of θ, random draws for N may be obtained from the negative binomial distribution. A further description of the Metropolis-Hastings algorithm is given by KFV08, and our MHA is an extension of the MHA described in KFV08. For further details on the MHA see Chib & Greenberg (1995) or Gelman et al. (2004).

As in KFV08, we denote the current value of a parameter by placing a ~ over its symbol, and we denote the proposal value by placing a $*$ in the superscript. For example, if one were updating α_0, then $\tilde{\alpha}_0$ denotes the current value of α_0 in the random walk, α_0^* denotes the proposed value of α_0, $\tilde{\theta}$ denotes the current value of θ, and θ^* denotes the proposed value of θ, i.e., $\theta^* = (\alpha_0^*, \tilde{\alpha}_m, \tilde{\sigma}_l^2, \tilde{\beta}_0, \tilde{\sigma}_{BL}^2, \tilde{\pi}, \tilde{\mu}, \tilde{\Sigma}, \tilde{\mu}_0, \tilde{A}, \tilde{T})$. Here, μ_0, A and T are the parameters for the prior distribution on the mixture of Gaussian functions parameter (see KFV08). In addition, for ease of notation we define $x_{obs} = (\mathbf{v}_{obs}, L_{obs}, z_{obs})$ to be the set of observable quantities.

1. Start with initial guesses for $\alpha_0, \alpha_m, \sigma_l^2, \beta_0, \sigma_{BL}^2, \pi, \mu, \Sigma, \mu_0$, and A.

2. Draw a proposal value for α_0 and α_m from a 2-dimensional normal distribution centered at the current values of α_0 and α_m with set covariance matrix, Σ_α. The proposal values of α_0 and α_m are then simulated as $(\alpha_0^*, \alpha_m^*) \sim N_2([\tilde{\alpha}_0, \tilde{\alpha}_m], \Sigma_\alpha)$. If $p(\theta^*|x_{obs}) > p(\tilde{\theta}|x_{obs})$ then set $\tilde{\alpha}_0 = \alpha_0^*$ and $\tilde{\alpha}_m = \alpha_m^*$. Otherwise, calculate the ratio $r_\alpha = p(\theta^*|x_{obs})/p(\tilde{\theta}|x_{obs})$ and draw a random number uniformly distributed between 0 and 1, denoted as u. If $u < r_\alpha$ then set $\tilde{\alpha}_0 = \alpha_0^*$ and $\tilde{\alpha}_m = \alpha_m^*$. Otherwise, if $u > r_\alpha$, the values of $\tilde{\alpha}_0$ and $\tilde{\alpha}_m$ remain unchanged.

3. Draw a proposal value for $\log \sigma_l^2$ as $\log \tilde{\sigma}_l^2 \sim N(2\log \sigma_l^*, \sigma_{\sigma_l}^2)$, where $\sigma_{\sigma_l}^2$ is some set variance. Similar to before, calculate the ratio $r_\sigma = \sigma_l^* p(\theta^*|x_{obs})/\tilde{\sigma}_l p(\tilde{\theta}|x_{obs})$. Here, the term $\sigma_l^*/\tilde{\sigma}_l$ arises because the MHA acceptance rule must be corrected for the asymmetry in the log-normal jumping distribution used for σ_l^2. If $r_\sigma \geq 1$ then set $\tilde{\sigma}_l = \sigma_l^*$, otherwise set $\tilde{\sigma}_l = \sigma_l^*$ with probability r_σ. This is done by drawing a uniformly distributed random variable as in step 2.

4. Draw a proposal value for β_0 from a normal distribution centered at the current value of β_0 with set variance, σ_β^2. If $p(\theta^*|x_{obs}) > p(\tilde{\theta}|x_{obs})$ then set $\tilde{\beta}_0 = \beta_0^*$. Otherwise, calculate the ratio $r_\beta = p(\theta^*|x_{obs})/p(\tilde{\theta}|x_{obs})$ and draw a random number uniformly distributed between 0 and 1, denoted as u. If $u < r_\beta$ then set $\tilde{\beta}_0 = \beta_0^*$. Otherwise, if $u > r_\beta$, then the value of $\tilde{\beta}_0$ remain unchanged. If one is employing multiple emission lines to estimate the BHMF, then we have found it faster to simulate proposed values of β_0 for each emission line simultaneously from a multivariate normal density.

5. Draw a proposal value for $\log \sigma_{BL}^2$ as $\log \tilde{\sigma}_{BL}^2 \sim N(2\log \sigma_{BL}^*, \sigma_{\sigma_{BL}}^2)$, where $\sigma_{\sigma_{BL}}^2$ is some set variance. Similar to the update for σ_l^2, calculate the ratio $r_{BL} = \sigma_{BL}^* p(\theta^*|x_{obs})/\tilde{\sigma}_{BL} p(\tilde{\theta}|x_{obs})$. If $r_{BL} \geq 1$ then set $\tilde{\sigma}_{BL} = \sigma_{BL}^*$, otherwise set $\tilde{\sigma}_{BL} = \sigma_{BL}^*$ with probability r_σ. This is done by drawing a uniformly distributed random variable as in step 2. If one is employing multiple emission lines to estimate the BHMF, then we have found it faster to simulate proposed values of $\log \sigma_{BL}^2$ for each emission line simultaneously from a multivariate normal density.

6. Draw new values of the Gaussian mixture model parameters according to steps 2–6 in the MHA described in KFV08.

One then repeats steps 2–6 until the MCMC converges, saving the values of $\tilde{\theta}$

at each iteration. After convergence, the MCMC is stopped, and the values of $\tilde{\theta}$ may be treated as a random draw from the marginal posterior distribution of θ, $p(\theta|x_{obs})$. Techniques for monitering convergence of the Markov Chains can be found in Gelman et al. (2004). Given the values of θ obtained from the MCMC, one can then draw values of N from the negative binomial density (cf. Eq.[5.16]).

Having obtained random draws of N and θ from $p(\theta, N|\mathbf{v}_{obs}, L_{obs}, z_{obs})$, one can then use these values to calculate an estimate of $\phi(M_{BH}, z)$, and its corresponding uncertainty. This is done by plugging in each of the MCMC draws of θ and N directly into Equation (5.18). The posterior distribution of $\phi(M_{BH}, z)$ can be estimated for any value of M_{BH} and z by plotting a histogram of the values of $\phi(M_{BH}, z)$ obtained from the MCMC values of θ and N. KFV08 illustrates in more detail how to use the MHA results to perform statistical inference.

5.7 APPLICATION TO SIMULATED DATA

As an illustration of the effectiveness of our method, we applied it to a simulated data set. Because we will eventually apply this method to the BHMF for the SDSS DR3 quasar catalogue (Schneider et al., 2005), we assume the effective survey area and selection function reported for the DR3 quasar sample (Richards et al., 2006).

5.7.1 Construction of the Simulated Sample

We construct our simulated survey in a manner very similar to that used by KFV08. We first drew a random value of N_{Ω} quasars from a binomial distribution with probability of success $\Omega/4\pi = 0.0393$ and number of trials $N = 2 \times 10^5$. Here, $\Omega = 1622 \deg^2$ is the effective sky area for our simulated survey, and we chose the total number of quasars to be $N = 2 \times 10^5$ in order to produce a value of $n \sim 1000$ observed sources after including the flux limit. While this produces a much smaller sample than the actual sample of $\sim 1.5 \times 10^4$ quasars from the SDSS

DR3 luminosity function work (Richards et al., 2006), we chose to work with this smaller sample to illustrate the effectiveness of our method on more moderate sample sizes. This first step of drawing from a binomial distribution simulates a subset of N_Ω sources randomly falling within an area Ω on the sky, where the total number of sources is N. Note that we have not included any flux limits yet.

For each of these $N_\Omega \sim 8000$ sources, we simulated values of M_{BH} and z. We first simulated values of $\log z$ from a distribution of the form

$$f(\log z) = \frac{4\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\exp(a\zeta^*)}{(1+\exp(\zeta^*))^{a+b}}, \tag{5.62}$$

where $\zeta^* = 4(\log z - 0.4)$. The parameters $a = 2$ and $b = 3$ were chosen to give an observed redshift distribution similar to that seen for SDSS DR3 quasars (e.g., Richards et al., 2006).

For each simulated value of z, we simulated a value of M_{BH} by taking the distribution of M_{BH} at a given redshift to be a smoothly-connected double power-law. In this case, the conditional distribution of $\log M_{BH}$ at a given z is

$$f(\log M_{BH}|z) \quad \propto \quad M_{BH}^{\gamma(z)/\ln 10} \left[1 + \left(\frac{M_{BH}}{M_{BH}^*(z)}\right)^{(\gamma(z)+\delta(z))/\ln 10}\right]^{-1} \tag{5.63}$$

$$\gamma(z) \quad = \quad 2.5 + 0.5\log z \tag{5.64}$$

$$\delta(z) \quad = \quad 4.75 + 2\log z \tag{5.65}$$

$$\log M_{BH}^*(z) \quad = \quad 7.5 + 3\log(1+z), \tag{5.66}$$

where $\log M_{BH}^*(z)$ approximately marks the location of the peak in $f(\log M_{BH}|z)$, $\gamma(z)$ is the slope of $\log f(\log M_{BH}|z)$ for $M_{BH} \lesssim M_{BH}^*(z)$, and $\delta(z)$ is the slope of $\log f(\log M_{BH}|z)$ for $M_{BH} \gtrsim M_{BH}^*(z)$. For our simulation, both the peak and logarithmic slopes of the BHMF evolve.

The joint probability distribution of $\log M_{BH}$ and $\log z$ is $f(\log M_{BH}, \log z) = f(\log M_{BH}|z)f(\log z)$, and therefore Equations (5.60) and (5.61) imply that the true

BHMF for our simulated sample is

$$\phi_0(M_{BH}, z) \propto \frac{N}{zM_{BH}} \left(\frac{dV}{dz}\right)^{-1} f(\log M_{BH}|z) f(\log z). \tag{5.67}$$

The constant of proportionality in Equation (5.65) can be calculated by noting that

$\int \int \phi_0(M_{BH}, z) \, dM_{BH} \, dV = N$. Figure 5.2 shows $\phi_0(M_{BH}, z)$ at several redshifts. Also shown in Figure 5.2 is the best fit for a mixture of $K = 4$ gaussian functions. Despite the fact that $\phi_0(M_{BH}, z)$ has a rather complicated parameteric form, a mixture of four gaussian functions is sufficient to achieve a good approximation to $\phi_0(M_{BH}, z)$.

For each simulated black hole mass and redshift, we simulated a luminosity according to Equation (5.20). However, unlike the Gaussian distribution assumed in this work (see Eq.[5.19]), we assume an asymetric distribution of Eddington ratios that evolves at $\Gamma_{Edd} \propto \sqrt{1 + z}$. We do this in order to test the robustness of our simple assumption that the distribution of L_λ at a given M_{BH} is independent of redshift and given by a normal distribution. In this simulated 'universe', the distribution of Γ_{Edd} does not evolve strongly, as is implied by observations (e.g., Vestergaard, 2004; Kollmeier et al., 2006).

To simulate values of luminosity at a given black hole mass, we first simulated values of the Eddington ratio from a skew-normal distribution as

$$\log \Gamma_{Edd} = 0.2\epsilon - 0.75|\delta| - 0.3 + 0.5 \log(1 + z). \tag{5.68}$$

Here, ϵ and δ are both random deviates independently drawn from the standard normal distribution, i.e., $\epsilon, \delta \sim N(0, 1)$. Figure 5.3 shows the distribution of Γ_{Edd} at a few different redshifts. Values of λL_λ we then calculated according to Equation (5.20) assuming a constant bolometric correction of $C_\lambda = 10$ (e.g., Kaspi et al., 2000). For simplicity, we only use a constant bolometric correction for all
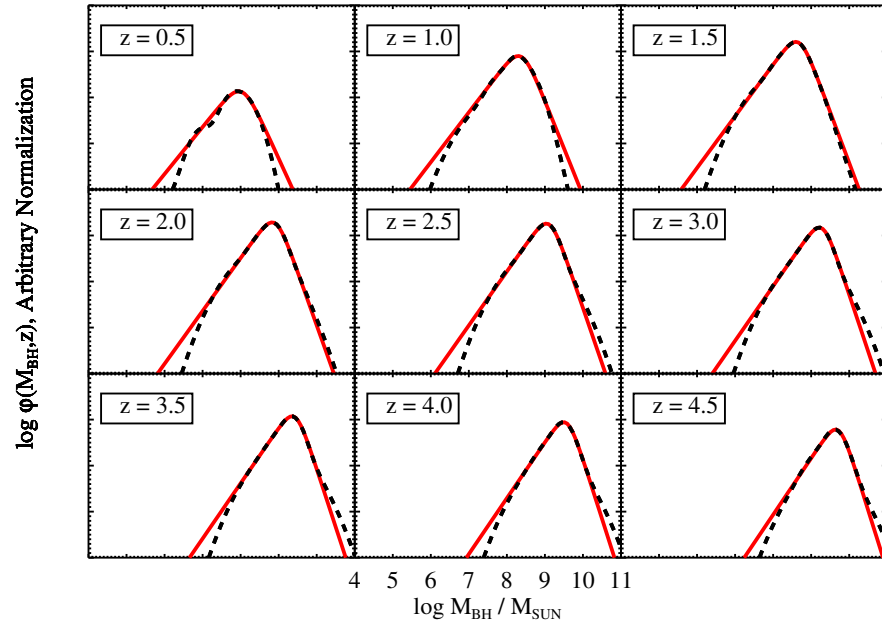
Figure 5.2 The true BHMF (solid red line) at several values of z, and the best $K = 4$ Gaussian function fit (dashed black line). In this case, approximating the BHMF with $K = 4$ 2-dimensional Gaussian functions provides an excellent fit.

.
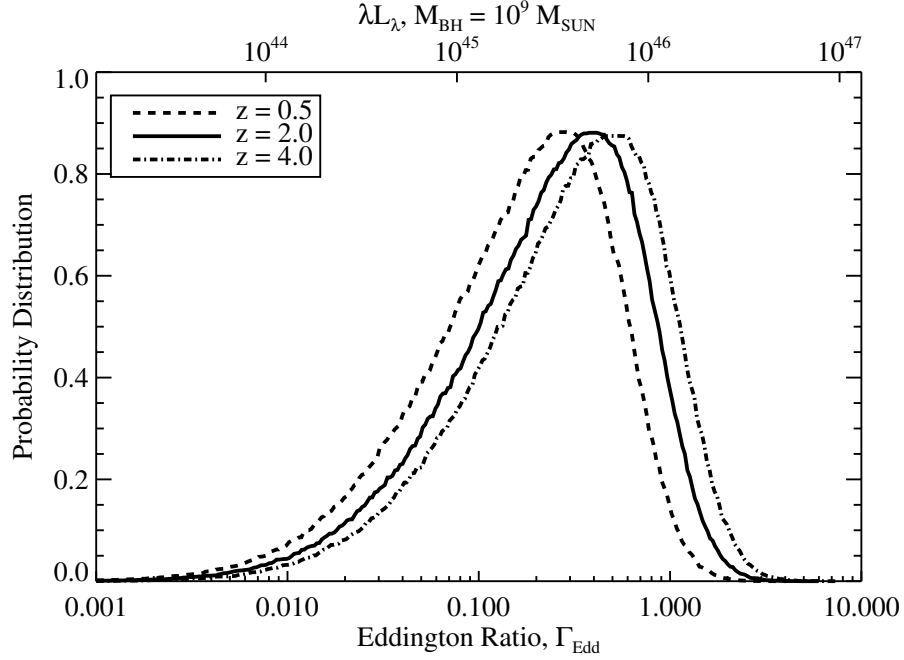
Figure 5.3 Distribution of Eddington ratios, Γ_{Edd}, for our simulated survey at $z = 0.5, 2$, and 4. The corresponding values of λL_λ are shown along the top of the plot for a black hole with $M_{BH} = 10^9 M_\odot$.

simulated quasars. In our simulation we take $\lambda = 2500$Å; the choice of λ is arbitrary and has not material effect on our results. The median Eddington ratio for our simulated sample is $\Gamma_{Edd} \approx 0.25$, and the dispersion in $\log \Gamma_{Edd}$ is ≈ 0.5 dex. Because the mean Γ_{Edd} evolves in our simulation, and because the mean M_{BH} evolves, Γ_{Edd} and M_{BH} are slightly correlated due to the shared correlation with z: $\Gamma_{Edd} \propto M_{BH}^{0.09}$. Therefore, $L_\lambda \propto M_{BH}^{1.09}$. Comparison with Equation (5.19) suggest that we would expect $\alpha_0 \sim 36$, $\alpha_m \sim 1.09$, and $\sigma_l \sim 0.5$ dex.

For each simulated black hole mass and luminosity, we simulated broad emission line widths for Hβ, Mg II, and C IV according to Equation (5.24). We simulated values of the Hβ line width for $0 < z < 0.9$, values of the Mg II line width

for $0.4 < z < 2.2$, and values of the C IV line width for $1.6 < z < 4.5$. Note that for this simulation both Hβ and Mg II are available at $0.4 < z < 0.9$, and both Mg II and C IV are available at $1.6 < z < 2.2$. Based on the most recent reverberation mapping data (Kaspi et al., 2005; Bentz et al., 2006), we took $R \propto L_\lambda^{1/2}$ ($\beta_l = 0.5$) for all emission lines. In addition, we set $\beta_0 = 10.6, 10.6$, and 10.7 for the Hβ, Mg II, and C IV emission lines, respectively; these values were chosen to give emission line $FWHM$ with typical values of several thousand km s^{-1}. The dispersion in the logarithm of the emission line width at a given luminosity and black hole mass was taken to be $\sigma_{BL} = 0.25, 0.225$, and 0.2 for Hβ, Mg II, and C IV, respectively. These values of σ_{BL} were chosen to give broad line mass estimate statistical uncertainties similar to that found from the reverberation mapping data (Vestergaard & Peterson, 2006).

We randomly kept each source, where the probability of including a source given its luminosity and redshift was taken to be the SDSS DR3 Quasar selection function, as reported by Richards et al. (2006). In addition, we only kept sources with at least one emission line having a line width $1000 \text{ km s}^{-1} < v < 1.8 \times 10^4 \text{ km s}^{-1}$. Sources with $v < 1000$ were assumed to be indistinguishable from narrow-line AGN, and sources with $v > 1.8 \times 10^4$ were assumed to be too difficult to distinguish from the underlying continuum and iron emission, and are thus too broad to be able to obtain a reliable estimate of the line width. After simulating the effects of the selection function, we were left with a sample of $n \sim 1000$ sources. Therefore, our simulated survey was only able to detect $\sim 0.5\%$ of the $N = 2 \times 10^5$ total quasars in our simulated 'universe'.

The distributions of M_{BH}, z, L_λ, and v are shown in Figure 5.4 for both the detected sources and the full sample. As can be seen, the majority of sources are missed by our simulated survey, and that the fairly 'hard' limit on luminosity

corresponds to a much 'softer' limit on M_{BH}. In particular, almost all simulated quasars with $M_{BH} \lesssim 10^8 M_\odot$ are missed at $z \gtrsim 1$, and all simulated quasars with $M_{BH} \lesssim 10^7 M_\odot$ are missed at any redshift.

To simulate the effects of using values of β_0 and σ_{BL} derived from a reverberation mapping sample, we simulated a sample of 25 low-z sources with known M_{BH}; these low-z sources were simulated in the same manner as described above. We then used these 25 'reverberation mapping' sources to fit β_0 and σ_{BL}. The fitted values were then used for our prior distribution on β_0 and σ_{BL} as described in § 5.5.3.

5.7.2   Performing Statistical Inference on the BHMF with the MCMC Output

We performed the MHA algorithm described in § 5.6 to obtain random draws from the posterior distribution for this sample, assuming the Gaussian mixture model described in § 5.5. We performed 10^4 iterations of burn-in, and then ran the markov chains for an additional 3×10^4. We ran five chains at the same time in order to monitor convergence (e.g., see Gelman et al., 2004) and explore possible multimodality in the posterior. The chains had converged after 4×10^4 total iterations, leaving us with $\sim 1.5 \times 10^5$ random draws from the posterior distribution, $p(\theta, N | \mathbf{v}_{obs}, L_{obs}, z_{obs})$.

In Figure 5.5 we show $\phi(\log M_{BH}, z)$ at several different redshifts, on both a linear scale and a logarithmic scale. In general, we find it easier to work with $\phi(\log M_{BH}, z) = \ln 10 M_{BH} \phi(M_{BH}, z)$, as $\phi(M_{BH}, z)$ can span several orders of magnitude in M_{BH}. Figure 5.5 shows the true value of the BHMF, $\phi_0(\log M_{BH}, z)$, the best-fit estimate of $\phi(\log M_{BH}, z)$ based on the mixture of Gaussian functions model, and the regions containing 90% of the posterior probability. Here, as well as throughout this work, we will consider the posterior median of any quantity to be the 'best-fit' for that quantity. In addition, in this work we will report errors
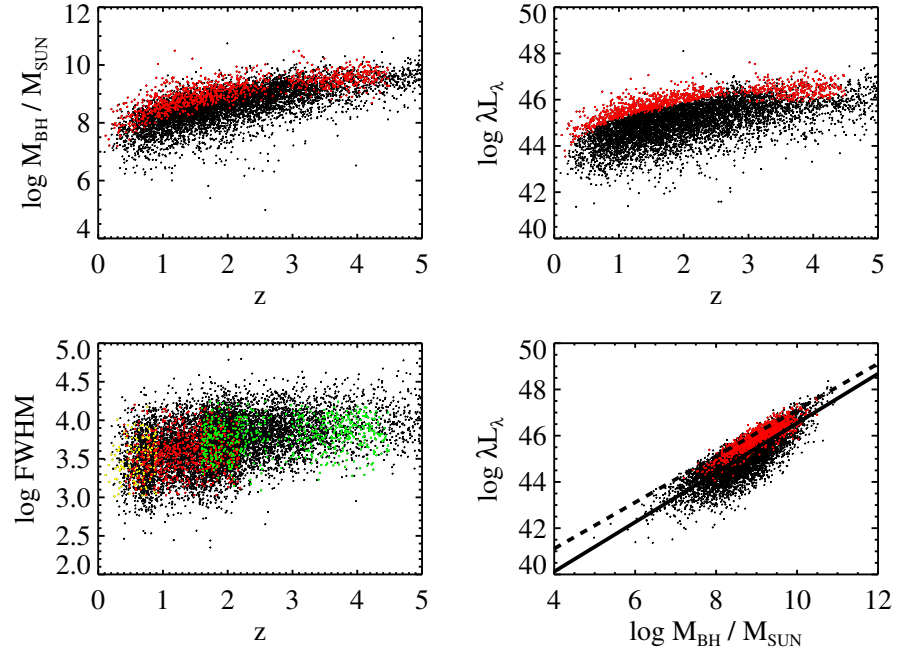
Figure 5.4 The distribution of M_{BH}, L_{λ}, and $\mathbf{v}$ for our simulated sample. Red dots denote sources included in the sample, and black dots denote sources not included in the sample. In the plot of $\mathbf{v}$ as a function of z, yellow dots denote sources with $H\beta$ measurements, red dots denote sources with Mg II measurements, and green dots denots sources with C IV measurements. In the plot of L_{λ} as a function of M_{BH}, the solid line shows the best linear regression of $\log L_{\lambda}$ as a function of $\log M_{BH}$, and the dashed line shows the Eddington limit assuming a bolometric correction of $C_{\lambda} = 10$.

at the 90% level unless specified otherwise, and therefore the regions containing 90% of the posterior probability can be loosely interpreted as asymmetric error bars of length $\approx 1.65\sigma$. As can be seen, the true value of $\phi(\log M_{BH}, z)$ is contained within the 90% probability region for all almost values of m, even those below the survey detection limit.

We compare our method with an estimate of the BHMF obtained by combining the broad line mass estimates with the more traditional $1/V_a$ estimator, developed for luminosity function estimation. We do this primarily to illustrate the pitfalls that can arise from employing broad line mass estimates and not properly accounting for the black hole mass selection function. Following Fan et al. (2001), we denote the effective volume of the ith source as V_a^i. If the ith source lies in a redshift bin of width Δz and has a luminosity $L_{\lambda,i}$, then

$$V_a^i = \int_{\Delta z} s(L_{\lambda,i}, z) \left(\frac{dV}{dz} \right) dz. \tag{5.69}$$

Dividing up the $(\log M_{BH}, z)$ plane into bins of width $\Delta \log M_{BH} \times \Delta z$, one may be tempted to calculate an estimate of $\phi(\log M_{BH}, z)$ based on the broad line estimates of $\log M_{BH}$ as

$$\hat{\phi}_{BL}(\log M_{BH}, z) = \frac{1}{\Delta \log M_{BH}} \sum_i \frac{1}{V_a^i}. \tag{5.70}$$

Here, the sum is over all sources with broad lines estimates $\log M_{BH} \leq \log \hat{M}_{BL,i} \leq \log M_{BH} + \Delta \log M_{BH}$ and $z \leq z_i \leq z + \Delta z$.

Figure 5.5 also displays the expected value of $\hat{\phi}_{BL}$ for $z = 0.5, 1.5, 2.5, 3.5$ and 4.5. In order to estimate the expected value of $\hat{\phi}_{BL}$ at each z, we simulated 10^7 quasars at each redshift interval. This produces extremely small error bars on $\hat{\phi}_{BL}$ and allows us to estimate the value of $\hat{\phi}_{BL}$ that would be obtained on average, i.e., in the limit of an infinitely large sample. As can be seen, $\hat{\phi}_{BL}$ is a biased
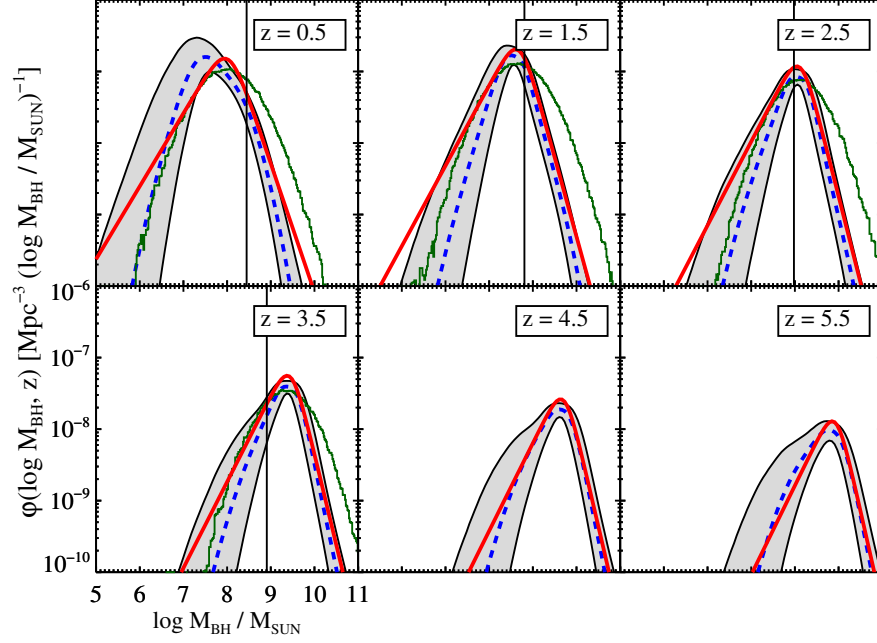
Figure 5.5 The true BHMF (solid red line) at several redshifts. The axis labels are the same for all panels, but for clarity we only label the bottom left panel. Also shown is the posterior median estimate of the BHMF based on the mixture of Gaussian functions model (dashed blue line), the region containing 90% of the posterior probability (shaded region), and the expected value for a $1/V_a$-type binned estimate based on the broad emission line estimates, $\hat{\phi}_{BL}$ (thin bumpy solid green line). The bayesian mixture of Gaussian functions model is able to accurately constrain the BHMF, even below the survey detection limit. However, $\hat{\phi}_{BL}$ provides a biased estimate of the BHMF.

estimate of the BHMF. The poor performance of $\hat{\phi}_{BL}$ lies in the use of the luminosity selection function and the broad line mass estimates. The large statistical uncertainties on the broad line mass estimates broaden the inferred BHMF, and therefore $\hat{\phi}_{BL}$ significantly overestimates the BHMF at the high mass end, while underestimating the BHMF near its peak. In addition, $\hat{\phi}_{BL}$ underestimates the BHMF at the low mass end due to the inability of the $1/V_a$ technique to correct for incompleteness. The end result is a systematic shift in the inferred BHMF toward higher M_{BH}, and a similar effect has been noted by Shen et al. (2007). The effective volume in Equation (5.67) is defined based on the detection probability as a function of luminosity, and not black hole mass. As mentioned in § 5.3.3, in order to correctly apply the $1/V_a$ estimator for BHMF estimation it is necessary to obtain the black hole mass selection function, given by Equation (5.11). However, this requires knowledge of $p(L_\lambda|M_{BH}, z)$. Furthermore, even if there were no selection effects, $\hat{\phi}_{BL}$ would still be biased because of the significant uncertainty (~ 0.4 dex) on $\log \hat{M_{BH\,BL}}$.

As in KFV08, we can use the MCMC output to constrain various quantities of interest calculated from the BHMF. Figure 5.6 compares the true integrated $z < 6$ number distribution of $\log M_{BH}$, $n(\log M_{BH}, z < 6)$, with the mixture of Gaussian functions estimate. The quantity $n(\log M_{BH}, z < 6)d\log M_{BH}$ is the number of quasars at $z < 6$ with black hole masses between $\log M_{BH}$ and $\log M_{BH} + d\log M_{BH}$. KFV08 give an equation for calculating $n(\log L, z < z_0)$ based on the mixture of Gaussian functions model (see their Eq.[42]), and $n(\log M_{BH}, z < z_0)$ is calculated in an equivalent manner. Similar to Figure 5.5, the true value of $n(\log M_{BH}, z < 6)$ is contained within the 90% probability region for all values of M_{BH}, even those below the survey detection limit.

In addition, in Figure 5.6 we show the comoving number density of broad line
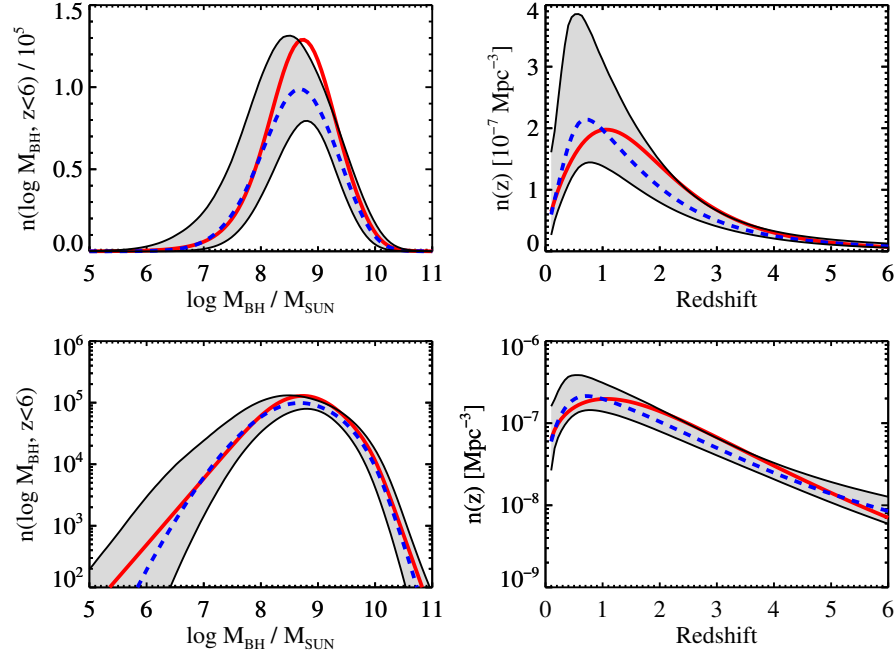
Figure 5.6 The integrated $z < 6$ quasar number density (number per $\log M_{BH}$ interval, left two panels) and the comoving quasar number density as a function of z (number per mpc^3, right two panels). The top two panels show a linear stretch and the bottom two panels show a logarithmic stretch. As with Figure 5.5, the solid red line denotes the true value for the simulation, the dashed blue line denotes the posterior median for the mixture of Gaussian functions model, and the shaded regions contain 90% of the posterior probability. The posterior median provides a good fit to the true values, and the uncertainties derived from the MCMC algorithm based on the Gaussian mixture model are able to accurately constrain the true values of these quantities, despite the flux limit.

AGN as a function of redshift, $n(z)$. This is obtained by integrating $\phi(M_{BH}, z)$ over all possible values of M_{BH}, given by Equation (45) of KFV08. As before, the true value of $n(z)$ is contained within the 90% probability region, despite the fact that the integration extends over *all* M_{BH}, even those below the detection limit. The wider confidence regions reflect additional uncertainty in $n(z)$ resulting from integration over those M_{BH} below the detection limit. In particular, the term dV/dz becomes small at low redshift, making the estimate of $n(z)$ more unstable as $z \to 0$, and thus inflating the uncertainties at low z.

Two other potentially useful quantities are the comoving black hole mass density for quasars, $\rho_{BH}^{QSO}(z)$, and its derivative. The comoving black hole mass density is given by $\rho_{BH}^{QSO}(z) = \int_0^\infty M_{BH}\phi(M_{BH}, z)\ dM_{BH}$. The quantity $\rho_{BH}^{QSO}(z)$ is given by Equation (47) of KFV08 and replacing luminosity with black hole mass. We calculate the derivative of $\rho_{BH}^{QSO}(z)$ numerically. Figure 5.7 compares the true values of $\rho_{BH}^{QSO}(z)$ and its derivative with the posterior distribution for $\rho_{BH}^{QSO}(z)$ inferred from the mixture model, both as a function of z and the age of the universe at redshift z, $t(z)$. Comparison with Figure 5.6 reveals that the comoving quasar black hole mass density, $\rho_{BH}^{QSO}(z)$, is a better constrained quantity than the comoving quasar number density, $n(z)$. Furthermore, $n(z)$ appears to peak later than $\rho_{BH}^{QSO}(z)$. We can correctly infer that the quasar comoving black hole mass density reaches it point of fastest growth at $t(z) \lesssim 1$ Gyr, and its point of fastest decline at $t(z) \sim 4$ Gyr.

Figure 5.8 quantifies the suggestion that $n(z)$ peaks later than $\rho_{BH}^{QSO}(z)$ by displaying the posterior distribution for the location of the respective peaks in $n(z)$ and $\rho_{BH}^{QSO}(z)$. While the location of the peak in $n(z)$ is highly uncertain we can still constrain it to be $z \lesssim 1.5$, whereas the location of the peak in $\rho_{BH}^{QSO}(z)$ is constrained to occur earlier at $2 \lesssim z \lesssim 4$. This is a consequence of the fact that
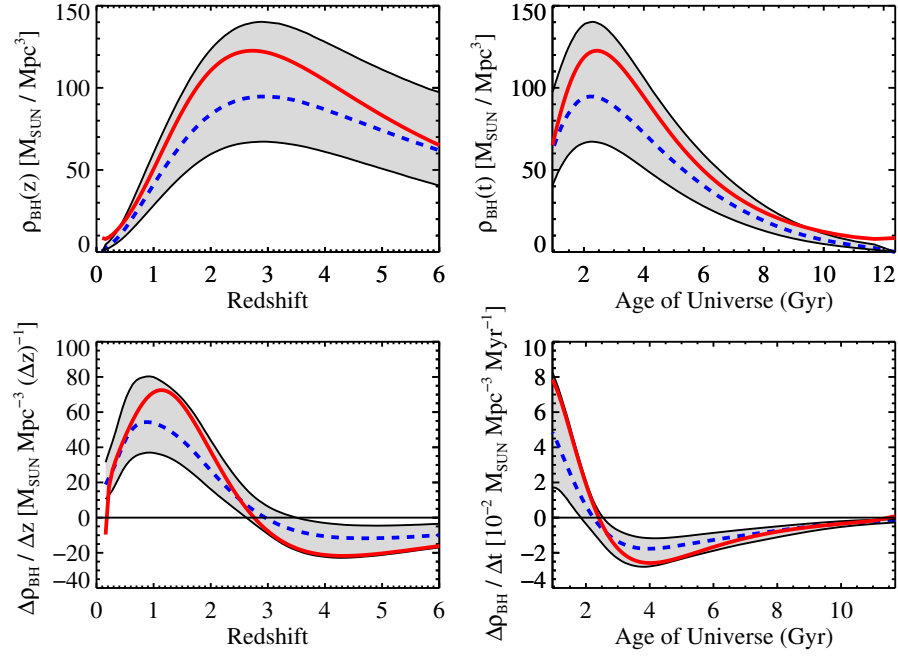
Figure 5.7 Comoving broad line quasar black hole mass density (top two panels) and its derivative (bottom two panels), shown as a function of redshift (left two panels) and cosmic age (right two panels). The plotting symbols are the same as in Figure 5.6. As in the previous figures, the Gaussian mixture model is able to provide an accurate fit to the true values of $\rho_{BH}^{QSO}(z)$, and the bayesian MCMC approach is able to provide accurate constraints on $\rho_{BH}^{QSO}(z)$ and $d\rho_{BH}^{QSO}/dz$, despite the fact that the integral used for calculating these quanties extends below the survey detection limit.

while there were more quasars at $z \sim 1$ per comoving volume, their black hole masses were much higher at higher redshift. This evolution in characteristic M_{BH} is quantified in Figure 5.9, which summarizes the posterior distribution for the location of the peak in $\phi(\log M_{BH}, z)$ as a function of redshift and $t(z)$. As can be seen, the location of the peak in the BHMF shows a clear trend of increasing 'characteristic' M_{BH} with increasing z, although the mixture of Gaussian functions fit has difficulty constraining the location of the peak at low redshift.

As noted in § 5.5.2, we can use the values of α_0 and σ_l to estimate the average Eddington ratio and the dispersion in $\log \Gamma_{Edd}$. We find $\alpha_0 = 35.7^{+0.9}_{-1.1}, \alpha_m = 1.11^{+0.12}_{-0.10}$, and $\sigma_l = 0.31^{+0.06}_{-0.05}$, where the errors are at 95% confidence. For a bolometric correction of $C_\lambda = 10$, and assuming that Γ_{Edd} is independent of M_{BH}, this implies that our inferred typical Eddington ratio is $\Gamma_{Edd} = 0.040^{+0.278}_{-0.036}$ at 95% confidence; the estimated dispersion in $\log \Gamma_{Edd}$ is simply given by $\sigma_l, \sim 0.3$ dex. While the typical Eddington ratio that we infer from α_0 is roughly consistent with the actual median Γ_{Edd} of 0.25, our estimated dispersion in Γ_{Edd} underestimates the true value of 0.5 dex. This is because we incorrectly assume that the M_{BH}–L relationship is described by Equation (5.19). Our inference regarding the Eddington ratio distribution is therefore biased because we assume that the distribution of Γ_{Edd} does not evolve, and that the distribution is Gaussian. In particular, the bias resulting from the assumption of Gaussian dispersion appears to significantly effect the estimated dispersion in $\log \Gamma_{Edd}$ more than the estimated typical value of Γ_{Edd}, at least for our simulation. This is largely because the distribution in Γ_{Edd} is skewed toward lower values of Γ_{Edd}. However, because of the flux limit, sources with low values of Γ_{Edd} are undetectable. Because the dispersion in $\log \Gamma_{Edd}$ is estimated from the detected sources, in combination with the assumption of a Gaussian distribution, Equation (5.19) is not able to pick up the
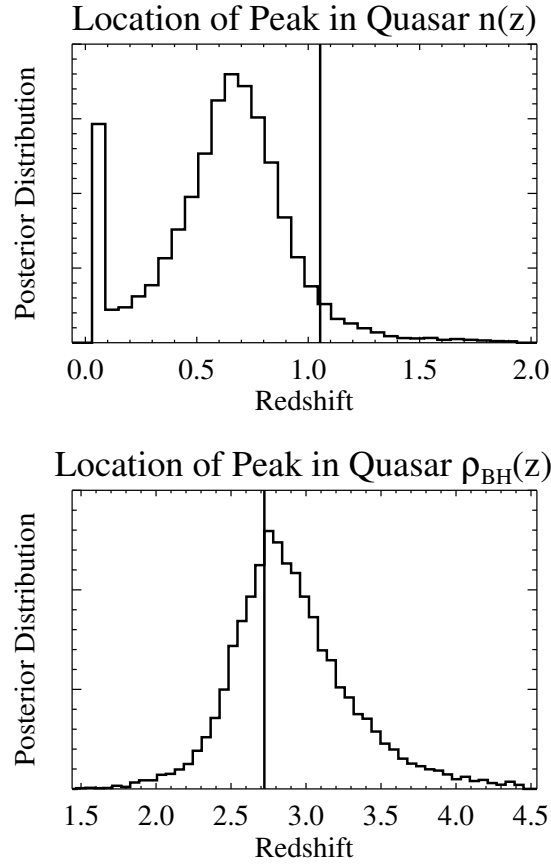
Figure 5.8 Posterior distribution for the redshift location of the peak in the comoving number density of quasars ($n(z)$, top) and the peak in the comoving quasar black hole mass density ($\rho_{BH}^{QSO}(z)$, bottom). The spike in the posterior at $z \approx 0$ for values of the peak in $n(z)$ arises because the term $(dV/dz)^{-1}$ becomes very large at low z. The vertical lines denote the true values. The posterior distribution inferred from the MCMC output is able to accurately constrain the true values of the argumentative maximum in $n(z)$ and $\rho_{BH}^{QSO}(z)$.
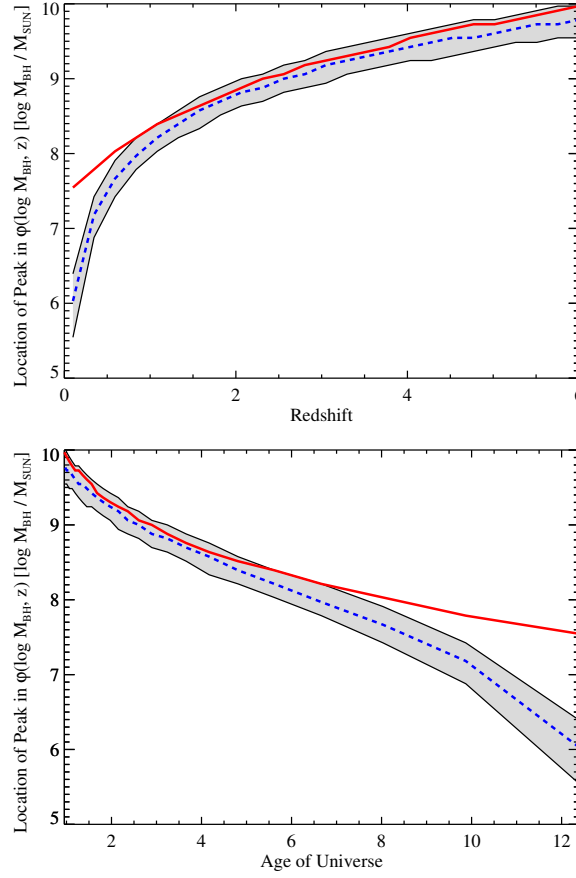
Figure 5.9 Location of the peak in the BHMF as a function of z (top) and cosmic age (bottom). The plot symbols are the same is in Figure 5.6. In general the posterior median of the Gaussian mixture model provides a good estimate of the true peak locations, although some bias is exhibited at the lowest redshifts. It is clear from these plots that the location of the peak in $\phi(M_{BH}, z)$ evolves.

additional skew at low $\log \Gamma_{Edd}$. As a result, the estimated dispersion in $\log \Gamma_{Edd}$ is underestimated when assuming a Gaussian distribution.

In order to assess how the inferred BHMF depends on the sample size, we simulated a second data set in the sammer manner as described above, but used $N = 2 \times 10^6$ sources for the BHMF normalization. This gave use $n \sim 10^4$ detected quasars. In Figure 5.10 we compare the estimated BHMF at $z = 2.5$ for the survey with $n \sim 1000$ sources and $n \sim 10^4$ sources. The uncertainties are lower for the survey with more sources, where the most noticeable improvement occurs at low M_{BH}. However, the increased sample size did not offer a significant amount of improvement at high M_{BH}, where sources are more easily detected. This is likely because the uncertainty in the broad line mass estimate normalization, β_0, and intrinsic scatter, σ_{BL}, dominates the uncertainty in the BHMF at high M_{BH}. Because we cannot constrain β_0 and σ_{BL} from the distribution of line widths and luminosities, the data do not contain any information on β_0 and σ_{BL}. Therefore, the likelihood function is unable to convey any information on β_0 and σ_{BL}, and all of the information comes from the prior distribution. As a result, our ability to constrain the BHMF is limited by the 'systematic' uncertainty on β_0 and σ_{BL}, and an increase in the sample size will eventually not result in a decrease in the uncertainty on the BHMF. The only way to reduce the uncertainty on the BHMF for large surveys is to better constrain the broad line mass estimate normalization and intrinsic scatter, most likely by increasing the sample of AGN with reverberation mapping data.

5.7.3 Using the MCMC Output to Evaluate the BHMF Fit

Throughout this section we have been analyzing the MCMC results by comparing to the true BHMF. However, in practice we do not have access to the true BHMF, and thus a method is needed for assessing the quality of the fit. As in
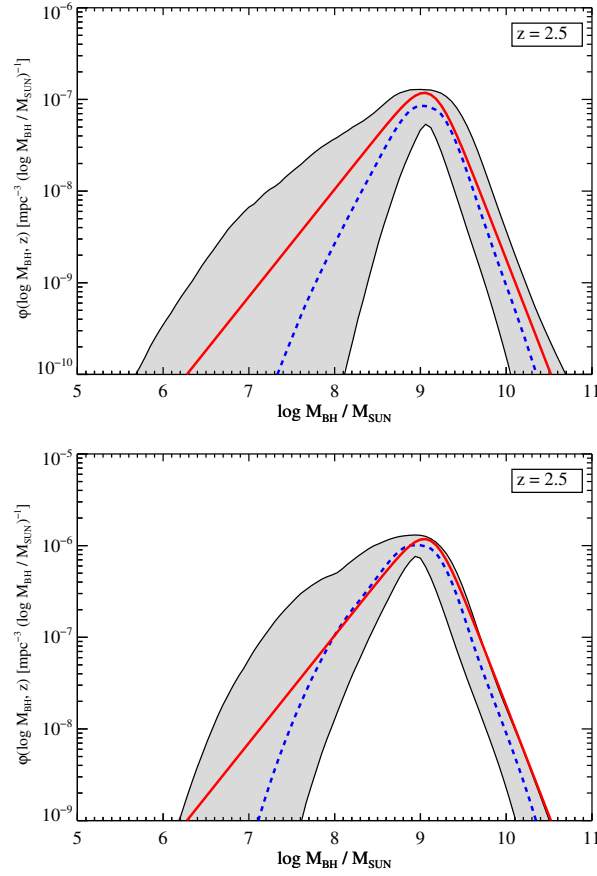
Figure 5.10 BHMF at $z = 2.5$ for the simulated sample with $n \sim 1000$ detected sources (top) and $n \sim 10^4$ detected sources (bottom); the top panel is the same as the $z = 2.5$ BHMF shown in Figure 5.5. The uncertainties derived for the $n \sim 10^4$ are smaller than for the $n \sim 1000$ sample, particularly at low M_{BH} where the survey becomes incomplete. However, the uncertainties for the $n \sim 10^4$ survey at high M_{BH}, where the survey is complete, are not considerably smaller than those for the $n \sim 1000$ survey. This is because the BHMF estimate is limited by the systematic uncertainty in the broad line mass estimate normalization, derived from β_0, and the broad line mass estimate statistical error, derived from σ_{BL}. Because the observed distribution of luminosities and line widths does not convey any information on these two quantities, increasing the sample size will not reduce the uncertainty on the BHMF beyond the systematic uncertainty on β_0 and σ_{BL}.

KFV08, the statistical model may be checked using a technique known as posterior predictive checking (e.g., Rubin, 1981, 1984; Gelman, Meng, & Stern, 1998). Here, the basic idea is to use each of the MCMC outputs to simulate a new random observed data set. The distributions of the simulated observed data sets are then compared to the true observed data in order to assess whether the statistical model gives an accurate representation of the observed data. It is important to construct simulated data sets for each of the MCMC draws in order to incorporate our uncertainty in the model parameters.

Random draws for M_{BH} and z for each MCMC draw may be obtained according to the procedure outlined in § 7.3 of KFV08, after replacing L with M_{BH}. Once one obtains a random draw of M_{BH} and z, simulated values of L_λ may be obtained using Equation (5.19) with α_0, α_m, and σ_l. Then, given these values of L_λ and M_{BH}, values of v for each emission line can be simulated from Equation (5.24) using the values of β_0, β_l, and σ_{BL}. Simulation from Equation (5.24) requires a value of α_λ in order to convert L_λ to L_λ^{BL}. In order to account for the intrinsic scatter in continuum slopes, we randomly draw of value of α_λ from our data set and use this value to convert to L_λ^{BL}. These simulated values of L_λ, z, and $\mathbf{v}$ are then folded through the selection function, leaving one with a simulated observed data set $(\mathbf{v}_{obs}, L_{obs}, z_{obs})$. This process is repeated for all values of N and θ obtained from the MCMC output, leaving one with simulated observed data sets of $(\mathbf{v}_{obs}, L_{obs}, z_{obs})$. These simulated observed data sets can then be compared with the true distribution of $\mathbf{v}_{obs}, L_{obs}$, and z_{obs} to test the statistical model for any inconsistencies.

In Figure 5.11 we show histograms for the observed distributions of z, $\log L_\lambda$, and $\log FWHM$ for the Hβ, Mg II, and C IV emission lines. These histograms are compared with the posterior median of the observed distributions based on

the mixture of Gaussian functions model, as well as error bars containing 90% of the simulated observed values. As can be seen, the distributions of the observed data sets simulated from our assumed statistical model are consistent with the distributions of the true observed data, and therefore there is no reason to reject the statistical model as providing a poor fit.

5.8   APPLICATION TO BQS QUASARS

As a final illustration of our method we used it to estimate the low redshift active BHMF from the 87 $z < 0.5$ quasars from the Bright Quasar Survey (BQS, Schmidt & Green, 1983). The Hβ line widths and continuum luminosities for 71 of the BQS quasars are taken from Table 7 of Vestergaard & Peterson (2006), and 16 of the quasars in the Boroson & Green (1992) sample have black hole mass estimates from reverberation mapping (Peterson et al., 2004). For each source with reverberation mapping data, we used the first entry of $\lambda L_\lambda(5100\text{Å})$ in Table 1 of Vestergaard & Peterson (2006) as the single-epoch luminosity; these values were based on continuum luminosities reported by Boroson & Green (1992) or Marziani et al. (2003). We assumed measurement errors of 10% on the emission line $FWHM$. The BQS sample covers an area of $\Omega = 10,714 \text{ deg}^2$ and is selected with an average flux limit of $B = 16.16$ (Schmidt & Green, 1983), with no apparent correlation with redshift and $U - B$ color (Jester, 2005). We converted the $B = 16.16$ flux limit to a flux limit at 5100Å assuming a power law continuum, $f_\nu \propto \nu^{-\alpha}$, with $\alpha = 0.5$ (Richards et al., 2001). We used $K = 3$ Gaussian functions to fit $\phi(M_{BH}, z)$ for $z < 0.5$.

Because we are including the actual values of M_{BH} for the 16 reverberation mapping sources, the contribution to the posterior for these sources is

$$p(\theta|M_{BH}, L_\lambda, z) = \prod_{i=1}^{16} p(\log L_{\lambda,i}|M_{BH,i}, \theta) p(\log M_{BH,i}, \log z_i|\theta). \qquad (5.71)$$
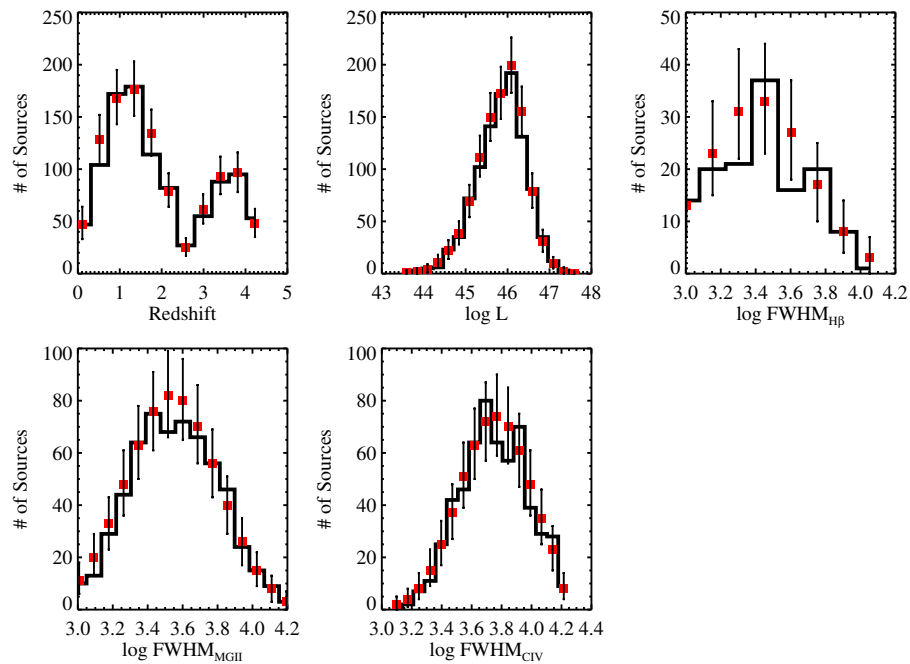
Figure 5.11 Posterior predictive check for the Gaussian mixture model (see
§ 5.7.3). The histograms show the actual distributions of $\log L_{obs}$, z_{obs}, and $\log \mathbf{v}_{obs}$,
the red squares denote the posterior medians for the number of sources in each
respective bin, and the error bars contain the inner 90% of the histogram values
for the samples simulated from the posterior. The mixture of Gaussian functions
model is able to provide an accurate prediction of the observed distribution of
luminosity, redshift, and line widths, and thus there is not any evidence to reject
it as providing a poor fit.

Here, $p(L_{\lambda,i}|M_{BH,i}, \theta)$ is given by Equation (5.19) and $p(\log M_{BH,i}, \log z_i|\theta)$ is given by Equation (5.17). The product in Equation (5.69) is only over the quasars with M_{BH} estimated from reverberation mapping, whereas the contribution to the posterior for the BQS sources without reverberation mapping is given by Equation (5.15). The posterior for the complete BQS sample is then the product of Equation (5.69) and Equation (5.15).

In Figure 5.12 we show the $z = 0.17$ BHMF derived from the BQS sample. Also shown is the binned BHMF for the BQS sources, calculated directly from the broad line mass estimates by Vestergaard (2006). We show the BHMF at $z = 0.17$ because the average redshift of the BQS sources is $z \approx 0.17$, therefore allowing a more direct comparison between the binned BHMF and the BHFM derived using our mixture of Gaussian functions approach. In addition, the uncertainties on our estimated BHMF are smallest at $z \approx 0.17$. We are able to place some constraints on the local BHMF, despite the fact that the BQS sample only contains 87 sources and has a very shallow flux limit. The $z \sim 0.2$ quasar BHMF appears to fall off as a power law above $M_{BH} \gtrsim 10^8 M_\odot$. Unfortunately, our estimate of the local BHMF becomes considerably uncertain below $M_{BH} \lesssim 10^8 M_\odot$, so it is unclear to what degree the power law trend continues below this point. In addition, the binned estimate overestimates the BHMF at the high M_{BH} end due to the intrinsic uncertainty in the broad line mass estimates, and underestimates the BHMF at the low M_{BH} end due to incompleteness, in agreement with our simulations (see § 5.7.2).

In Figure 5.12 we also compare our estimate of the BHMF at $z = 0.5$ with the $z = 0.5$ BHMF as reported by Vestergaard et al. (2008). Vestergaard et al. (2008) estimated the $z = 0.5$ BHMF by binning estimates of M_{BH} derived from the Hβ broad emission line over the redshift range $0.3 < z < 0.68$, using the SDSS DR3
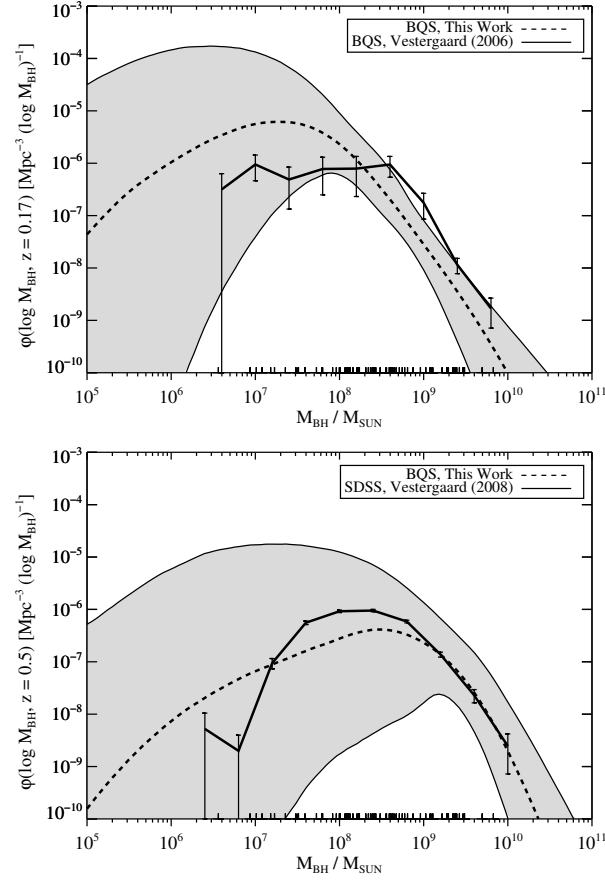
Figure 5.12 The $z = 0.17$ (top) and $z = 0.5$ (bottom) broad line quasar BHMF as estimated from the BQS sample. The dashed line denotes the posterior median for the mixture of Gaussian functions model, the shaded region contains 90% of the posterior probability, and the tick marks along the x-axis mark the locations of the broad line mass estimates. The estimate of the $z = 0.17$ BHMF becomes significantly uncertain at $M_{BH} \lesssim 10^8 M_\odot$, and the $z = 0.17$ BHMF appears to fall off as a power law above $M_{BH} \gtrsim 10^8 M_\odot$. The $z = 0.5$ BHMF is not very well constrained, but there is evidence for a shift in the BHMF toward higher M_{BH} from $z = 0.17$ to $z = 0.5$. For comparison, we show the BHMF estimated by Vestergaard (2006) using the BQS sources (top, solid line with error bars), and the BHMF estimated by Vestergaard et al. (2008) using the SDSS DR3 quasars (bottom, solid line with error bars). The shift in the BHMF inferred from the binned mass estimates is apparent in the BQS sample, while the SDSS and BQS $z = 0.5$ BHMF estimates agree fairly well.

quasar catalogue (Schneider et al., 2005). Despite the differences in approach and survey selection, the two estimates of the $z = 0.5$ BHMF agree fairly well. However, because $z = 0.5$ defines the upper redshift limit of our BQS sample, the uncertainties on the BHMF derived from the BQS quasars are very large. In addition, incompleteness in M_{BH} likely affects the low M_{BH} bins of the Vestergaard et al. (2008), causing the Vestergaard et al. (2008) $z = 0.5$ BHMF to underestimate the true $z = 0.5$ BHMF in these bins. However, a direct comparison between our Bayesian approach and the Vestergaard et al. (2008) estimate is difficult, due to the different redshift ranges used to estimate the BHMF, and the different selection methods of the BQS and the SDSS.

Although the BQS has a small sample size and probes narrow range in z, we can attempt to quantify any evolution in the local BHMF by comparing the ratio of the comoving number density of quasars at two different values of M_{BH}. Comparison of the estimated BHMF at $z = 0.17$ and $z = 0.5$ suggests a shift in the BHMF toward large M_{BH}. In Figure 5.13 we show the best fit values of the ratio of $\phi(\log M_{BH}, z)$ at $M_{BH} = 5 \times 10^8 M_\odot$ to $\phi(\log M_{BH}, z)$ at $M_{BH} = 5 \times 10^9 M_\odot$ as a function of z, as well as the 90% confidence interval. The logarithm of this ratio gives the slope of a power-law between $M_{BH} = 5 \times 10^8 M_\odot$ and $M_{BH} = 5 \times 10^9 M_\odot$, and therefore allows us to probe evolution in the shape of the quasar BHMF at the high M_{BH} end. In general, the ratio is fairly flat, implying no evolution in the high M_{BH} slope of the BHMF. However, at $z \gtrsim 0.3$ there is marginal evidence for an increase in the comoving number density of quasars hosting $M_{BH} \sim 5 \times 10^9 M_\odot$ SMBHs, as compared to the comoving number density of quasars hosting $M_{BH} \sim 5 \times 10^8 M_\odot$. The values of this ratio imply that the BHMF at the high M_{BH} end falls off as a power-law with slope ~ 2, although slopes of ~ 1 and ~ 3 are also consistent with the BQS quasars.
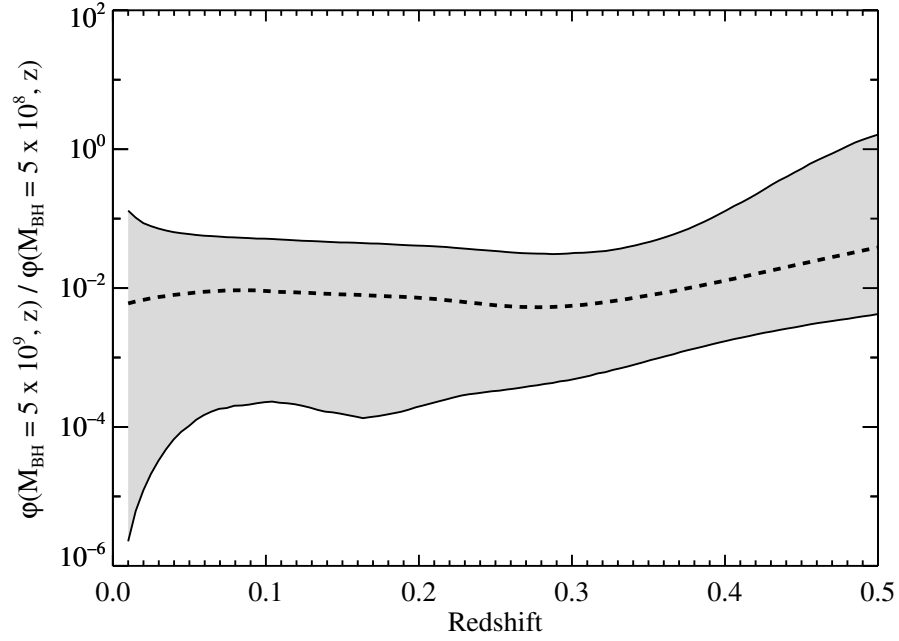
Figure 5.13 The ratio of the broad line quasar BHMF at $M_{BH} = 5 \times 10^9 M_\odot$ compared to the BHMF at $M_{BH} = 5 \times 10^8 M_\odot$, as a function of z and estimated from the BQS quasars. The dashed line is the posterior median, and the shaded region contains 90% of the probability. Assuming that the BHMF is a power-law from $M_{BH} = 5 \times 10^8 M_\odot$ to $M_{BH} = 5 \times 10^9 M_\odot$, the logarithm of this ratio is the slope of the BHMF. The high M_{BH} BHMF slope appears to be fairly constant for $z \lesssim 0.3$ with a slope of ~ 2, and there is marginal evidence for a flattening of the high M_{BH} slope at $z \gtrsim 0.3$.

In figure 5.14 we summarize the posterior probability distribution for the parameters governing the distribution of L_λ at a given M_{BH} (see Eq.[5.19]). Based on the MCMC results, we can constrain the M_{BH}–$\lambda L_\lambda(5100\text{Å})$ relationship at $z < 0.5$ to be

$$\lambda L_\lambda(5100\text{Å}) = 5.18^{+429}_{-5.14} \times 10^{36} \left(\frac{M_{BH}}{M_\odot}\right)^{0.92\pm0.24} \; [\text{erg s}^{-1}], \qquad (5.72)$$

where we have quoted the errors at 95% confidence. The dispersion in L_{5100} at a given M_{BH} is estimated to be $\sigma_l = 0.35^{+0.13}_{-0.08}$. Assuming that the bolometric correction is on average $C_{5100} \sim 10$ (e.g., Kaspi et al., 2000), comparison of Equation (5.70) with Equation (5.20) suggests that $z < 0.5$ broad line AGN have typical Eddington ratios of $\Gamma_{Edd} \sim 0.4$. As argued in § 5.5.2, the scatter in the distribution of $\log L_\lambda$ at a given M_{BH} is the convolution of the scatter in $\log \Gamma_{Edd}$ with the scatter in $\log C_\lambda$. Therefore, the dispersion in L_λ at a given M_{BH} is a combination of the dispersion in Eddington ratio and bolometric correction. As a result, we are unable to estimate the dispersion in Eddington ratios at a given M_{BH} from σ_l. However, if the bolometric correction to L_{5100} increases with increasing Eddington ratio, as found by Vasudevan & Fabian (2007), or if the bolometric correction is independent of Γ_{Edd}, then the dispersion in Γ_{Edd} must be less than σ_l. Therefore, because we infer that $\sigma_l \lesssim 0.5$ dex, our results imply that the dispersion in Eddington ratios at a given M_{BH} is $\lesssim 0.5$ dex for $z < 0.5$ broad line quasars. These results on the Eddington ratio distribution are consistent with previous work (e.g., McLure & Dunlop, 2004; Vestergaard, 2004; Kollmeier et al., 2006); however, they may be biased because of our assumption of a Gaussian and non-evolving Eddington ratio distribution. In particular, if the distribution of Eddington ratios is skewed toward low $\log \Gamma_{Edd}$, then we will have underestimated the intrinsic dispersion in $\log \Gamma_{Edd}$.
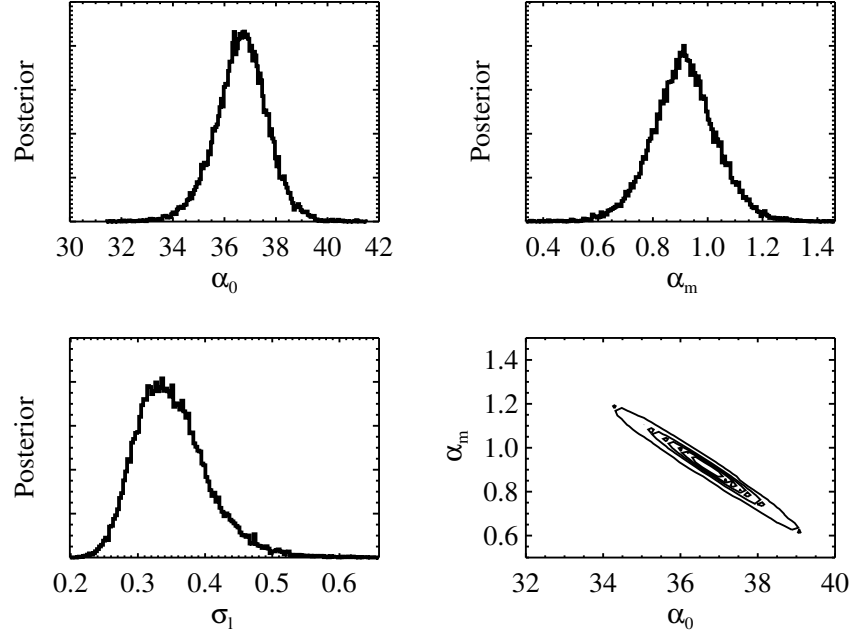
Figure 5.14 Posterior distributions of the parameters for the distribution of luminosities at a given M_{BH}, as estimated from the $z < 0.5$ BQS quasars. The uncertainty on α_0 and α_m is highly correlated. Assuming a bolometric correction of $C_{5100} \sim 10$, the values of α_0 and σ_l imply that the $z < 0.5$ distribution of broad line quasar Eddington ratios has a mean of $\Gamma_{Edd} \sim 0.4$ and a dispersion of ~ 0.5 dex.

5.9 SUMMARY

We have derived the observed data likelihood function which relates the quasar BHMF to the observed distribution of redshifts, luminosities, and broad emission line widths. This likelihood function is then used in a Bayesian approach to estimating the BHMF, where the BHMF is approximated as a mixture of Gaussian functions. Because much of this work was mathematically technical, we summarize the important points here.

- Using methods developed for luminosity function estimation (e.g., $1/V_a$-type estimators) without modification will lead to errors in black hole mass function estimation, as the black hole mass selection function is not equivalent to the flux selection function. In addition, using broad line estimates of M_{BH} will lead to a broader inferred BHMF if one does not correct for the intrinsic uncertainty in the broad line mass estimates. This causes one to overestimate $\phi(M_{BH}, z)$ in the tails of the distribution, and underestimate $\phi(M_{BH}, z)$ near the peak of the distribution. However, because low M_{BH} AGN are more likely to be missed by flux-limited surveys, $\phi(M_{BH}, z)$ will be underestimated at low M_{BH} due to incompleteness. The end result is a spurious shift in the inferred BHMF toward higher M_{BH}: incompleteness at low M_{BH} causes one to low M_{BH} sources while the intrinsic statistical uncertainty on the broad line mass estimates causes one to overestimate the number of high M_{BH} black holes.

- Equations (5.15) and (5.16) give the posterior probability distribution of the BHMF parameters, given the observed data, for broad line mass estimates. These equations may be used to perform Bayesian inference on the BHMF for any assumed parameterization.

- In this work we describe a flexible parameteric model for the BHMF, where the BHMF is modelled as a mixture of Gaussian functions. The distribution of luminosities is modelled as a linear regression of $\log L_\lambda$ as a function of $\log M_{BH}$, where the intrinsic scatter in $\log L_\lambda$ at a given M_{BH} was assumed to follow a normal distribution. The distribution in line widths at a given L_λ and M_{BH} is also assumed to have the form of a linear regression, where the parameters are based on the most recent broad line mass estimates. Equation (5.18) gives the BHMF under the mixture of Gaussian function model.

  Equations (5.30) and (5.46) define the likelihood function for broad line mass estimates under the mixture of Gaussian functions model if only one emission line at a given z is used to estimate M_{BH}. Otherwise, if multiple emission lines are used for a single quasar, then Equation (5.41) must be used. The posterior is then found by pluggin the prior distribution and likelihood function into Equations (5.15) and (5.16).

- In § 5.5.5 we modify the likelihood function to include measurement error in the emission line width. We show that if the measurement errors on the line width are much smaller than the intrinsic physical dispersion in line widths, then measurement error may be neglected. However, if measurement error on the line width is a concern, Equations (5.56)–(5.58) should be used for Equation (5.29) instead of Equation (5.30).

- We describe in § 5.6 a Metropolis-Hastings algorithm (MHA) for obtaining random draws from the posterior distribution of the BHMF under the mixture of Gaussian functions model. These random draws may be used to estimate the posterior distribution for the BHMF, as well as to estimate the posterior for any quantities calculated from the BHMF. The posterior

provides statistically accurate uncertainties on the BHMF and related quantities, even below the survey detection limits. We use simulation in § 5.7 to illustrate the effectiveness of our statistical method, as well as to give an example on how to use the MHA output to perform statistical inference.

- We concluded by applying our method to obtain an estimate of the local unobscured quasar BHMF from the $z < 0.5$ BQS quasar sample. Although there is little information in the BQS quasars on the BHMF at $M_{BH} \lesssim 10^8 M_\odot$, the mixture of Gaussian functions estimate suggests that the local quasar BHMF falls off approximately as a power law with slope ~ 2 for $M_{BH} \gtrsim 10^8 M_\odot$ at $z \approx 0.2$. The local quasar BHMF appears to shift toward larger M_{BH} at higher z, and there is marginal evidence for a flattening of the high mass BHMF slope at $z \gtrsim 0.3$. We estimate that at a given M_{BH}, $z < 0.5$ broad line quasars have a typical Eddington ratio of ~ 0.4 and a dispersion in Eddington ratio of $\lesssim 0.5$ dex. However, the estimate of the dispersion in Eddington ratio could be biased toward smaller values if the true distribution of Eddington ratios is significantly skewed toward lower values.

CHAPTER 6

EVOLUTION OF THE X-RAY EMISSION OF RADIO-QUIET QUASARS

6.1   CHAPTER ABSTRACT

We report new *Chandra* observations of seven optically faint, $z \sim 4$ radio-quiet quasars. We have combined these new observations with previous *Chandra* observations of radio-quiet quasars to create a sample of 174 sources. These sources have $0.1 < z < 4.7$, and $10^{44}\,\mathrm{ergs\,s^{-1}} < \nu L_\nu(2500\text{Å}) < 10^{48}\,\mathrm{ergs\,s^{-1}}$. The X-ray detection fraction is 90%. We find that the X-ray loudness of radio-quiet quasars decreases with UV luminosity and increases with redshift. The model that is best supported by the data has a linear dependence of optical-to-X-ray ratio, α_{ox}, on cosmic time, and a quadratic dependence of α_{ox} on $\log L_{UV}$, where α_{ox} becomes X-ray quiet more rapidly at higher $\log L_{UV}$. We find no significant evidence for a relationship between the X-ray photon index, Γ_X, and the UV luminosity, and we find marginally significant evidence that the X-ray continuum flattens with increasing z (2σ). The Γ_X–z anti-correlation may be the result of X-ray spectral curvature, redshifting of a Compton reflection component into the observed *Chandra* band, and/or redshifting of a soft excess out of the observed *Chandra* band. Using the results for Γ_X, we show that the α_{ox}–z relationship is unlikely to be a spurious result caused by redshifting of the observable X-ray spectral region. A correlation between α_{ox} and z implies evolution of the accretion process. We present a qualitative comparison of these new results with models for accretion disk emission.

6.2   CHAPTER INTRODUCTION

It is widely accepted that the extraordinary activity associated with quasars involves accretion onto a supermassive black hole, with the UV/optical emission

arising from a geometrically thin, optically thick cold accretion disk, and the X-ray continuum arising from a hot, optically thin corona that Compton upscatters the disk's UV photons. The geometry of the X-ray emitting region is uncertain, but possibilities include an accretion disk that evaporates into a hot inner flow (e.g., Shapiro et al., 1976; Zdziarski et al., 1999), a hot ionized 'skin' that sandwiches the cold disk (e.g., Bisnovatyi-Kogan & Blinnikov, 1977; Liang & Price, 1977; Nayakshin, 2000), a combination of a hot inner flow and a corona that sandwiches the disk (e.g., Poutanen et al., 1997; Sobolewska et al., 2004a), or a patchy corona, consisting of a number of hot spots above the accretion disk (e.g., Galeev et al., 1979; Malzac et al., 2001; Sobolewska et al., 2004b). In addition, the UV and X-ray producing processes may be coupled as a result of radiation pressure from the UV photons driving a flow from the disk into the corona (Proga, 2005). Investigating the relationships between the UV and X-ray emission is an important step towards understanding the origin of the X-ray emission. Furthermore, learning how the X-ray and UV emission change with z provides insight into evolution of the accretion process, quasar black hole mass, and accretion rate.

Many studies have investigated whether α_{ox}, the ratio of X-ray to UV/optical flux, depends on redshift or UV luminosity, L_{UV} (e.g., Avni & Tananbaum, 1982; Wilkes et al., 1994; Yuan et al., 1998a; Bechtold et al., 2003; Vignali et al., 2003b; Strateva et al., 2005; Steffen et al., 2006). The parameter α_{ox} is a simple measure of the amount of X-ray radiation, dominated by non-thermal processes, in respect to the amount of UV radtion, dominated by thermal processes. Most studies have concluded that there is no evidence for a redshift dependence of α_{ox} (e.g., Avni & Tananbaum, 1986; Wilkes et al., 1994; Strateva et al., 2005), although Bechtold et al. (2003) argued that α_{ox} is significantly correlated with both z and UV luminosity, and Yuan et al. (1998a) found evidence for a slight dependence of α_{ox} on

redshift at $z < 0.5$. Vignali et al. (2003b), Strateva et al. (2005, S05), and Steffen et al. (2006, S06) used a partial correlation and regression analysis to conclude that there is no evidence for a dependence of $\alpha_{\rm ox}$ on z, after accounting for the $\alpha_{\rm ox}$–L_{UV} and L_{UV}–z correlations. These authors also found evidence that RQQs become more X-ray quiet with increasing UV luminosity.

Previous investigations of the X-ray photon index, Γ_X, have also produced mixed results. Bechtold et al. (2003) used a sample of *ROSAT* observations over from Yuan et al. (1998a) and *Chandra* observations of high redshift quasars to conclude that Γ_X is correlated with both luminosity and z. A Γ_X–L_{UV} correlation was also seen by Dai et al. (2004), using a sample of gravitationally-lensed sources with *XMM-Newton* and *Chandra* data. Some evidence for an anti-correlation between Γ_X and z has also been found using *ASCA* observations (Reeves et al., 1997; Vignali et al., 1999) and *XMM-Newton* observations (Page et al., 2003). However, other investigations based on *XMM-Newton* data (e.g., Risaliti & Elvis, 2005) and spectral fitting of composite spectra from *Chandra* observations (e.g., Vignali et al., 2003c, 2005; Shemmer et al., 2006) have not revealed any evidence for a relationship between Γ_X, L_{UV}, and z. Similarly, other *ASCA* observations have also not produced evidence for a Γ_X–z correlation (Reeves & Turner, 2000). In addition, there has been evidence for a correlation between Γ_X and the Eddington ratio (e.g., Lu & Yu, 1999; Wang et al., 2004; Shemmer et al., 2006), and Gallagher et al. (2005) found evidence for an anti-correlation between Γ_X and the UV spectral slope.

All of these conclusions are by necessity based on flux-limited samples. Flux-limited samples typically suffer from an artificial correlation between z and L_{UV}, making it difficult to disentangle which parameter is more important in determining X-ray properties. To help break the L_{UV}–z degeneracy, we observed

seven optically faint $z \sim 4$ radio-quiet quasars with *Chandra*. We combined these sources with *Chandra* data of optically-selected radio-quiet quasars, drawn mostly from the Sloan Digital Sky Survey (SDSS, York et al., 2000), to create a flux-limited sample of 174 sources, 90% of which have detections. Because the X-ray emission in radio-loud sources can have an additional component from the jet (e.g., Zamorani et al., 1981; Wilkes & Elvis, 1987), we focus our analysis on the radio-quiet majority. We use these sources to perform a multivariate analysis of $\alpha_{\mathrm{ox}}, \Gamma_X, L_{UV}$, and z in a manner that allows us to effectively separate the dependence of the X-ray spectral properties on L_{UV} and z.

We adopt a cosmology based on the the WMAP best-fit parameters ($h = 0.71, \Omega_m = 0.27, \Omega_\Lambda = 0.73$, Spergel et al., 2003). For ease of notation, we define $l_{UV} \equiv \log \nu L_\nu(2500\text{Å})$, and $l_X \equiv \log \nu L_\nu(2\,\mathrm{keV})$.

6.3   OBSERVATIONS AND COMPARISON SAMPLE

The new observations targeted seven non-BAL RQQs with $z > 4$ from the literature. These were known to be among the faintest ($\nu L_\nu(2500\text{Å}) \lesssim 3 \times 10^{46}\,\mathrm{ergs\,s^{-1}}$) $z \sim 4$ optically-selected quasars to be observed thus far by *Chandra* or *XMM-Newton*. All seven were observed on-axis on the ACIS-S3 chip with exposure times 10–23 ksec. The exposure times were chosen in order to ensure that the X-ray source would be detected if $\alpha_{ox} < 1.9$. All targets were, in fact, detected. The new observations are summarized in Table 6.1.

The other $z \gtrsim 4$ sources were selected from the literature (Bechtold et al., 2003; Vignali et al., 2001, 2003a), and had been observed as targeted observations with Chandra. The $z \lesssim 4$ sources were found by cross-correlating the SDSS DR3 quasar catalogue (Schneider et al., 2005) with the *Chandra* public archive as of 2005 February 22. We selected those SDSS DR3 quasars that were serendipitiously

Table 6.1. List of New Observations

Source	RA	DEC	z	r^{a}	OBSID	Counts[b]	$f_X{}^{\text{c}}$	Exp. Time	Spec. Ref.[d]	Phot. Ref.[e]
	J2000	J2000					10^{-15} ergs cm^{-2} s^{-1}	ksec		
SDSS 0050-0053	00 50 06.3	-00 53 19.0	4.331	20.13	4825	26.3	6.91 ± 1.38	13.0	1	1
Q 0910+564	09 14 39.3	+56 13 21.0	4.035	20.87	4821	13.4	1.79 ± 0.57	23.0	2	2
SDSS 1321+0038	13 21 10.8	+00 38 22.0	4.716	21.30	4824	19.8	3.88 ± 0.87	17.8	1	1
SDSS 1413+0000	14 13 15.3	+00 00 32.0	4.078	19.75	4823	23.0	4.33 ± 1.02	12.5	1	1
SDSS 1444-0123	14 44 28.7	-01 23 44.0	4.179	19.64	4826	12.1	4.12 ± 1.26	10.0	1	1
PC 1450+3404	14 53 00.6	+33 52 06.0	4.191	20.81	4822	20.2	6.20 ± 1.36	14.8	3	4
SDSS 2357+0043	23 57 18.3	+00 43 50.0	4.362	19.92	4827	19.7	5.20 ± 1.23	12.7	1	1

[a]r-band apparent magnitude.

[b]Number of observed background-subtracted source counts in the range 0.3–7.0 keV.

[c]Unabsorbed 2–10 keV flux, assuming a power law with $\Gamma_X = 1.9$.

[d]Reference for the observed frame optical spectrum.

[e]Reference for the r-band magnitude.

References. — (1) SDSS; (2) Schneider et al. (1991); (3) Constantin et al. (2002); (4) Schneider et al. (1997)

within $12'$ of a *Chandra* target. The radio-quiet sources were selected to have $R_i = 0.4(i - t_{1.4\mathrm{GHz}}) < 1.5$ (Ivezić et al., 2004), where, $t_{1.4\mathrm{GHz}}$ is the FIRST 1.4 GHz AB magnitude, and i is the SDSS i-band magnitude. The radio-loud sources were omitted because such sources have an additional component of X-ray emission arising from the jet (e.g., Wilkes & Elvis, 1987; Worrall et al., 1987). Almost all of the $z < 4$ quasars have their *Chandra* data reported here for the first time.

The optical/UV spectra for each source were inspected by eye to exclude the BALs or any sources that had significant absorption. It is necessary to remove the BAL QSOs because their high column density gives them the appearance of being X-ray weak (e.g., Green et al., 2001; Gallagher et al., 2002, 2006), potentially biasing our analysis. We are unable to remove the high-ionization BAL quasars for $z < 1.5$, as their identification requires observations of the C IV line. We are able to remove low-ionization BALs at $0.45 < z < 2.25$ based on Mg II absorption. Reichard et al. (2003) found the fraction of BALs in the SDSS to be $\sim 14\%$, and therefore we expect there to be 13 ± 3 BALs in our sample at $z < 1.5$. We did not include seven sources with an obvious contribution in their spectra from the host-galaxy. Host-galaxy contamination is likely negligible for all included sources, except for possibly the lowest luminosity quasars, since $\nu L_\nu^* \sim 10^{44}\ \mathrm{ergs\ s^{-1}}$ at $2500 \mathring{A}$ for galaxies (Budavári et al., 2005).

We visually inspected the *Chandra* events files to find those sources that fell on an ACIS chip. We did not include any sources that were observed on chip S4 due to higher read-out noise (Data Caveats on CIAO web pages [1]). All X-ray sources reported are within 1"–2" of the optical position. The archival sources and their X-ray properties are listed in Table 6.2.

---

[1]http://cxc.harvard.edu/ciao/caveats

Table 6.2.   List of Archival Sources

Source	RA J2000	DEC J2000	z	r[a]	OBSID	θ[b]	Counts[c]	Exp. Time ksec	f_X[d] 10^{-14} ergs cm^{-2} s^{-1}	Spec. Ref.[e]	Phot. Ref.[f]
0002+0049	00 02 30.7	+00 49 59.0	1.352	18.12	4861	0.58	188.0	5.7	10.52 ± 1.07	1	1
0006-0015	00 06 54.1	-00 15 33.4	1.725	18.20	4096	0.58	46.2	4.5	3.35 ± 0.49	1	1
0022+0016	00 22 10.0	+00 16 29.3	0.574	18.06	2252	7.12	343.4	71.2	2.82 ± 0.48	1	1
0027+0026	00 27 52.4	+00 26 15.7	0.205	18.18	4080	4.48	54.4	1.6	11.66 ± 4.47	1	1
0031+0034	00 31 31.4	+00 34 20.2	1.735	18.79	2101	1.20	73.9	6.7	3.10 ± 0.51	1	1
0057+1450	00 57 01.1	+14 50 03.0	0.623	18.81	865	1.15	4.6	4.6	< 0.91	1	1
0059+0003	00 59 22.8	+00 03 01.0	4.178	19.5	2179	0.58	9.0	2.7	1.21 ± 0.38	2	5
0106+0048	01 06 19.2	+00 48 22.0	4.437	19.1	2180	0.60	24.2	3.7	2.04 ± 0.41	2	5
0113+1531	01 13 05.7	+15 31 46.5	0.576	18.30	3219	2.09	1291.8	58.5	12.41 ± 0.53	1	1
0113+1535	01 13 09.1	+15 35 53.6	1.806	18.33	3219	6.29	528.3	58.5	6.20 ± 0.29	1	1
0115+0020	01 15 37.7	+00 20 28.7	1.275	18.72	3204	6.79	406.2	37.6	6.77 ± 0.66	1	1
0133+0400	01 33 40.4	+04 00 59.0	4.150	18.0	3152	0.59	39.4	6.1	2.03 ± 0.33	3	3
0134+3307	01 34 21.5	+33 07 56.6	4.530	18.9	3018	0.60	22.7	5.0	1.38 ± 0.32	3	3
0148+0001	01 48 12.2	+00 01 53.3	1.704	17.67	4098	0.57	42.3	3.7	3.57 ± 0.55	1	1
0148-0002	01 48 21.0	-00 02 25.8	0.930	18.39	4098	5.41	94.7	3.7	7.96 ± 1.53	1	1
0152+0105	01 52 58.7	+01 05 07.4	0.647	19.22	1448	4.13	104.6	7.9	6.86 ± 1.30	1	1
0153+0052	01 53 09.1	+00 52 50.1	1.161	18.81	3580	7.26	58.1	19.9	2.27 ± 0.37	1	1
0156+0053	01 56 50.3	+00 53 08.5	1.652	18.65	4100	0.59	95.2	5.6	4.58 ± 0.49	1	1
0159+0023	01 59 50.2	+00 23 40.8	0.162	15.97	4104	0.58	5708.2	9.7	189.5 ± 8.19	1	1
0201-0919	02 01 18.7	-09 19 35.8	0.660	17.61	3772	7.83	379.2	19.7	11.42 ± 1.17	1	1
0208+0022	02 08 45.5	+00 22 36.1	1.885	17.08	4099	0.58	58.2	3.5	4.69 ± 0.65	1	1
0209+0517	02 09 44.7	+05 17 14.0	4.140	17.8	3153	0.60	30.2	5.8	1.66 ± 0.31	3	3
0232-0731	02 32 17.7	-07 31 19.9	1.163	19.19	3030	10.1	0.0	4.2	< 1.85	1	1
0241-0811	02 41 05.8	-08 11 53.2	0.978	19.87	385	1.69	5.8	2.3	< 1.56	1	1
0241+0023	02 41 10.0	+00 23 01.4	0.790	20.47	4011	9.33	15.8	5.0	2.35 ± 0.57	1	1
0244-0134	02 44 01.9	-01 34 03.0	4.053	18.4	875	0.64	17.6	7.4	0.55 ± 0.15	2	6
0248+1802	02 48 54.3	+18 02 49.9	4.430	18.4	876	0.66	18.1	1.7	3.55 ± 0.83	2	5
0259+0048	02 59 59.7	+00 48 13.6	0.892	19.44	4145	0.54	40.6	4.7	3.13 ± 0.47	1	1
0311-1722	03 11 15.2	-17 22 47.3	4.000	18.0	3154	0.57	8.9	6.1	< 1.07	3	3
0314-0111	03 14 27.5	-01 11 52.3	0.387	18.05	4084	4.48	114.8	1.9	23.9 ± 5.29	1	1

Table 6.2—Continued

Source	RA J2000	DEC J2000	z	r[a]	OBSID	θ[b]	Counts[c]	Exp. Time ksec	f_X[d] 10^{-14} ergs cm^{-2} s^{-1}	Spec. Ref.[e]	Phot. Ref.[f]
0403-1703	04 03 56.6	-17 03 24.1	4.236	18.7	2182	0.59	13.9	3.8	1.08 ± 0.29	2	6
0419-5716	04 19 50.9	-57 16 13.1	4.460	18.7	4066	0.58	9.3	4.0	0.71 ± 0.24	3	3
0755+2203	07 55 02.1	+22 03 46.9	0.399	18.99	647, 3767	3.85, 7.26	457.0	156.2	5.83 ± 0.41	1	1
0755+4058	07 55 35.6	+40 58 03.0	2.417	18.85	3032	9.48	15.6	7.3	1.37 ± 0.36	1	1
0755+4111	07 55 40.0	+41 11 19.1	0.967	17.86	3032	10.0	9.6	7.3	1.13 ± 0.35	1	1
0755+4056	07 55 45.6	+40 56 43.6	2.348	19.17	3032	9.06	19.8	7.3	1.49 ± 0.37	1	1
0819+3649	08 19 51.4	+36 49 50.8	0.736	19.19	4119	11.5	86.6	7.3	11.22 ± 2.71	1	1
0832+5243	08 32 06.0	+52 43 59.3	1.572	19.47	1643	3.80	7.6	9.1	< 0.56	1	1
0845+3431	08 45 26.6	+34 31 02.0	2.046	19.88	818	10.1	3.4	4.5	< 2.13	1	1
0849+4457	08 49 05.1	+44 57 14.8	1.259	20.00	927, 1708	3.26, 3.26	938.2	186.5	2.37 ± 0.11	1	1
0849+4500	08 49 43.7	+45 00 24.3	1.592	18.39	927, 1708	10.4, 10.4	1149.6	186.5	3.87 ± 0.13	1	1
0910+5427	09 10 29.0	+54 27 19.0	0.525	18.76	2227	7.57	3327.4	105.7	19.95 ± 0.74	1	1
0912+0547	09 12 10.3	+05 47 42.1	3.240	18.06	419, 1629	10.9, 11.0	53.9	38.6	0.88 ± 0.14	1	1
0918+5139	09 18 28.6	+51 39 32.1	0.185	17.46	533	4.51	147.5	11.3	5.05 ± 3.14	1	1
0918+0647	09 18 47.5	+06 47 04.7	0.821	18.80	3563	11.1	78.0	4.9	5.62 ± 1.63	1	1
0933+5515	09 33 59.3	+55 15 50.8	1.863	19.08	805	1.27	779.7	40.8	4.57 ± 0.24	1	1
0941+5948	09 41 33.7	+59 48 11.3	0.967	16.38	3035	2.83	398.1	4.2	25.16 ± 2.86	1	1
0950+5619	09 50 24.0	+56 19 46.7	1.912	20.53	4151	6.76	29.0	8.9	1.87 ± 0.34	1	1
0951+5940	09 51 30.2	+59 40 37.1	1.056	18.74	3036	5.39	45.9	5.1	4.48 ± 0.66	1	1
0951+5944	09 51 51.6	+59 44 30.0	2.338	19.79	3036	1.12	28.5	5.1	1.26 ± 0.28	1	1
0952+5152	09 52 40.2	+51 52 50.0	0.553	18.47	3195	2.42	1487.6	26.9	12.81 ± 0.94	1	1
0952+5151	09 52 43.0	+51 51 21.1	0.861	17.34	3195	3.25	1688.2	26.9	14.79 ± 0.76	1	1
0955+5935	09 55 05.6	+59 35 17.6	0.912	18.91	3156	4.84	38.3	5.7	3.50 ± 0.56	1	1
0955+5940	09 55 11.3	+59 40 32.2	4.340	18.58	3156	0.60	17.8	5.7	0.90 ± 0.21	1	1
0955+4116	09 55 42.1	+41 16 55.3	3.420	19.36	5294	7.45	31.2	17.3	1.01 ± 0.20	1	1
0955+4109	09 55 48.1	+41 09 55.3	2.307	18.74	5294	2.67	71.9	17.3	2.47 ± 0.30	1	1
0956+4110	09 56 40.4	+41 10 43.5	1.887	20.49	5294	7.23	12.8	17.3	0.58 ± 0.15	1	1
0958+0734	09 58 20.5	+07 34 36.1	1.885	18.44	2990	9.38	296.9	14.1	14.01 ± 1.66	1	1
0958+0747	09 58 22.6	+07 47 47.7	3.218	20.07	2990	8.48	10.6	14.1	0.72 ± 0.20	1	1
0958+0745	09 58 36.6	+07 45 56.3	1.487	19.17	2990	6.23	192.1	14.1	4.70 ± 0.63	1	1

Table 6.2—Continued

Source	RA J2000	DEC J2000	z	r[a]	OBSID	θ[b]	Counts[c]	Exp. Time ksec	f_X[d] 10^{-14} ergs cm^{-2} s^{-1}	Spec. Ref.[e]	Phot. Ref.[f]
1002+5542	10 02 05.4	+55 42 57.9	1.151	18.03	2038	2.84	161.6	26.6	3.44 ± 0.62	1	1
1003+4736	10 03 52.8	+47 36 53.4	2.934	19.72	4152	0.59	64.5	13.7	1.24 ± 0.16	1	1
1013-0052	10 13 14.9	-00 52 33.6	0.275	17.78	4085	4.49	192.3	2.0	32.75 ± 5.50	1	1
1019+4737	10 19 02.0	+47 37 14.6	2.944	19.19	4153	0.58	28.7	8.0	0.98 ± 0.19	1	1
1023+0415	10 23 50.9	+04 15 42.0	1.809	19.40	1651, 909	6.66, 6.66	240.7	211.7	2.19 ± 0.14	1	1
1030+0524	10 30 31.6	+05 24 54.9	1.182	17.74	3357	0.88	11.2	8.0	0.57 ± 0.15	1	1
1032+5738	10 32 27.9	+57 38 22.5	1.968	20.58	3345, 3344	8.17, 8.18	589.7	77.0	3.79 ± 0.16	1	1
1032+5800	10 32 36.2	+58 00 34.0	0.686	19.83	3343	7.90	1.7	37.0	< 0.25	1	1
1036-0343	10 36 23.8	-03 43 20.0	4.509	18.5	877	0.65	15.7	3.4	1.29 ± 0.33	2	6
1038+4727	10 38 08.7	+47 27 34.9	1.047	18.56	4154	3.96	15.6	9.8	0.65 ± 0.17	1	1
1042+0100	10 42 30.7	+01 00 01.6	1.400	18.40	4086	4.87	21.2	1.7	4.05 ± 1.27	1	1
1044+5921	10 44 54.9	+59 21 34.1	1.291	19.03	5030	7.31	265.5	65.7	2.51 ± 0.27	1	1
1049+5750	10 49 21.5	+57 50 36.6	1.106	18.81	1673	8.73	23.9	4.9	2.92 ± 0.62	1	1
1050+5702	10 50 15.6	+57 02 55.7	3.273	20.17	1679, 1680	10.9, 8.19	24.0	9.4	1.47 ± 0.33	1	1
1050+5738	10 50 50.1	+57 38 20.0	1.281	19.09	1678	3.98	32.8	4.7	3.49 ± 0.62	1	1
1052+5724	10 52 39.6	+57 24 31.4	1.111	17.79	1683	2.96	90.8	4.7	8.81 ± 1.64	1	1
1053+5735	10 53 16.8	+57 35 50.8	1.204	19.08	1683, 1684	9.42, 7.40	161.8	9.4	10.6 ± 0.96	1	1
1054+5740	10 54 04.1	+57 40 19.8	1.100	18.04	1688	5.25	24.7	4.7	2.61 ± 0.54	1	1
1054+5720	10 54 22.6	+57 20 31.0	2.972	19.85	1687	7.13	3.1	4.7	< 1.99	1	1
1055+5704	10 55 18.1	+57 04 23.6	0.695	18.73	1686, 1691	8.96, 8.35	86.7	9.4	6.43 ± 1.06	1	1
1056+5722	10 56 44.5	+57 22 33.5	0.286	18.90	1693	4.96	26.2	5.7	2.56 ± 0.52	1	1
1057+4555	10 57 56.4	+45 55 52.0	4.100	17.48	878	0.65	34.8	2.8	3.12 ± 0.53	2	1
1109+0900	11 09 05.3	+09 00 48.7	1.674	19.42	3252	7.34	22.1	10.0	1.40 ± 0.31	1	1
1111+5532	11 11 32.1	+55 32 40.3	1.004	18.44	2025	7.89	333.8	59.4	3.70 ± 0.37	1	1
1114+5315	11 14 52.8	+53 15 31.7	1.213	19.02	3253, 3321	4.01, 8.41	39.8	13.6	4.60 ± 0.46	1	1
1115+5309	11 15 20.7	+53 09 22.1	0.877	18.05	3321	1.18	2.6	4.8	< 0.55	1	1
1129-0137	11 29 43.9	-01 37 52.3	1.294	18.15	2082	5.66	128.7	4.8	7.96 ± 1.14	1	1
1129-0150	11 29 51.2	-01 50 37.3	1.784	20.24	2082	7.89	23.7	4.8	2.86 ± 0.56	1	1
1136+0159	11 36 21.2	+01 59 27.9	0.766	19.24	4833	3.15	85.1	5.9	5.28 ± 0.98	1	1
1136+0158	11 36 31.9	+01 58 01.1	1.470	17.85	4833	0.57	9.8	5.9	0.74 ± 0.21	1	1

Table 6.2—Continued

Source	RA J2000	DEC J2000	z	r [a]	OBSID	θ [b]	Counts[c]	Exp. Time ksec	f_X [d] 10^{-14} ergs cm^{-2} s^{-1}	Spec. Ref.[e]	Phot. Ref.[f]
1136+0207	11 36 33.1	+02 07 47.7	0.239	18.07	4833	9.31	65.4	5.9	5.48 ± 2.06	1	1
1202-0129	12 02 26.8	-01 29 15.3	0.150	17.13	4108	0.59	2361.2	9.4	54.36 ± 3.27	1	1
1204+0150	12 04 36.6	+01 50 25.6	1.927	18.63	3234	5.49	148.6	30.0	1.85 ± 0.19	1	1
1208+0016	12 08 29.6	+00 16 42.7	1.063	18.97	2083	5.86	34.8	4.6	3.53 ± 0.61	1	1
1213+0252	12 13 43.0	+02 52 48.9	0.641	19.30	4110	4.55	57.9	10.0	1.69 ± 0.47	1	1
1214+0055	12 14 15.2	+00 55 11.5	0.395	18.35	4087	4.48	137.3	2.0	21.42 ± 3.75	1	1
1215-0034	12 15 40.5	-00 34 33.8	0.757	19.46	4201	3.45	551.2	44.5	7.16 ± 0.56	1	1
1218+0546	12 18 36.1	+05 46 28.1	0.795	18.89	3322	0.98	72.5	4.6	3.53 ± 0.60	1	1
1220-0025	12 20 04.4	-00 25 39.1	0.421	18.96	3141	0.58	2380.4	19.7	30.45 ± 1.26	1	1
1223+1034	12 23 07.5	+10 34 48.2	2.747	18.59	3232	4.65	302.6	30.1	3.02 ± 0.20	1	1
1226-0011	12 26 52.0	-00 11 59.6	1.175	17.88	4865	0.58	137.7	4.9	8.07 ± 1.05	1	1
1228+4413	12 28 18.0	+44 13 02.0	0.662	18.05	2031	5.81	3.0	26.6	< 0.27	1	1
1228+4411	12 28 53.7	+44 11 52.9	1.276	18.79	2031	9.09	96.6	26.6	2.71 ± 0.30	1	1
1230+0302	12 30 05.8	+03 02 04.2	1.604	18.91	4040	3.01	34.1	3.5	3.07 ± 0.56	1	1
1230+0305	12 30 25.9	+03 05 35.4	1.055	19.45	4040	3.17	35.9	3.5	4.72 ± 0.80	1	1
1230+0306	12 30 27.4	+03 06 27.5	0.628	18.65	4040	4.05	105.6	3.5	14.99 ± 2.32	1	1
1230+0308	12 30 39.9	+03 08 57.3	1.843	19.50	4040	8.02	12.7	3.5	1.73 ± 0.54	1	1
1230+0305	12 30 54.7	+03 05 37.2	0.427	19.19	4040	9.76	16.8	3.5	3.05 ± 0.78	1	1
1236+6215	12 36 22.9	+62 15 26.6	2.587	20.44	580, 2423, 2344, 3409, 967, 966, 3389, 957, 3408, 2233, 2232, 2386, 3388, 2421, 2234, 3293, 3294, 3390, 3391, 1671	6.83, 1.67, 5.38, 1.79, 5.46, 5.45, 4.11, 1.73, 4.12, 1.67, 1.69, 5.38, 4.12, 1.64, 1.64, 4.12, 1.79, 1.79, 1.79, 5.37	3741.3	1961.0	0.90 ± 0.02	1	1
1237+6203	12 37 16.0	+62 03 23.4	2.068	19.86	580, 2344, 967, 966, 2386, 1671	9.23, 9.77, 9.59, 9.62, 9.78, 9.78	198.9	431.9	0.30 ± 0.03	1	1
1242+0249	12 42 55.3	+02 49 57.0	1.458	19.21	323, 3926	8.17, 6.56	657.2	127.1	3.250 ± 0.17	1	1
1245-0027	12 45 41.0	-00 27 44.9	1.693	18.58	4018	7.82	88.7	4.9	11.24 ± 1.21	1	1
1255+5652	12 55 35.1	+56 52 39.6	1.803	19.13	1031	6.00	339.3	39.3	1.79 ± 0.18	1	1
1255+5650	12 55 36.2	+56 50 00.1	1.373	19.83	1031	6.32	0.0	39.3	< 0.14	1	1
1258-0143	12 58 49.8	-01 43 03.3	0.967	17.06	4178	4.58	2024.9	27.3	21.09 ± 1.00	1	1

Table 6.2—Continued

Source	RA J2000	DEC J2000	z	r^a	OBSID	θ^b	Counts[c]	Exp. Time ksec	$f_X{}^d$ $10^{-14}\,\mathrm{ergs\,cm^{-2}\,s^{-1}}$	Spec. Ref.[e]	Phot. Ref.[f]
1259+0102	12 59 43.6	+01 02 55.0	0.394	18.34	4088	4.48	4.0	1.9	< 2.47	1	1
1311+0031	13 11 08.5	+00 31 51.7	0.429	17.92	4089	4.48	14.9	1.7	2.86 ± 0.76	1	1
1317+3531	13 17 43.2	+35 31 31.1	4.360	19.1	879	0.64	5.7	2.8	< 1.78	2	5
1344-0000	13 44 25.9	-00 00 56.2	1.095	18.58	2251	0.39	26.2	9.6	0.82 ± 0.16	1	1
1411+5217	14 11 04.1	+52 17 55.6	2.882	19.07	2254	6.15	247.3	90.9	1.48 ± 0.11	1	1
1411+5205	14 11 04.9	+52 05 16.8	1.083	18.97	2254	7.19	41.1	90.9	1.20 ± 0.68	1	1
1417+4456	14 17 00.8	+44 56 06.4	0.113	16.32	541	8.26	5313.3	31.2	92.89 ± 4.08	1	1
1419+4709	14 19 51.9	+47 09 01.4	2.288	17.37	3076	0.59	174.3	7.7	6.77 ± 0.53	1	1
1424+4214	14 24 14.1	+42 14 00.1	1.608	19.03	3077	5.91	133.1	5.9	6.81 ± 0.79	1	1
1424+4210	14 24 36.0	+42 10 30.5	2.217	17.51	3077	0.58	137.0	5.9	5.35 ± 0.57	1	1
1432-0059	14 32 44.4	-00 59 15.2	1.026	17.26	907	7.56	1715.9	21.4	43.53 ± 2.13	1	1
1433+0227	14 33 35.3	+02 27 18.3	2.072	19.94	3959	3.85	17.0	3.5	2.57 ± 0.63	1	1
1434+0227	14 34 07.5	+02 27 04.6	1.710	19.41	3959	4.24	28.4	3.5	2.91 ± 0.56	1	1
1438+0341	14 38 42.0	+03 41 10.4	1.737	18.27	3290	5.97	296.9	57.6	3.11 ± 0.20	1	1
1438+0335	14 38 59.1	+03 35 47.5	0.733	18.43	3290	8.23	498.2	57.6	7.25 ± 0.50	1	1
1442+0110	14 42 31.7	+01 10 55.3	4.560	20.90	3960	0.58	43.9	11.0	1.37 ± 0.20	4	1
1443+5856	14 43 40.8	+58 56 53.2	4.260	18.28	3160	0.58	16.3	5.8	0.71 ± 0.19	1	1
1445+0129	14 45 54.8	+01 29 03.3	1.845	20.00	2112	2.40	38.9	5.9	2.22 ± 0.34	1	1
1448+4738	14 48 53.4	+47 38 21.3	2.894	19.39	4155	0.58	19.0	6.9	0.93 ± 0.21	1	1
1448+0015	14 48 56.7	+00 15 10.3	0.832	18.80	4092	7.59	19.2	2.1	4.64 ± 1.08	1	1
1449+0024	14 49 13.5	+00 24 06.9	0.440	19.13	4092	3.10	48.0	2.1	7.82 ± 1.18	1	1
1452+4304	14 52 15.6	+43 04 48.7	0.296	18.89	1048, 2424	3.14, 3.11	785.1	47.2	5.010 ± 0.49	1	1
1452+4308	14 52 40.9	+43 08 14.4	1.704	19.41	1048	8.36	139.8	17.7	4.46 ± 0.38	1	1
1511+5659	15 11 26.5	+56 59 34.8	1.031	17.55	3334	9.81	3.9	4.9	< 4.30	1	1
1515+5521	15 15 04.9	+55 21 07.3	1.844	20.40	3006	9.83	18.2	9.6	1.02 ± 0.39	1	1
1539+4313	15 39 47.6	+43 13 41.6	0.347	18.75	2993	1.30	870.1	14.8	15.96 ± 1.28	1	1
1543+5405	15 43 16.4	+54 05 26.1	0.245	18.11	822	8.39	132.3	4.5	14.35 ± 3.03	1	1
1545+4846	15 45 30.2	+48 46 09.1	0.399	16.44	3339	8.93	245.8	4.9	34.29 ± 5.79	1	1
1605-0109	16 05 17.8	-01 09 55.5	1.572	19.14	2086	5.42	80.4	4.6	7.08 ± 0.91	1	1
1618+3456	16 18 34.0	+34 56 25.6	1.922	18.73	3341	5.29	17.4	4.9	1.83 ± 0.46	1	1
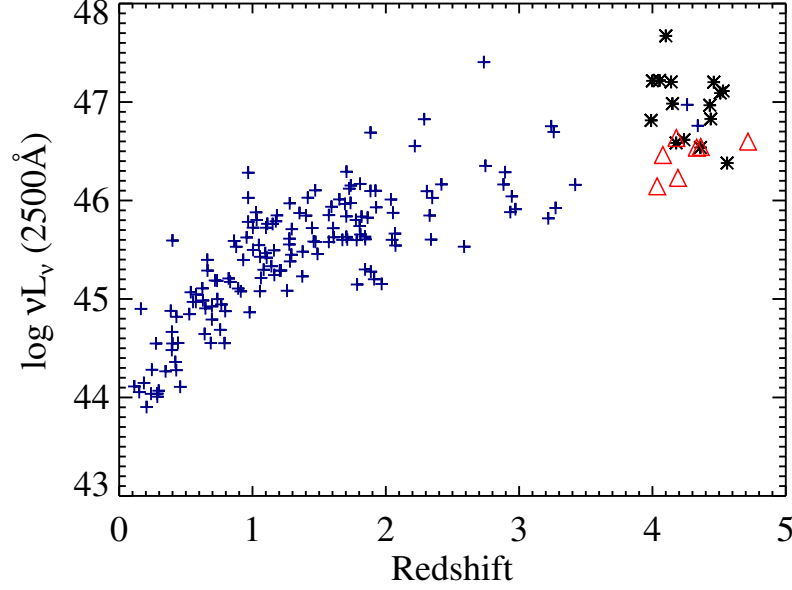
Figure 6.1 The (L_{UV}, z) distribution of our sample. The seven new observations are denoted by red triangles, the SDSS sources are denoted by blue crosses, and the non-SDSS $z \gtrsim 4$ sources are denoted by black asterisks. There appears to be data points for only six RQQs with new *Chandra* observations because sources 0050-0053 and 2357+0053 have almost the same L_{UV} and z, causing their symbols to overlap.

Altogether, the sample consists of 174 radio-quiet quasars, with a broad range in redshift $(0.1 < z < 4.7)$ and luminosity $(10^{44} \text{ ergs s}^{-1} \lesssim \nu L_\nu(2500\text{Å}) \lesssim 10^{48} \text{ ergs s}^{-1})$. All sources have been observed with the *Chandra X-ray Observatory* using the ACIS-S or ACIS-I detectors, and 157 (90%) of them are detected. The (L_{UV}, z) distribution of the sample is shown in Figure 6.1.

Table 6.2—Continued

Source	RA J2000	DEC J2000	z	r[a]	OBSID	θ[b]	Counts[c]	Exp. Time ksec	f_X[d] 10^{-14} ergs cm^{-2} s^{-1}	Spec. Ref.[e]	Phot. Ref.[f]
1640+4644	16 40 25.0	+46 44 49.1	0.537	18.38	896	1.17	1049.2	42.3	5.17 ± 0.41	1	1
1641+4649	16 41 10.6	+46 49 11.9	0.695	19.21	896	9.98	366.1	42.3	4.83 ± 0.48	1	1
1641+4000	16 41 54.2	+40 00 33.1	1.002	17.81	3575	1.29	311.8	46.5	3.09 ± 0.38	1	1
1657+3524	16 57 13.2	+35 24 39.4	2.328	19.37	3662	8.88	102.8	49.6	1.16 ± 0.14	1	1
1701+6412	17 01 00.6	+64 12 09.0	2.735	16.00	547	4.98	364.7	49.5	3.69 ± 0.21	1	1
1702+3405	17 02 24.5	+34 05 39.0	2.038	18.93	4179	5.25	97.8	57.0	0.65 ± 0.06	1	1
1703+6045	17 03 55.8	+60 45 11.7	0.284	18.77	435	4.97	250.3	9.1	12.0 ± 2.11	1	1
1708+6154	17 08 17.9	+61 54 48.6	1.414	17.84	4864	0.58	239.4	4.1	16.4 ± 1.83	1	1
1719+2732	17 19 27.3	+27 32 46.8	1.446	18.72	3245	10.1	92.7	10.0	7.27 ± 1.29	1	1
1720+2638	17 20 26.5	+26 38 16.0	1.141	19.13	3224, 4361	4.43, 5.43	170.6	49.5	2.66 ± 0.27	1	1
1735+5355	17 35 51.9	+53 55 15.7	0.955	17.91	4863	0.58	268.2	5.4	14.86 ± 1.76	1	1
1737+5828	17 37 16.6	+58 28 39.5	1.775	19.05	3038	3.28	21.2	4.6	3.71 ± 0.83	1	1
1738+5837	17 38 36.2	+58 37 48.6	1.279	17.71	4860	0.58	10.1	3.9	0.76 ± 0.26	1	1
2215-1611	22 15 27.1	-16 11 33.0	3.990	18.1	2185	0.59	16.6	3.2	1.35 ± 0.37	2	6
2238-0921	22 38 19.8	-09 21 06.0	3.259	18.04	2411	6.07	1.3	5.9	< 5.97	1	1
2238-0937	22 38 54.7	-09 37 36.2	1.472	19.14	2411	12.7	4.0	5.8	< 17.5	1	1
2239-0933	22 39 17.3	-09 33 40.9	1.817	19.28	2414	11.3	9.9	5.7	< 14.6	1	1
2249-0808	22 49 03.3	-08 08 41.7	0.457	19.42	583	8.04	304.5	11.7	13.6 ± 1.97	1	1
2337+0025	23 37 18.1	+00 25 50.7	2.053	19.28	3248	7.59	20.6	9.2	1.85 ± 0.40	1	1
2337+0022	23 37 22.0	+00 22 38.9	1.376	19.27	3248	4.50	8.0	9.2	0.97 ± 0.33	1	1
2337+0026	23 37 39.1	+00 26 56.2	1.703	18.85	3248	7.44	61.7	9.2	4.13 ± 0.58	1	1
2348+0107	23 48 40.1	+01 07 53.5	0.718	18.50	861	10.8	779.4	74.2	7.08 ± 0.46	1	1

Note. — For sources with multiple observation IDs, the off-axis angles are reported for each. However, the reported counts and exposure time are summed over the observation IDs.

[a] r-band apparent magnitude.

[b] Off-axis angle, in arcmin.

[c] Number of observed background-subtracted source counts in the range 0.3–7.0 keV.

[d] Unabsorbed 2–10 keV flux, assuming a power-law. A photon index of $\Gamma_X = 1.9$ was assumed for those sources with < 50 counts. A "$<$"denotes an upper limit.

[e] Reference for the observed frame optical spectrum.

[f] Reference for the r-band magnitude.

References. — (1) SDSS; (2) Constantin et al. (2002); (3) Péroux et al. (2001); (4) Anderson et al. (2001); (5) Kennefick et al. (1995); (6) Storrie-Lombardi et al. (1996)
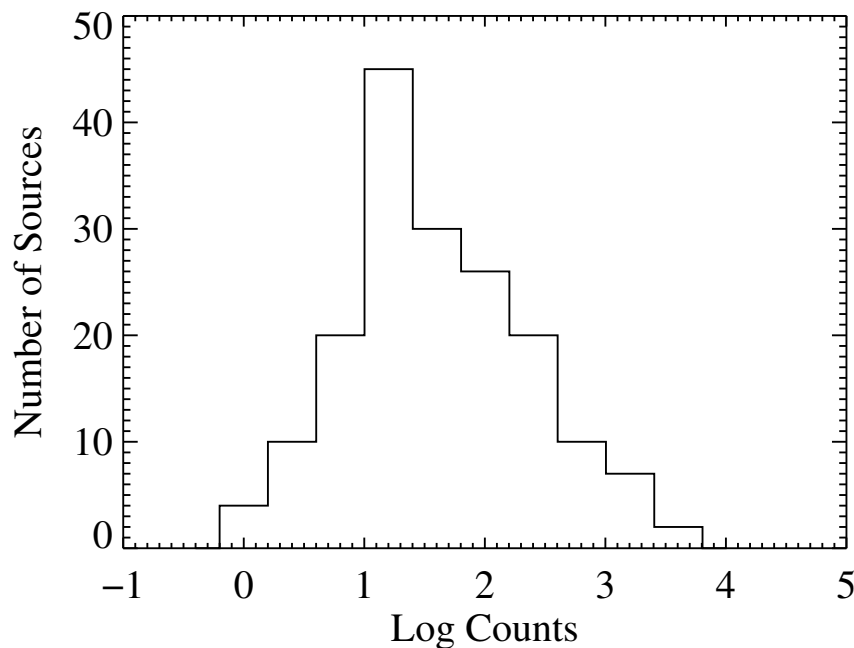
Figure 6.2 The distribution of observed, background-subtracted, 0.3–7.0 keV photon counts for our sample.

6.3.1 X-ray Spectra

Source extraction was done using *CIAO 3.2.2*[2] and *CALDB 3.1*. We extracted the PHA spectrum for all 174 sources using a circular aperture with radius chosen to include 95% of 3.5 keV photons. Typical extraction regions range from 5″ for on-axis sources, to 10″–30″ for most off-axis sources. The background was extracted from an annular region centered on the source, and any nearby sources were removed from the extraction regions. A correction for pileup was necessary for only one source in our sample, MRK 1014. The distribution of source counts is shown in Figure 6.2.

Spectral fitting was done using the *CIAO* tool *SHERPA* (Freeman et al., 2001).

---

[2]Chandra Interactive Analysis of Observations (CIAO), http://cxc.harvard.edu/ciao/

We estimated the parameters for a power law of the form

$$N(E) = n_0 \left(\frac{E}{1 \text{ keV}} \right)^{-\Gamma_X},$$

(6.1)

where Γ_X is the photon index and n_0 is the normalization at 1 keV, in units of photons keV^{-1} cm^{-2} s^{-1}. We restricted our fits to energies 0.3–7.0 keV, and included Galactic absorption with N_H fixed to that inferred from 21 cm maps (COLDEN[3] Dickey & Lockman, 1990). If the source had < 200 counts, we fit the unbinned spectrum using the Cash statistic (Cash, 1979). For these sources, the background was fit simultaneously with Equation (6.1) using an empirically determined background for the ACIS-S and ACIS-I, respectively. If the source had < 50 counts, we calculated n_0 by fixing $\Gamma_X = 1.9$, a typical value for RQQs (Reeves & Turner, 2000; Piconcelli et al., 2003). We also estimate Γ_X for these sources, but fix $\Gamma_X = 1.9$ when calculating n_0 to stabilize the estimates of L_X and α_{ox}. We fit n_0 and Γ_X simultaneously for sources with counts between 50 and 200. If the source had > 200 counts, we fit the binned spectrum by minimizing χ^2 and included an intrinsic neutral absorber if justified by the data; only two of the sources with > 200 counts showed evidence for intrinsic absorption. These were 1438+0335 and 0958+0734, with intrinsic $N_H = 5.34^{+0.745}_{-0.203} \times 10^{21}$ cm^{-2} and $N_H = 1.363^{+2.034}_{-0.626} \times 10^{22}$ cm^{-2}, respectively; the errors are at 95% confidence. The background for the > 200 count sources was binned and subtracted before spectral fitting. There were 86 sources with < 50 counts, 44 sources with between 50 and 200 counts, and 44 sources with > 200 counts.

We included an intrinsic neutral absorber for sources 0259+0048, 0918+5139, 1002+5542, and 1411+5205, despite the fact that they have < 200 counts. These sources initially exhibited unusually hard X-ray spectra ($\Gamma_X \lesssim 1$), so we fit Γ_X and N_H simultaneously to test if these low values of Γ_X were caused by unrecog-

---

[3]For COLDEN, see http://cxc.harvard.edu/toolkit/colden.jsp

nized absorption. Even after including an absorber at the quasar redshift, sources 0259+0048, 0918+5139, and 1002+5542 still have rather hard X-ray continua, with photon induces of 1.02 ± 0.24, 0.92 ± 0.33, and 1.27 ± 0.15, respectively. Such hard X-ray spectra could be the result of more complex absorption, such as an ionized or partial covering absorber. However, the low number of counts for these sources preclude obtaining meaningful spectral fits for more complex models. In addition, the mean Γ_X for our sample is ≈ 2 and the observed dispersion in Γ_X for our sample is 0.44. Thus, we might expect to observe a few sources with $\Gamma_X \sim 1$ in a sample with 157 X-ray detected sources, and therefore these three sources are probably not outliers but just represent the tail of the RQQ Γ_X distribution.

Source 1411+5205 shows evidence for considerable absorption, $N_H \sim 10^{23} \text{ cm}^{-2}$, and may be a BAL QSO. A 95% confidence interval on the column density for this source is $1.8 \times 10^{22} \text{ cm}^{-2} < N_H < 2.45 \times 10^{23} \text{ cm}^{-2}$.

We used the projection method in *Sherpa* to estimate a 3σ confidence interval on n_0. Those sources that did not contain $n_0 = 0$ in their 3σ confidence interval were considered detected, otherwise we set a 3σ upper limit on n_0. For those sources with < 50 counts, the projection method was used to calculate the 68% (1σ) individual confidence intervals on the power law parameters. We calculated the covariance matrix of the parameters for those sources with > 50 counts.

Fifteen sources were in multiple observations. For each of these sources, all of the observations were fit simultaneously assuming the same power-law spectrum.

Based on our X-ray spectral fits, we estimate the mean value of Γ_X for our sample to be $\bar{\Gamma}_X = 2.033 \pm 0.034$, with an observed dispersion of 0.44. After accounting for the additional scatter in Γ_X caused by measurement error, we find that the intrinsic dispersion of Γ_X is ~ 0.31. This is consistent with the dispersion

found in other studies (e.g., Piconcelli et al., 2005; Brocksopp et al., 2006; Grupe et al., 2006). The intrinsic dispersion in Γ_X is similar to the intrinsic dispersion in the optical/UV spectral slope (Richards et al., 2001).

6.3.2   Optical/UV Spectra

Optical spectra were obtained for most sources from the SDSS. We also obtained spectra for some of the high redshift quasars from Anderson et al. (2001), Péroux et al. (2001), and Constantin et al. (2002).

We corrected the optical spectra for Galactic absorption using the $E(B-V)$ values taken from Schlegel et al. (1998), as listed in the NASA/IPAC Extragalactic Database (NED), and the extinction curve of Cardelli et al. (1989), assuming a value of $A_V/E(B-V) = 3.1$. We model the continuum as a power law of the form $f_\nu \propto \nu^{-\alpha}$, and the Fe emission as a scaled and broadened iron template extracted from I Zw I in the UV by Vestergaard & Wilkes (2001). The continuum and iron emission were fit simultaneously using the Levenberg-Marquardt method for nonlinear χ^2-minimization. The median value of α for our sample is 0.602, and the dispersion in α is ≈ 0.4.

We were not able to obtain a spectrum for Q 0910+564. For this source, we calculated the flux density at $2500\mathring{A}$ from the AB magnitude at $1450(1+z)\mathring{A}$ and the spectral index reported by (Schneider et al., 1991).

We were not able to use a power-law fit to calculate l_{UV} for the $z < 0.3$ sources, as the SDSS spectral range for these sources does not contain the rest-frame UV continuum. We therefore performed a linear regression of the dependence of l_{UV} on $\log \nu L_\nu(5100\mathring{A})$ and α_{opt} for higher redshift sources ($0.3 < z < 1.2$) for which we had all three quantities. Here α_{opt} is the spectral index of the optical continuum. Using the regression results, l_{UV} was then estimated for the $z < 0.3$ sources based on their optical luminosity and spectral index. The optical continuum parameters

were found in the same manner as for the UV continuum, except that we used the Fe emission template from Véron-Cetty et al. (2004). For sources within $0.3 < z < 1.2$ the scatter about the regression fit resulted in a 'measurement' error on l_{UV} of ≈ 0.07 dex. For comparison, typical measurement errors on l_{UV} for the $z > 0.3$ sources are ≈ 0.001–0.01 dex, ignoring variability.

6.3.3 α_{ox}

We calculate the ratio of optical to X-ray flux (Tananbaum et al., 1979) as

$$\alpha_{\mathrm{ox}} = -\frac{\log(f_X/f_{UV})}{\log(\nu_X/\nu_{UV})}, \tag{6.2}$$

where f_X and f_{UV} are the rest-frame flux densities at 2 keV and $2500\mathring{A}$, respectively. If the flux density from $2500\mathring{A}$ to 2 keV is a simple power law, then α_{ox} is the spectral slope of this continuum, and thus α_{ox} may be thought of as a crude estimate of the shape of the ionizing continuum. The parameter α_{ox} is an important parameter for model comparison, as it summarizes the amount of energy emitted in the X-ray region (most likely a Comptonized component), compared with that emitted in the optical-UV (accretion disk component). The mean α_{ox} of our sample is $\bar{\alpha}_{\mathrm{ox}} = 1.49 \pm 0.01$, and the dispersion of α_{ox} is estimated to be $\sigma_{\mathrm{ox}} \approx 0.19$. Because some of the data points are censored, these estimates of the mean and dispersion of α_{ox} were obtained by maximum-likelihood assuming a normal density. The Kaplan-Meier estimate of the mean (e.g., Feigelson & Nelson, 1985), a non-parametric estimate, gives $\bar{\alpha}_{\mathrm{ox}} = 1.49 \pm 0.06$.

In Figure 6.3 we show the distributions of l_X and α_{ox} as functions of l_{UV} and z. We report the X-ray and UV parameters in Table 6.3.
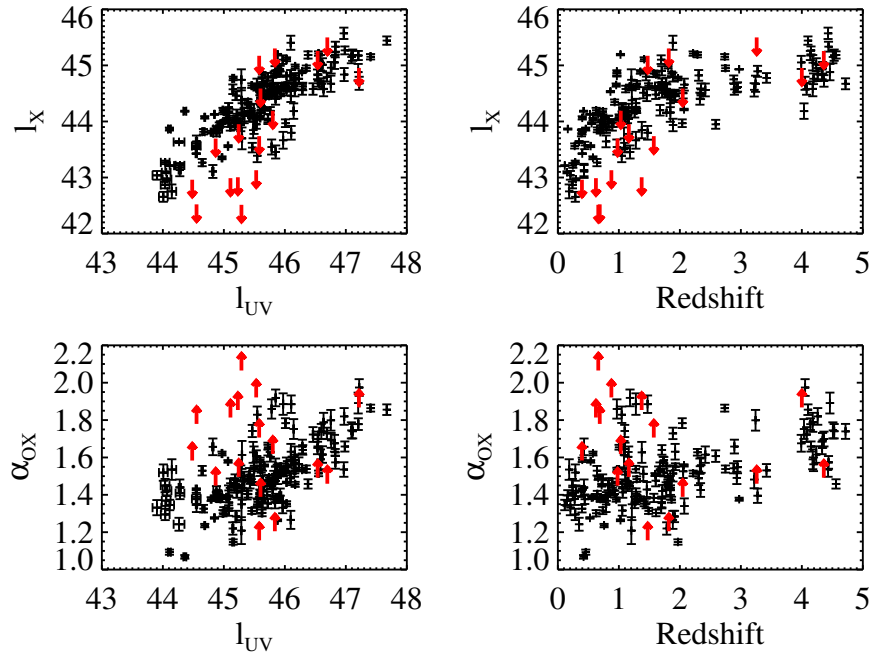
Figure 6.3 The distribution of l_X and α_{ox} as functions of l_{UV} and z. Red arrows denote upper limits for l_X and lower limits of α_{ox}.

Table 6.3.   X-ray and UV Parameters

Source	Redshift	Gal. N_H [a] 10^{20} cm^{-2}	n_0 [b] 10^{-6} cm^{-2} s^{-1} keV^{-1}	Γ_X [c]	$\log \nu L_\nu (2500\text{Å})$ ergs s^{-1}	$\log \nu L_\nu (2\text{ keV})$ ergs s^{-1}	$\alpha_{\rm ox}$	α_{UV} [d]
0002+0049	1.353	2.82	41.87 ± 3.08	2.20 ± 0.11	45.87	44.88	1.38	1.33
0006-0015	1.725	3.16	$13.26^{+2.04}_{-1.85}$	$1.81^{+0.22}_{-0.22}$	46.12	44.61	1.58	0.74
0022+0016	0.575	2.70	11.40 ± 0.88	2.30 ± 0.20	44.97	43.35	1.62	0.52
0027+0026	0.205	3.01	47.63 ± 7.30	1.68 ± 0.27	43.90	43.04	1.33	. . .
0031+0034	1.735	2.41	12.24 ± 1.49	2.21 ± 0.18	45.98	44.62	1.52	0.15
0050-0053[e]	4.332	2.70	$2.66^{+0.56}_{-0.50}$	$2.02^{+0.33}_{-0.32}$	46.54	44.84	1.65	0.98
0057+1450	0.624	4.35	< 3.69	. . .	45.11	< 42.99	> 1.81	0.64
0059+0003	4.178	3.01	$4.65^{+1.62}_{-1.33}$	$0.91^{+0.44}_{-0.44}$	46.58	45.05	1.59	1.12
0106+0048	4.437	3.16	$7.85^{+1.70}_{-1.49}$	$1.95^{+0.31}_{-0.30}$	46.83	45.33	1.57	0.60
0113+1531	0.576	4.40	50.18 ± 3.63	1.91 ± 0.08	45.05	44.04	1.39	0.53
0113+1535	1.807	4.38	24.49 ± 1.37	2.29 ± 0.10	46.17	44.98	1.46	0.53
0115+0020	1.276	3.34	26.95 ± 1.77	1.92 ± 0.11	45.55	44.61	1.36	0.64
0133+0400	4.150	3.01	$7.84^{+1.34}_{-1.21}$	$2.16^{+0.28}_{-0.27}$	46.98	45.27	1.66	1.00
0134+3307	4.530	4.67	$5.29^{+1.31}_{-1.14}$	$2.24^{+0.42}_{-0.40}$	47.11	45.18	1.74	0.65
0148+0001	1.705	2.88	$14.12^{+2.29}_{-2.07}$	$2.39^{+0.24}_{-0.24}$	46.29	44.62	1.64	0.16
0148-0002	0.930	2.75	31.91 ± 3.49	2.31 ± 0.19	45.40	44.34	1.40	0.54
0152+0105	0.647	2.80	27.70 ± 2.75	2.50 ± 0.18	44.90	43.85	1.40	0.18
0153+0052	1.161	2.69	9.06 ± 1.44	2.19 ± 0.26	45.49	44.05	1.56	0.59
0156+0053	1.652	2.69	18.15 ± 2.12	1.36 ± 0.15	46.01	44.63	1.53	1.38
0159+0023	0.163	2.13	775.4 ± 12.4	2.39 ± 0.03	44.90	43.86	1.40	. . .
0201-0919	0.661	2.08	46.05 ± 2.87	2.16 ± 0.12	45.40	44.12	1.49	0.90

Table 6.3—Continued

Source	Redshift	Gal. N_H [a] 10^{20} cm^{-2}	n_0 [b] 10^{-6} cm^{-2} s^{-1} keV^{-1}	Γ_X [c]	$\log \nu L_\nu (2500\text{Å})$ ergs s^{-1}	$\log \nu L_\nu (2 \text{ keV})$ ergs s^{-1}	α_{ox}	α_{UV} [d]
0208+0022	1.885	2.78	18.49 ± 2.65	1.56 ± 0.20	46.69	44.79	1.73	0.49
0209+0517	4.140	4.57	$6.39^{+1.27}_{-1.12}$	$2.72^{+0.33}_{-0.31}$	47.20	45.18	1.78	0.68
0232-0731	1.164	3.30	< 7.39	. . .	45.24	< 43.95	> 1.50	1.41
0241-0811	0.979	2.90	< 6.24	. . .	44.87	< 43.70	> 1.45	0.63
0241+0023	0.790	3.37	$9.46^{+2.45}_{-2.12}$	$2.01^{+0.48}_{-0.44}$	44.55	43.64	1.35	0.78
0244-0134	4.053	3.52	$2.14^{+0.62}_{-0.53}$	$1.52^{+0.41}_{-0.39}$	47.22	44.68	1.97	1.06
0248+1802	4.430	9.02	$13.64^{+3.43}_{-2.94}$	$1.97^{+0.39}_{-0.37}$	46.97	45.57	1.54	0.34
0259+0048	0.892	7.22	$12.56^{+2.18}_{-1.94}$	$1.02^{+-1.0}_{-1.02}$	45.11	43.90	1.46	0.58
0311-1722	4.000	3.81	< 4.14	. . .	47.21	< 44.95	> 1.87	1.90
0314-0111	0.387	5.78	97.12 ± 9.62	2.75 ± 0.17	44.88	43.79	1.42	. . .
0403-1703	4.236	2.29	$4.16^{+1.20}_{-1.01}$	$1.39^{+0.37}_{-0.37}$	46.62	45.01	1.62	1.26
0419-5716	4.460	1.67	$2.73^{+1.01}_{-0.82}$	$2.13^{+0.51}_{-0.49}$	47.20	44.88	1.89	2.18
0755+2203	0.400	5.60	23.69 ± 2.84	1.38 ± 0.10	44.55	43.43	1.43	. . .
0755+4058	2.417	4.78	$5.36^{+1.52}_{-1.31}$	$3.22^{+0.57}_{-0.55}$	46.16	44.56	1.62	0.27
0755+4111	0.967	4.98	$4.53^{+1.51}_{-1.27}$	$2.09^{+0.61}_{-0.55}$	45.78	43.54	1.86	0.20
0755+4056	2.348	4.73	$5.86^{+1.55}_{-1.34}$	$2.75^{+0.56}_{-0.53}$	46.03	44.57	1.56	0.45
0819+3649	0.736	4.82	45.20 ± 5.24	3.08 ± 0.26	45.00	44.18	1.32	0.38
0832+5243	1.573	3.87	< 2.23	. . .	45.58	< 43.74	> 1.71	0.30
0845+3431	2.046	3.41	< 8.40	. . .	45.60	< 44.59	> 1.39	0.62
0849+4457	1.259	2.75	9.43 ± 0.37	2.07 ± 0.07	45.08	44.15	1.36	0.38
0849+4500	1.592	2.70	15.33 ± 0.57	2.10 ± 0.06	45.94	44.61	1.51	0.53

<center>Table 6.3—Continued</center>

Source	Redshift	Gal. N_H[a] 10^{20} cm^{-2}	n_0[b] 10^{-6} cm^{-2} s^{-1} keV^{-1}	Γ_X[c]	$\log \nu L_\nu (2500\text{Å})$ ergs s^{-1}	$\log \nu L_\nu (2\text{ keV})$ ergs s^{-1}	α_{ox}	α_{UV}[d]
0910+5427	0.526	2.03	80.75 ± 1.73	2.11 ± 0.04	44.85	44.13	1.28	-0.3
0912+0547	3.241	3.65	3.40 ± 0.56	1.94 ± 0.31	46.75	44.67	1.80	0.67
0914+5613[e]	4.035	2.89	$0.69^{+0.24}_{-0.20}$	$1.04^{+0.55}_{-0.52}$	46.14	44.18	1.75	1.43
0918+5139	0.185	1.36	20.66 ± 9.26	0.92 ± 0.33	44.15	42.75	1.54	...
0918+0647	0.821	3.65	22.58 ± 3.32	2.81 ± 0.28	45.21	44.03	1.45	0.68
0933+5515	1.864	2.00	18.05 ± 1.18	2.08 ± 0.11	45.83	44.85	1.37	0.81
0941+5948	0.968	2.19	100.8 ± 5.90	2.44 ± 0.12	46.28	44.89	1.54	0.48
0950+5619	1.913	1.19	$7.36^{+1.40}_{-1.25}$	$2.87^{+0.38}_{-0.36}$	45.20	44.46	1.28	0.68
0951+5940	1.057	1.67	$17.92^{+2.76}_{-2.51}$	$1.80^{+0.25}_{-0.24}$	45.43	44.23	1.46	0.55
0951+5944	2.339	1.51	$4.95^{+1.17}_{-1.01}$	$2.01^{+0.38}_{-0.36}$	45.60	44.49	1.43	0.21
0952+5152	0.554	0.88	51.81 ± 1.52	2.29 ± 0.06	44.97	43.97	1.38	0.60
0952+5151	0.862	0.86	59.38 ± 1.62	2.16 ± 0.05	45.59	44.53	1.41	0.88
0955+5935	0.912	1.50	$14.05^{+2.37}_{-2.13}$	$2.20^{+0.30}_{-0.30}$	45.08	43.97	1.42	0.69
0955+5940	4.340	1.33	$3.47^{+0.89}_{-0.76}$	$1.87^{+0.34}_{-0.33}$	46.76	44.96	1.69	0.47
0955+4116	3.420	0.64	$3.93^{+0.83}_{-0.73}$	$2.15^{+0.38}_{-0.36}$	46.16	44.78	1.53	0.23
0955+4109	2.308	0.59	9.68 ± 1.36	2.13 ± 0.22	46.09	44.82	1.49	0.25
0956+4110	1.887	0.72	$2.28^{+0.63}_{-0.54}$	$2.50^{+0.53}_{-0.50}$	45.28	43.94	1.51	-0.5
0958+0734	1.885	2.95	59.53 ± 11.9	2.22 ± 0.26	46.10	45.40	1.27	0.72
0958+0747	3.219	2.97	$2.77^{+0.82}_{-0.70}$	$2.56^{+0.57}_{-0.54}$	45.82	44.56	1.48	0.44
0958+0745	1.488	3.05	18.66 ± 1.47	2.20 ± 0.19	45.46	44.64	1.31	0.72
1002+5542	1.151	0.84	13.75 ± 2.00	1.27 ± 0.15	45.76	44.19	1.60	1.04

Table 6.3—Continued

Source	Redshift	Gal. N_H[a] 10^{20} cm^{-2}	n_0[b] 10^{-6} cm^{-2} s^{-1} keV^{-1}	Γ_X[c]	$\log \nu L_\nu (2500\text{Å})$ ergs s^{-1}	$\log \nu L_\nu (2\text{ keV})$ ergs s^{-1}	α_{ox}	α_{UV}[d]
1003+4736	2.934	0.93	4.83 ± 0.64	1.53 ± 0.17	45.88	44.60	1.49	0.52
1013-0052	0.276	3.49	133.5 ± 10.1	2.74 ± 0.12	44.55	43.56	1.38	. . .
1019+4737	2.945	0.99	$3.82^{+0.78}_{-0.69}$	$2.31^{+0.28}_{-0.27}$	46.04	44.61	1.55	0.53
1023+0415	1.809	2.89	8.66 ± 0.68	2.05 ± 0.11	45.65	44.50	1.45	0.55
1030+0524	1.183	2.72	$2.27^{+0.67}_{-0.56}$	$1.21^{+0.40}_{-0.39}$	45.85	43.46	1.92	0.48
1032+5738	1.969	0.59	14.95 ± 0.81	1.73 ± 0.07	45.15	44.77	1.15	0.73
1032+5800	0.687	0.61	< 1.00	. . .	44.55	< 42.52	> 1.78	0.57
1036-0343	4.509	4.77	$4.96^{+1.38}_{-1.18}$	$2.67^{+0.50}_{-0.45}$	47.09	45.15	1.74	0.62
1038+4727	1.047	1.47	$2.61^{+0.74}_{-0.64}$	$0.92^{+0.40}_{-0.40}$	45.55	43.39	1.83	0.69
1042+0100	1.401	3.95	$16.10^{+3.88}_{-3.36}$	$2.25^{+0.36}_{-0.35}$	45.85	44.51	1.51	0.61
1044+5921	1.292	0.70	10.00 ± 0.80	1.85 ± 0.11	45.45	44.19	1.48	0.41
1049+5750	1.106	0.60	$11.66^{+2.65}_{-2.31}$	$1.79^{+0.40}_{-0.38}$	45.42	44.10	1.51	0.72
1050+5702	3.273	0.59	$5.70^{+1.38}_{-1.20}$	$2.45^{+0.49}_{-0.46}$	45.92	44.89	1.40	0.81
1050+5738	1.281	0.59	$13.89^{+2.61}_{-2.32}$	$2.68^{+0.36}_{-0.35}$	45.38	44.33	1.41	0.53
1052+5724	1.112	0.60	35.22 ± 4.58	1.80 ± 0.18	45.76	44.58	1.45	0.72
1053+5735	1.205	0.59	42.26 ± 4.19	1.86 ± 0.15	45.29	44.74	1.21	1.37
1054+5740	1.101	0.59	$10.45^{+2.30}_{-2.01}$	$2.12^{+0.37}_{-0.36}$	45.72	44.04	1.64	0.54
1054+5720	2.972	0.59	< 7.75	. . .	45.91	< 44.93	> 1.38	0.62
1055+5704	0.696	0.60	25.93 ± 3.29	2.12 ± 0.20	44.92	43.93	1.38	0.87
1056+5722	0.286	0.60	$10.45^{+2.24}_{-1.97}$	$0.28^{+0.32}_{-0.33}$	44.01	42.66	1.52	. . .
1057+4555	4.100	1.16	$12.05^{+2.17}_{-1.94}$	$2.08^{+0.27}_{-0.27}$	47.67	45.44	1.86	2.14

Table 6.3—Continued

Source	Redshift	Gal. N_H[a] 10^{20} cm^{-2}	n_0[b] 10^{-6} cm^{-2} s^{-1} keV^{-1}	Γ_X[c]	$\log \nu L_\nu(2500\text{Å})$ ergs s^{-1}	$\log \nu L_\nu(2\text{ keV})$ ergs s^{-1}	$\alpha_{\rm ox}$	α_{UV}[d]
1109+0900	1.674	2.79	$5.56^{+1.29}_{-1.13}$	$1.67^{+0.38}_{-0.36}$	45.60	44.20	1.54	0.85
1111+5532	1.004	0.78	14.82 ± 0.91	3.23 ± 0.10	45.50	44.10	1.54	0.71
1114+5315	1.213	0.96	18.34 ± 2.19	1.65 ± 0.16	45.29	44.38	1.35	2.11
1115+5309	0.877	0.97	< 2.21	$\ldots$	45.53	< 43.13	> 1.92	0.56
1129-0137	1.295	3.58	31.69 ± 3.31	1.54 ± 0.15	45.71	44.67	1.40	0.75
1129-0150	1.785	3.56	$11.30^{+2.37}_{-2.09}$	$2.11^{+0.37}_{-0.35}$	45.15	44.57	1.22	0.97
1136+0159	0.766	2.50	21.23 ± 2.44	2.03 ± 0.19	44.94	43.96	1.38	0.73
1136+0158	1.471	2.61	$2.95^{+0.89}_{-0.75}$	$1.45^{+0.39}_{-0.40}$	46.10	43.80	1.88	0.57
1136+0207	0.239	2.61	22.37 ± 4.05	1.63 ± 0.26	44.04	42.87	1.45	$\ldots$
1202-0129	0.150	2.22	222.5 ± 5.86	3.02 ± 0.04	44.05	43.09	1.37	$\ldots$
1204+0150	1.927	1.88	7.29 ± 0.63	2.10 ± 0.15	45.93	44.50	1.55	0.29
1208+0016	1.063	1.99	$14.11^{+2.56}_{-2.29}$	$1.25^{+0.28}_{-0.27}$	45.21	44.14	1.41	0.92
1213+0252	0.641	1.74	6.82 ± 0.99	2.10 ± 0.25	44.64	43.27	1.53	0.98
1214+0055	0.396	1.95	87.02 ± 7.77	2.44 ± 0.13	44.67	43.82	1.33	$\ldots$
1215-0034	0.758	2.07	28.80 ± 1.72	2.08 ± 0.11	44.69	44.07	1.24	0.86
1218+0546	0.795	1.57	14.21 ± 2.11	1.28 ± 0.20	44.88	43.86	1.39	1.73
1220-0025	0.421	2.02	123.6 ± 2.93	1.45 ± 0.03	44.36	44.18	1.07	$\ldots$
1223+1034	2.747	2.13	11.80 ± 0.81	1.59 ± 0.11	46.35	44.95	1.54	0.27
1226-0011	1.175	1.93	32.20 ± 2.91	2.43 ± 0.13	45.79	44.62	1.45	1.28
1228+4413	0.662	1.45	< 1.07	$\ldots$	45.29	< 42.51	> 2.07	0.59
1228+4411	1.277	1.34	10.81 ± 1.18	2.12 ± 0.18	45.61	44.21	1.54	0.12

Table 6.3—Continued

Source	Redshift	Gal. N_H[a] 10^{20} cm^{-2}	n_0[b] 10^{-6} cm^{-2} s^{-1} keV^{-1}	Γ_X[c]	log $\nu L_\nu(2500\text{Å})$ ergs s^{-1}	log $\nu L_\nu(2\text{ keV})$ ergs s^{-1}	α_{ox}	α_{UV}[d]
1230+0302	1.605	1.80	$11.27^{+2.17}_{-1.93}$	$1.61^{+0.29}_{-0.28}$	45.72	44.48	1.48	0.40
1230+0305	1.056	1.83	$18.88^{+3.37}_{-3.02}$	$2.37^{+0.32}_{-0.31}$	45.08	44.26	1.32	0.46
1230+0306	0.628	1.81	59.90 ± 6.60	2.03 ± 0.17	44.99	44.20	1.30	0.33
1230+0308	1.843	1.81	$6.85^{+2.35}_{-1.92}$	$2.61^{+0.59}_{-0.57}$	45.63	44.39	1.48	0.29
1230+0305	0.428	1.83	$12.38^{+3.41}_{-2.92}$	$1.73^{+0.52}_{-0.48}$	44.28	43.13	1.44	. . .
1236+6215	2.587	1.52	1.19 ± 0.12	1.79 ± 0.19	45.53	43.95	1.61	0.55
1237+6203	2.068	1.44	3.54 ± 0.08	1.85 ± 0.03	45.66	44.21	1.56	1.02
1242+0249	1.459	1.92	12.89 ± 0.58	2.32 ± 0.08	45.58	44.47	1.43	0.06
1245-0027	1.693	1.73	44.48 ± 6.08	1.80 ± 0.19	45.96	45.10	1.33	0.72
1255+5652	1.804	1.25	7.06 ± 0.48	2.44 ± 0.11	45.74	44.46	1.49	0.34
1255+5650	1.374	1.25	< 0.56	. . .	45.23	< 43.00	> 1.85	0.28
1258-0143	0.967	1.54	84.49 ± 2.06	2.30 ± 0.05	46.03	44.81	1.47	0.56
1259+0102	0.395	1.62	< 10.05	. . .	44.48	< 42.96	> 1.58	. . .
1311+0031	0.429	1.84	$11.60^{+3.33}_{-2.80}$	$2.58^{+0.46}_{-0.44}$	44.82	43.11	1.66	. . .
1317+3531	4.360	0.99	< 6.84	. . .	46.54	< 45.26	> 1.49	0.90
1321+0038[e]	4.716	1.88	$1.49^{+0.36}_{-0.31}$	$2.50^{+0.38}_{-0.37}$	46.60	44.67	1.74	2.04
1344-0000	1.096	1.89	$3.27^{+0.68}_{-0.60}$	$1.80^{+0.32}_{-0.30}$	45.47	43.53	1.74	0.43
1411+5217	2.883	1.33	5.79 ± 0.50	1.82 ± 0.14	46.16	44.75	1.54	0.57
1411+5205	1.084	1.40	4.80 ± 2.07	2.44 ± 0.49	45.30	43.70	1.61	0.73
1413+0000[e]	4.078	3.15	$1.67^{+0.42}_{-0.37}$	$1.93^{+0.43}_{-0.39}$	46.46	44.58	1.72	1.66
1417+4456	0.114	1.13	380.7 ± 5.53	2.38 ± 0.03	44.11	43.21	1.35	. . .

Table 6.3—Continued

Source	Redshift	Gal. N_H[a] 10^{20} cm^{-2}	n_0[b] 10^{-6} cm^{-2} s^{-1} keV^{-1}	Γ_X[c]	$\log \nu L_\nu (2500\text{Å})$ ergs s^{-1}	$\log \nu L_\nu (2\text{ keV})$ ergs s^{-1}	α_{ox}	α_{UV}[d]
1419+4709	2.288	1.56	26.56 ± 2.06	1.85 ± 0.12	46.83	45.19	1.63	1.25
1424+4214	1.608	1.25	26.00 ± 2.42	1.97 ± 0.15	45.63	44.84	1.30	-0.0
1424+4210	2.218	1.25	23.20 ± 2.15	2.39 ± 0.12	46.55	45.21	1.51	0.35
1432-0059	1.027	3.39	174.2 ± 5.15	2.06 ± 0.05	45.88	45.19	1.26	0.78
1433+0227	2.072	2.75	$9.59^{+2.54}_{-2.16}$	$2.36^{+0.46}_{-0.44}$	45.54	44.66	1.34	0.75
1434+0227	1.711	2.76	$10.73^{+2.19}_{-1.94}$	$2.05^{+0.34}_{-0.32}$	45.60	44.51	1.42	0.08
1438+0341	1.737	2.62	12.71 ± 0.93	2.11 ± 0.13	46.15	44.63	1.59	1.22
1438+0335	0.734	2.62	20.96 ± 3.70	1.70 ± 0.17	45.18	43.92	1.48	0.88
1442+0110	4.560	3.36	$5.25^{+0.82}_{-0.74}$	$1.98^{+0.23}_{-0.23}$	46.38	45.19	1.46	0.71
1443+5856	4.260	1.56	$2.74^{+0.80}_{-0.67}$	$2.29^{+0.42}_{-0.40}$	46.97	44.84	1.82	0.29
1444-0123[e]	4.179	4.03	$1.59^{+0.53}_{-0.44}$	$2.95^{+0.77}_{-0.66}$	46.63	44.58	1.79	1.60
1445+0129	1.846	3.48	$8.75^{+1.42}_{-1.28}$	$2.45^{+0.29}_{-0.28}$	45.61	44.50	1.43	0.17
1448+4738	2.894	2.05	$3.61^{+0.86}_{-0.74}$	$1.72^{+0.32}_{-0.31}$	46.29	44.57	1.66	1.12
1448+0015	0.832	3.58	$18.64^{+4.73}_{-4.05}$	$2.24^{+0.48}_{-0.45}$	45.17	44.00	1.45	0.60
1449+0024	0.441	3.58	$31.72^{+5.06}_{-4.58}$	$2.50^{+0.27}_{-0.27}$	44.55	43.57	1.38	. . .
1452+4304	0.296	1.69	20.43 ± 0.99	1.97 ± 0.07	44.07	42.97	1.42	. . .
1452+4308	1.704	1.64	17.64 ± 1.90	1.78 ± 0.15	45.63	44.70	1.36	0.20
1453+3352[e]	4.191	1.22	$2.39^{+0.56}_{-0.49}$	$1.38^{+0.35}_{-0.34}$	46.23	44.76	1.56	2.05
1511+5659	1.031	1.54	< 17.19	. . .	45.80	< 44.19	> 1.62	0.61
1515+5521	1.844	1.44	$4.02^{+1.63}_{-1.44}$	$4.56^{+0.90}_{-0.90}$	45.30	44.16	1.44	0.33
1539+4313	0.348	2.03	64.94 ± 2.34	1.96 ± 0.06	44.27	43.63	1.24	. . .

Table 6.3—Continued

Source	Redshift	Gal. N_H[a] 10^{20} cm^{-2}	n_0[b] 10^{-6} cm^{-2} s^{-1} keV^{-1}	Γ_X[c]	$\log \nu L_\nu(2500\text{Å})$ ergs s^{-1}	$\log \nu L_\nu(2\text{ keV})$ ergs s^{-1}	α_{ox}	α_{UV}[d]
1543+5405	0.245	1.31	58.57 ± 5.41	2.08 ± 0.15	44.28	43.22	1.41	. . .
1545+4846	0.400	1.61	139.3 ± 10.9	2.23 ± 0.13	45.59	44.06	1.59	. . .
1605-0109	1.573	8.88	28.07 ± 3.18	2.03 ± 0.18	45.85	44.85	1.38	0.62
1618+3456	1.922	1.46	$7.21^{+1.96}_{-1.66}$	$1.76^{+0.44}_{-0.42}$	46.10	44.46	1.63	1.60
1640+4644	0.537	1.74	20.91 ± 0.73	2.13 ± 0.06	45.07	43.56	1.58	0.36
1641+4649	0.695	1.77	19.47 ± 1.22	2.11 ± 0.11	44.79	43.81	1.38	0.67
1641+4000	1.003	1.02	12.38 ± 0.95	1.74 ± 0.12	45.72	44.02	1.65	0.63
1657+3524	2.329	1.75	4.55 ± 0.68	1.69 ± 0.24	45.85	44.41	1.55	0.45
1701+6412	2.736	2.59	14.40 ± 0.93	2.04 ± 0.10	47.41	45.15	1.86	0.37
1702+3405	2.038	2.04	2.55 ± 0.29	1.37 ± 0.16	46.01	43.97	1.78	1.11
1703+6045	0.285	2.32	49.18 ± 3.43	2.18 ± 0.13	44.04	43.27	1.30	. . .
1708+6154	1.415	2.49	65.18 ± 4.69	2.00 ± 0.14	46.03	45.11	1.35	1.15
1719+2732	1.447	3.68	28.89 ± 3.64	2.22 ± 0.21	45.72	44.80	1.35	1.11
1720+2638	1.141	3.86	10.63 ± 0.94	2.40 ± 0.15	45.33	44.10	1.47	0.20
1735+5355	0.956	3.39	59.55 ± 3.96	1.98 ± 0.12	45.62	44.65	1.37	1.01
1737+5828	1.776	3.51	$14.65^{+3.51}_{-3.03}$	$2.39^{+0.40}_{-0.38}$	45.80	44.68	1.43	0.79
1738+5837	1.279	3.59	$3.04^{+1.16}_{-0.94}$	$2.23^{+0.55}_{-0.53}$	45.97	43.66	1.89	1.41
2215-1611	3.990	2.65	$5.23^{+1.53}_{-1.29}$	$1.30^{+0.38}_{-0.38}$	46.81	45.05	1.68	0.80
2238-0921	3.259	4.64	< 23.19	. . .	46.70	< 45.50	> 1.46	0.31
2238-0937	1.472	4.78	< 69.50	. . .	45.58	< 45.17	> 1.16	1.07
2239-0933	1.818	4.63	< 57.74	. . .	45.84	< 45.30	> 1.21	0.63

6.4 DEPENDENCE OF α_{ox} ON L_{UV} AND z

6.4.1 Regression Results

In order to study the relationship between α_{ox}, optical/UV luminosity, and red-shift, we performed a multivariate regression of α_{ox} on z and l_{UV}. In our analysis, we perform the regression using several different parameteric models for the redshift and l_{UV} dependencies. We compare the different models simultaneously using the Kullback-Leibler information (KLI; Kullback & Leibler 1951), a well-studied method for comparing data to models from information theory. In Appendix D, we describe KLI minimization in detail, and compare this approach to classical statistical methods for testing for significance.

Currently, there is no *a priori* reason to assume a certain parameteric form for a dependence α_{ox} on L_{UV} and z. Initially, we are interested in testing for the existence of a dependence of α_{ox} on redshift and on UV luminosity, and we are not concerned with the particular parameteric forms of the possible redshift and luminosity dependencies. As described in Appendix D, the KLI is a particularly powerful tool for comparing and testing several different parameteric models simultaneously, and is valid even if the 'correct' parameterization is not among those considered.

We compare models with redshift dependencies of the form $L_X \propto e^{-t(z)/t_0}$, $L_X \propto e^{z/z_0}$, and $L_X \propto (1+z)^{\beta_\zeta}$, and L_{UV} dependencies of the form $L_X \propto L_{UV}^{\beta_l}$. Here, $t(z)$ is the age of the universe at z in units of Gyr. In addition, Steffen et al. (2006) found some evidence for a nonlinear dependence of α_{ox} on l_{UV}, and to test this we include models that contain a quadratic term for l_{UV}.

We also tested for including the UV spectral slope, α_{UV}, as one of the independent variables. However, we found no evidence that α_{ox} depended on α_{UV}.

Each of the statistical models considered here may be expressed as a normal

Table 6.3—Continued

Source	Redshift	Gal. N_H[a]	n_0[b]	Γ_X[c]	$\log \nu L_\nu(2500\text{Å})$	$\log \nu L_\nu(2\,\text{keV})$	α_{ox}	α_{UV}[d]
		10^{20} cm^{-2}	10^{-6} cm^{-2} s^{-1} keV^{-1}		ergs s^{-1}	ergs s^{-1}		
2249-0808	0.457	3.46	55.44 ± 4.29	1.81 ± 0.11	44.11	43.86	1.09	. . .
2337+0025	2.054	3.81	$6.89^{+1.59}_{-1.39}$	$2.29^{+0.47}_{-0.43}$	45.87	44.50	1.53	0.18
2337+0022	1.376	3.30	$3.87^{+1.46}_{-1.18}$	$2.93^{+0.70}_{-0.65}$	45.48	43.85	1.63	0.99
2337+0026	1.703	3.80	21.37 ± 3.03	2.67 ± 0.28	45.84	44.90	1.36	0.15
2348+0107	0.718	3.98	28.51 ± 1.28	2.05 ± 0.08	45.19	44.01	1.45	0.32
2357+0043[e]	4.362	3.33	$2.00^{+0.51}_{-0.44}$	$1.58^{+0.40}_{-0.38}$	46.54	44.72	1.70	1.74

Note. — Quoted errors are at 68% (1σ) confidence.

[a]Galactic N_H, inferred from COLDEN (Dickey & Lockman, 1990).

[b]When fitting n_0 for the sources with < 50 counts, we fix $\Gamma_X = 1.9$.

[c]The photon index could not be estimated for those sources with upper limits.

[d]The UV spectral slope could not be estimated for sources with $z < 0.5$.

[e]One of seven new *Chandra* observations.

density with variance σ^2 and mean $E(\alpha_{\mathrm{ox}}) = \bar{\alpha}_{\mathrm{ox}}(\gamma)$. Here, $E(\alpha_{\mathrm{ox}})$ is the expectation value of α_{ox} at a given l_{UV} and z, and γ denotes the regression coefficients. The five models considered differ in their description of $\bar{\alpha}_{\mathrm{ox}}(\gamma)$:

$$\mathcal{M}_z : \bar{\alpha}_{\mathrm{ox}}(\gamma) = \gamma_0 + \gamma_l l_{UV} + \gamma_z z \tag{6.3}$$

$$\mathcal{M}_\zeta : \bar{\alpha}_{\mathrm{ox}}(\gamma) = \gamma_0 + \gamma_l l_{UV} + \gamma_\zeta \log(1 + z) \tag{6.4}$$

$$\mathcal{M}_t : \bar{\alpha}_{\mathrm{ox}}(\gamma) = \gamma_0 + \gamma_l l_{UV} + \gamma_t t(z) \tag{6.5}$$

$$\mathcal{M}_l : \bar{\alpha}_{\mathrm{ox}}(\gamma) = \gamma_0 + \gamma_l l_{UV} + \gamma_{l^2} l_{UV}^2 \tag{6.6}$$

$$\mathcal{M}_{l+t} : \bar{\alpha}_{\mathrm{ox}}(\gamma) = \gamma_0 + \gamma_l l_{UV} + \gamma_{l^2} l_{UV}^2 + \gamma_t t(z). \tag{6.7}$$

Here, we have introduced the notation that $\mathcal{M}_z$ stands for the model that parameterizes the average value of α_{ox} as depending linearly on redshift, and similarly for the remaining four models. We do not include models with terms higher than quadratic in l_{UV} because such models were estimated to give a poorer fit to the data (cf. § 6.4.4). Model $\mathcal{M}_t$ is almost identical to the paramerization used by several other authors (e.g., Avni & Tananbaum, 1986; Wilkes et al., 1994, ,S06), with the exception that other authors have used the fractional cosmological look-back time, $\tau(z) = 1 - t(z)/t(0)$.

Because some of the values of l_X are only upper limits, we employ the Expectation-Maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977; Aitken, 1981) to calculate the maximum-likelihood solution for the regression parameters. These parameters are the regression coefficients, γ, and the intrinsic variance about the linear relationship, σ^2. The likelihood functions for these five models are normal densities with means given by Equations (6.3)–(6.7). The regression is carried out directly on $\alpha_{\mathrm{ox}} = 0.384(l_{UV} - l_X + 2.605)$ using computer routines coded by the authors.

The results of the regressions using the entire *Chandra* sample are

$$\bar{\alpha}_{\text{ox}} = (-5.148 \pm 1.293) + (0.147 \pm 0.029)l_{UV} - (0.014 \pm 0.018)z, \ \sigma = 0.156 \tag{6.8}$$

$$\bar{\alpha}_{\text{ox}} = (-7.048 \pm 1.443) + (0.190 \pm 0.033)l_{UV} - (0.293 \pm 0.137)\log(1+z),$$
$$\sigma = 0.155 \tag{6.9}$$

$$\bar{\alpha}_{\text{ox}} = (-8.816 \pm 1.473) + (0.223 \pm 0.031)l_{UV} + (3.061 \pm 0.890) \times 10^{-2}t(z),$$
$$\sigma = 0.151 \tag{6.10}$$

$$\bar{\alpha}_{\text{ox}} = (62.83 \pm 30.48) - (2.816 \pm 1.336)l_{UV} + (3.226 \pm 1.464) \times 10^{-2}l_{UV}^2,$$
$$\sigma = 0.155 \tag{6.11}$$

$$\bar{\alpha}_{\text{ox}} = (24.22 \pm 32.88) - (1.212 \pm 1.428)l_{UV} + (1.560 \pm 1.552) \times 10^{-2}l_{UV}^2 +$$
$$(2.684 \pm 0.965) \times 10^{-2}t(z), \ \sigma = 0.151. \tag{6.12}$$

For all of these models, α_{ox} increases (becomes more X-ray quiet) with increasing luminosity and decreases (becomes more X-ray loud) with increasing redshift. The intrinsic scatter about the relationships is estimated to be $\sigma \sim 0.15$.

6.4.2 Evidence for low-redshift BAL QSOs in the Sample

The estimates for the regression parameters are derived via maximum-likelihood. However, the likelihood functions assume that the residuals are normally distributed. If this assumption is not true, then it may bias our results. To test the assumption of normality in the residuals, in Figure 6.4 we compare the cumulative distribution function (CDF) of the standardized residuals for $\mathcal{M}_t$ with the standard normal. The standardized residuals are the residuals normalized by the intrinsic scatter. The CDFs for the other models were very similar. As can be seen, there is evidence for a violation of the assumption of normality. A Kolmogorov-Smirnov (KS) test confirmed this, finding a probability of ≈ 0.001–0.007 that the maximum difference between the CDFs of the standardized residuals for the

models and the standard normal is greater than that observed, assuming that the two distributions are the same. In addition, inspection of the residuals reveals that there are several sources that are significantly more X-ray quiet than would be expected from Equations (6.8)–(6.13), and are therefore outliers. These sources are all censored (i.e., not detected in the X-rays) and at $z < 1.5$. As noted in § 6.3, the high column density of BAL QSOs gives them the appearance of being X-ray weak. Furthermore, we are unable to remove BALs at $z < 1.5$. These two facts suggests that the regression outliers are BALs, as they are unusually X-ray weak and at $z < 1.5$.

To test the possibility that the $z < 1.5$ censored sources are dominated by BALs, and thus affecting our regression analysis, we removed these 10 sources and recalculated the regressions. Note that this number is consistent with the expected number of BALs in the $z < 1.5$ sample (cf., § 6.3). While this may remove some non-BALs from the $z < 1.5$ sample, it is unlikely that removing a few censored non-BALs from the fit will significantly affect the results, since these sources are not expected to be outliers and the regression is dominated by the detected sources. However, the BALs can have a non-negligible effect on the regression even if they are censored because they have an additional absorption component that contributes to the observed X-ray luminosity, and therefore are not expected to follow the functional form assumed by the regression and can be outliers. The existence of outliers has the effect of biasing the estimate of the intrinsic scatter upwards, inflating the uncertainties on the regression coefficients, and therefore reducing the statistical significance of the regression coefficients.

After removing the $z < 1.5$ censored data points, we were left with a sample of 164 sources, 157 (96%) of which are detected. Performing the regressions on
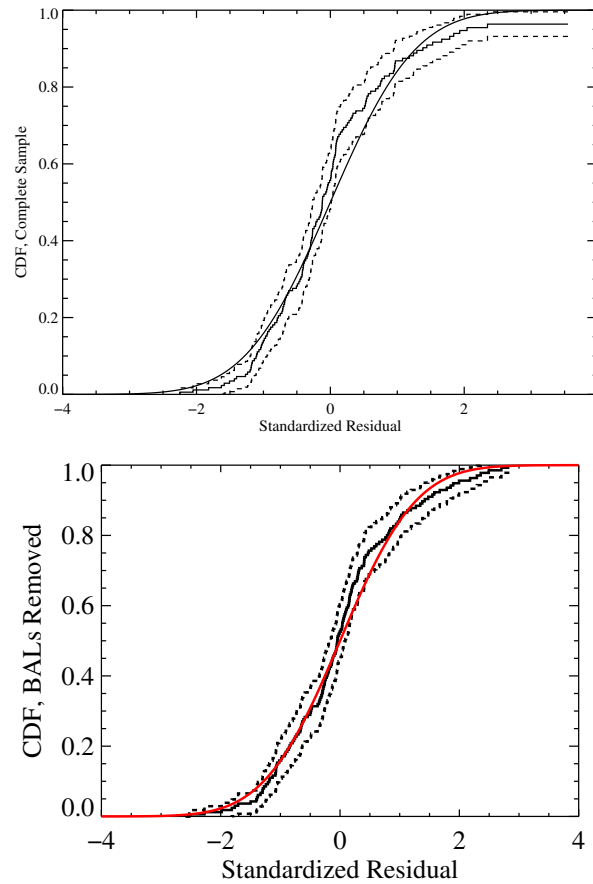
Figure 6.4 Empirical cumulative distribution functions (CDF) of the standardized residuals for the regression of α_{ox} on l_{UV} and $t(z)$ using the complete sample (top), and after removing suspected BALs (all $z < 1.5$ censored data points, bottom). The red line is the standard normal CDF, and the dashed lines denote the 95% pointwise confidence interval on the empirical CDF. The CDF of the standardized residuals for the full sample shows evidence of diverging from normality, while the CDF of the sample with suspected BALs removed is consistent with the assumption of normality.

this second sample, we find

$$\bar{\alpha}_{\mathrm{ox}} = (-5.340 \pm 1.097) + (0.150 \pm 0.025)l_{UV} - (0.004 \pm 0.015)z, \ \sigma = 0.131 \tag{6.13}$$

$$\bar{\alpha}_{\mathrm{ox}} = (-7.152 \pm 1.229) + (0.192 \pm 0.028)l_{UV} - (0.224 \pm 0.116)\log(1 + z),$$
$$\sigma = 0.129 \tag{6.14}$$

$$\bar{\alpha}_{\mathrm{ox}} = (-9.273 \pm 1.241) + (0.233 \pm 0.026)l_{UV} + (2.870 \pm 0.750) \times 10^{-2}t(z),$$
$$\sigma = 0.125 \tag{6.15}$$

$$\bar{\alpha}_{\mathrm{ox}} = (74.61 \pm 25.17) - (3.349 \pm 1.103)l_{UV} + (3.827 \pm 1.208) \times 10^{-2}l_{UV}^2,$$
$$\sigma = 0.127 \tag{6.16}$$

$$\bar{\alpha}_{\mathrm{ox}} = (40.97 \pm 27.34) - (1.949 \pm 1.187)l_{UV} + (2.370 \pm 1.288) \times 10^{-2}l_{UV}^2 \tag{6.17}$$
$$(2.263 \pm 0.813) \times 10^{-2}t(z),$$
$$\sigma = 0.124 \tag{6.18}$$

The results are very similar to Equations (6.8)–(6.19), but the intrinsic scatter has decreased and the significance levels of the regression coefficients are in general higher. Note that Equations (6.16) and (6.19) are equivalent to the form $L_X \propto e^{-t(z)/t_0}$, where the e-folding time is $t_0 = 5.75^{+4.98}_{-1.83}$ (95% confidence) Gyr for $\mathcal{M}_t$ and $t_0 = 7.25^{+14.4}_{-2.96}$ (95% confidence) Gyr for $\mathcal{M}_{l+t}$.

The CDF of the standardized residuals for Equation (6.16) is also shown in Figure 6.4. As can be seen, the residuals no longer show any evidence for a significant divergence from normality, suggesting that we have minimized BAL contamination by removing the $z < 1.5$ censored sources. A KS test also found that the empirical distribution of the standardized residuals for all parameterizations considered are not significantly different than the standard normal, having p-values of $p \sim 0.1$.

6.4.3 α_{ox} Depends on Both L_{UV} and z.

It is apparent from Equations (6.14)–(6.19) that there is statistically significant evidence for a dependence of α_{ox} on l_{UV} ($> 6\sigma$ significance). In addition, there is evidence from model $\mathcal{M}_t$ that α_{ox} depends on cosmic time (3.8σ significance), and evidence from model $\mathcal{M}_l$ that the α_{ox}–l_{UV} relationship is nonlinear (3.2σ significance). While the coefficients for both parameterizations imply that our data are inconsistent with the simple form $\bar{\alpha}_{\mathrm{ox}} = \gamma_0 + \gamma_l l_{UV}$, it is unclear which parameterization is the preferred one. In particular, because there is a strong correlation between l_{UV} and z, it is possible that the α_{ox}–$t(z)$ relationship is simply correcting for the nonlinearity in the α_{ox}–l_{UV} relationship, and is thus a spurious result. Because the models are not nested, (i.e. one is not merely a subset of the next), we cannot use classical statistical methods, such as the likelihood ratio or F-test, to compare their relative merits (e.g., Efron, 1984; Freeman et al., 1999; Protassov et al., 2002). Instead, we adopt an approach that attempts to find the model that minimizes the 'distance' to the true probability density that gives rise to the observed data. We do this by finding the model that minimizes the Kullback-Leibler information (KLI; see Appendix D).

We use the Akaike Information Criterion (AIC, Akaike, 1974) to estimate the difference in KLI between models. We estimate the difference in KLI between two models by multiplying their difference in AIC by $1/2$. Terms of order higher than l_{UV}^2 increased the AIC for models $\mathcal{M}_l$ and $\mathcal{M}_{l+t}$, and were not included in the analysis. Denoting the maximum likelihood estimate of the model parameters as $\hat{\theta}$, and the estimated KLI as $H(\hat{\theta})$, we find $H(\hat{\theta}_z) - H(\hat{\theta}_{l+t}) = 7.599$, $H(\hat{\theta}_\zeta) - H(\hat{\theta}_{l+t}) = 5.798$, $H(\hat{\theta}_l) - H(\hat{\theta}_{l+t}) = 2.770$, and $H(\hat{\theta}_t) - H(\hat{\theta}_{l+t}) = 0.677$. Here, $H(\theta_z)$ denotes the KLI for model $\mathcal{M}_z$, and likewise for $H(\theta_\zeta), H(\theta_l), H(\theta_t)$, and $H(\theta_{l+t})$. Model $\mathcal{M}_{l+t}$ is best supported by the empirical evidence.

Plots showing the residuals and partial residuals for the $\mathcal{M}_{l+t}$ regression are shown in Figure 6.5. To visualize the result, Figure 6.6 shows a 3-dimensional plot of the best fit for this model. The partial residual plots display the dependence of α_{ox} on L_{UV}, after accounting for the dependence on cosmic time, and the dependence of α_{ox} on z, after accounting for the dependence on L_{UV}. Both of these figures correspond to the regression results after removing the suspected BALs. Also shown in Figure 6.5 are non-parametric fits to the residuals, calculated using a locally-weighted average based on a Gaussian smoothing kernel; the kernel width was chosen using generalized cross-validation (e.g., see Hastie, Tibshirani, & Friedman, 2001). As can be seen, according to this model the nonlinearity in the α_{ox}–l_{UV} dependence, if real, is such that α_{ox} increases (becomes more X-ray quiet) faster at higher l_{UV}. This trend is in agreement with the results of Steffen et al. (2006), who found evidence that the slope of the α_{ox}–l_{UV} correlation may be steeper at higher l_{UV}.

Based on the AIC, model $\mathcal{M}_{l+t}$, which contains the $t(z)$ parameterization with a quadratic l_{UV} term, appears to provide the best description of our data, followed by the $t(z)$ parameterization with only a linear l_{UV} term.

6.4.4 Effects of Sampling and Nonlinear Dependence of α_{ox} on Luminosity

To assess how our estimate of the KLI varies under sampling from the underlying joint distribution of $(\alpha_{\mathrm{ox}}, l_{UV}, z)$, we use the non-parameteric bootstrap (Efron, 1979). We drew 10^4 bootstrap samples and performed the regression for each parameterization on each bootstrap sample. We then estimated the KLI in the same manner, with the exception that we now use the sample mean of the difference in log-likelihoods of the *original* sample, evaluated at the maximum likelihood estimate of θ based on the bootstrapped samples. The sampling distributions of the differences in KLI between $\mathcal{M}_z, \mathcal{M}_\zeta, \mathcal{M}_l$, and $\mathcal{M}_t$, with respect to $\mathcal{M}_{l+t}$, are
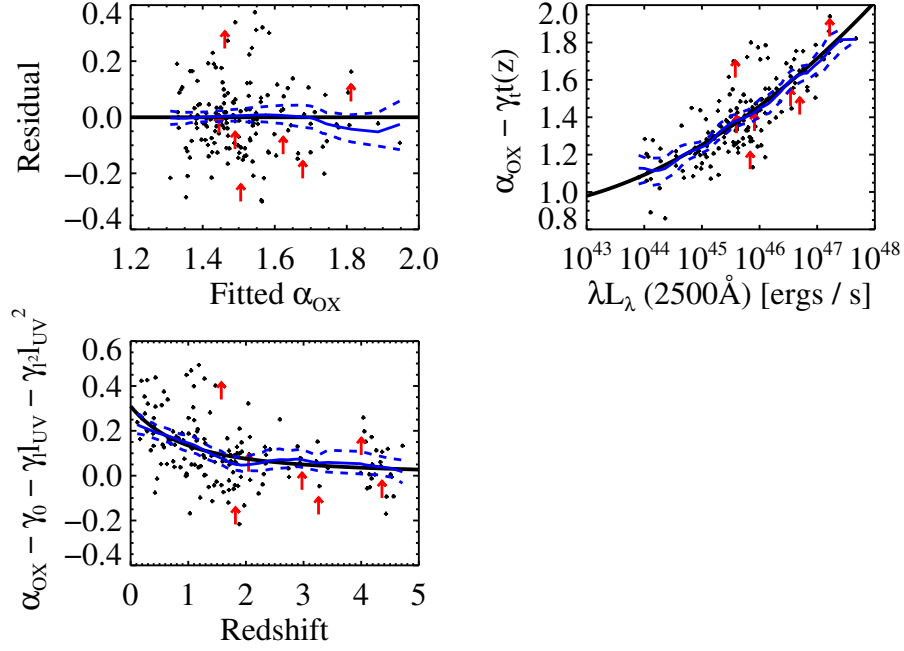
Figure 6.5 The residuals of the α_{ox} regression for model $\mathcal{M}_{l+t}$, shown as a function of the fitted α_{ox}, and partial residuals shown as functions of L_{UV} and z. The partial residual plots show the dependence of α_{ox} on L_{UV} or z, after accounting for the dependence of α_{ox} on z or L_{UV}. Also shown are kernel-smoother fits to the residuals (solid blue lines), as well as approximate 95% pointwise confidence intervals on the kernel-smoother fits (dashed blue lines).
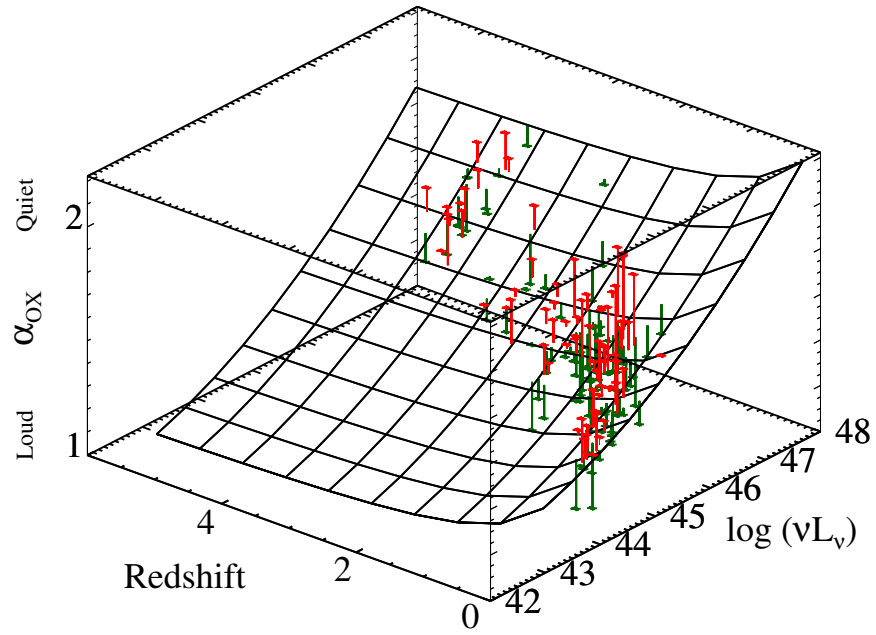
Figure 6.6 The 3-dimensional distribution of $\alpha_{\mathrm{ox}}, l_{UV}$, and z. The surface is the best fit to the data, obtained with model $\mathcal{M}_{l+t}$. Red denotes data points that fall above the fit, green denotes data points that fall below the fit.

shown in Figure 6.7. The $t(z)$ parameterization with quadratic l_{UV} term, $\mathcal{M}_{l+t}$, had the smallest estimated KLI for $\approx 99.9\%$ of the bootstrap samples. Therefore, the preference for $\mathcal{M}_{l+t}$ is unlikely to have resulted from fluctuations caused by random sampling, and thus it appears that this parameterization provides the best description of our data.

It is unlikely that the evidence for evolution is a result of nonlinearity in l_{UV}. Assuming that there is no dependence of α_{ox} on z, we can use a suitably large enough polynomial expansion of l_{UV} to approximate any smooth nonlinear dependence of α_{ox} on l_{UV}. However, as mentioned above, model $\mathcal{M}_l$ had the best AIC among the set of polynomial expansions in l_{UV}, and thus our data does not prefer terms of order higher than l_{UV}^2. Therefore, $\mathcal{M}_l$ should be viewed as the best approximation to the α_{ox}–l_{UV} relationship that is supported by our data without overfitting, and assuming that α_{ox} is independent of z. However, because the models that included $t(z)$ had an AIC lower than $\mathcal{M}_l$, and because $\mathcal{M}_l$ had an AIC lower than models that included terms of higher order, if follows that models $\mathcal{M}_{l+t}$ and $\mathcal{M}_t$ are preferred by our data over any polynomial expansion of α_{ox} as a function of l_{UV}. This is not to say that models $\mathcal{M}_{l+t}$ and $\mathcal{M}_t$ are preferred over any smooth nonlinear function of l_{UV}, but that if such a function exists, it is unlikely to differ significantly from $\mathcal{M}_l$. The nonlinear effects are not extreme, and in fact are not 'statistically significant' in the classical sense. However, while there is not enough evidence in the data to reject a null hypothesis that α_{ox} is linear in l_{UV} and $t(z)$ at, say, $> 2\sigma$ significance, the empirical evidence supports a nonlinear dependence of α_{ox} on l_{UV} at a given z over a linear dependence.

6.4.5   Effect of Variability and Measurement Error on the Results

Measurement error or variability may induce false correlations between parameters. In this section, we consider two effects. The first is especially important
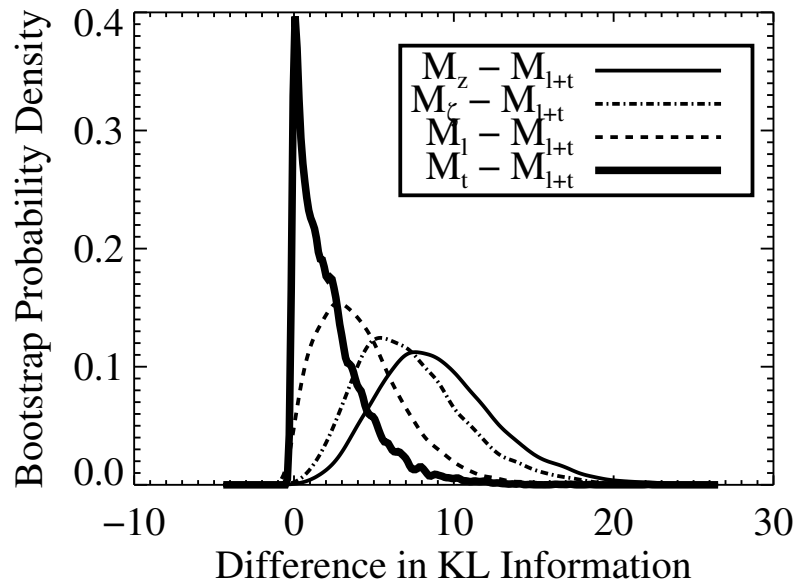
Figure 6.7 Sampling distributions of the difference in Kullback-Leibler informa-
tion relative to the $t(z)$ parameterization with quadratic l_{UV} term, $\mathcal{M}_{l+t}$, as deter-
mined from bootstrapping. Shown are the estimated distributions of the differ-
ence in KLI between $\mathcal{M}_z$ and $\mathcal{M}_{l+t}$ (thin solid line), $\mathcal{M}_\zeta$ and $\mathcal{M}_{l+t}$ (dot-dashed
line), $\mathcal{M}_l$ and $\mathcal{M}_{l+t}$ (dashed line), and $\mathcal{M}_t$ and $\mathcal{M}_{l+t}$ (thick solid line). As can
be seen, the $t(z)$ parameterization with quadratic l_{UV} term is almost always pre-
ferred.

when many sources are near the flux limit of the sample, and when the number of sources increases strongly with decreasing flux. In the second case, a bias can result even if all the sources are far above the flux limit.

For the first case, we argue that the expected tendancy would be for α_{ox} to *increase* with increasing redshift, the opposite of what we claim from the data. Consider the possiblity that the sources are variable. Assume for simplicity of argument that the sources vary around some mean flux in all spectral bands, and that a particular source spends an equal amount of time brighter than the mean and dimmer than the mean. Then the sources that were discovered by SDSS near the flux limit are preferentially observed in their 'bright' state. By the time we observed them with Chandra, they will likely no longer be in their 'bright' state, and thus may be systematically X-ray quieter. Thus we expect the sources near the flux limit to appear fainter in X-rays on average than they really are. Since most of the sources near the flux limit are at high redshift, the tendency may be for α_{ox} to increase (quasars are less X-ray bright) with redshift. Since this is the opposite of what we see, variability of sources near the flux limit is not producing the result.

For measurement error, the qualitative argument is similar. Near the flux limit, random errors in photon counts result in more sources just below the limit being randomly included in the sample than sources above the limit being randomly excluded, provided that the number of sources is an increasing function of decreasing flux limit (which is the case here). Thus, the SDSS selection would again be biased towards sources with optical fluxes that appear brighter in optical than they really are. Chandra then measured X-ray fluxes for essentially all the sources. Thus, we expect the sources near the flux limit to be systematically more X-ray faint than they really are. Again, this is the opposite of what we see

in the data, so this Malmquist-type bias is not important.

For the second case, measurement errors or variability may induce false correlations even if all sources are detected far above the flux limit. However in this case, Monte Carlo simulations can be carried out to see how important the effect might be. Measurement errors on the independent variables can bias the estimates of the regression coefficients (e.g., Akritas & Bershady, 1996; Fox, 1997), and errors on the dependent variable can bias the coefficient estimates for censored regression. When the dependent variable is measured with error and/or variability, the measurement error and variability inflate the observed variance in the regression residuals, biasing the estimate of the intrinsic scatter, σ^2. For ordinary least-squares this is not a problem, since the estimate for the intrinsic scatter, σ^2, and the estimates of the regression coefficients, γ, are statistically independent. However, for the censored regression model, the estimates of the intrinsic scatter and the coefficients are no longer statistically independent, and the bias in the intrinsic scatter estimate also carries over to the coefficient estimates (Stapleton & Young, 1984). These facts are confirmed by Monte Carlo simulations, which have shown that the *observed* relationship between l_X and l_{UV} can differ from the intrinsic relationship when the observed l_X and l_{UV} differ significantly from the intrinsic l_X and l_{UV} (e.g., Yuan et al., 1998b). Since the UV and X-ray data are measured with error, we are not fitting the intrinsic distribution of l_X given l_{UV} and z, but rather the distribution of $l_X + \epsilon_X$ at a given $l_{UV} + \epsilon_{UV}$ and z, where ϵ_X and ϵ_{UV} are random error terms. The errors for l_{UV} are the usual measurement errors from the continuum fitting and are very small in our analysis, with typical values of $\sigma_{UV} \approx 0.001$–$0.01$ dex. However, the errors for l_X include the contribution from measurement errors and from variability. The errors from variability of the X-ray emission arise from the fact that the X-ray and optical observations are

not simultaneous. We are interested in the distribution of l_X at a given l_{UV} and z; however, because the X-ray observations are not simultaneous with the optical, we do not observe the value of l_X given l_{UV} for each source, but some value of l_X which has varied from the original X-ray luminosity at the time of the optical observations.

Typical long-term X-ray variability for Seyfert 1s is 20%–40% with no obvious trend with luminosity (Grupe et al., 2001; Uttley et al., 2002; Markowitz et al., 2003). The measurement errors in l_X for our sample are typically ~ 0.07 dex. Assuming X-ray variability amplitudes of 30% for the sources in our sample, this implies typical uncertainties in the X-ray luminosity of ~ 0.15 dex. Correcting the scatter in α_{ox} for the contribution from X-ray variability and measurement error, we find an implied intrinsic scatter in l_X of $\sigma \approx 0.29$ dex.

To assess whether the observed dependence of α_{ox} on z is the result of bias arising from variability, we performed Monte Carlo simulations. Because we are interested in testing if a spurious redshift dependence may occur due to this type of bias, we simulate values of l_X, given l_{UV}, assuming $L_X \propto L_{UV}^{0.65}$. Within the framework of model $\mathcal{M}_t$, this form corresponds to assuming $\gamma_l = 0.134$ and $\gamma_t = 0$, and therefore α_{ox} depends only on UV luminosity for these simulations. The value of $\beta_l = 0.65$ was chosen because a linear regression of l_X on l_{UV} found $L_X \propto L_{UV}^{0.631 \pm 0.088}$, consistent with the work of Avni & Tananbaum (1986), Wilkes et al. (1994), Vignali et al. (2003b), Strateva et al. (2005), and Steffen et al. (2006).

The simulations were performed as follows. We first drew 164 values of l_{UV} and $t(z)$ from a kernel estimate of their joint distribution (Silverman, 1986), after removing the censored $z < 1.5$ sources. Then, we calculated values of l_X, assuming $L_X \propto L_{UV}^{0.65}$. The random Gaussian scatter in l_X about the l_{UV} dependence had a standard deviation of $\sigma = 0.30$ dex; this value was motivated by the re-

gression results. To simulate the upper limits, we randomly censored 6 of the values of l_X, and increased their censored values by a small random amount. We added random Gaussian noise to the uncensored values of l_X to simulate the effects of variability and measurement error, where the standard deviation of this noise was 0.15 dex. To simulate the effect of the measurement errors on l_{UV}, we also added random Gaussian noise of standard deviation 0.005 dex to the values of l_{UV}. We then performed censored regression on the simulated values. We repeated this procedure for 10^4 simulations, and calculated the average simulated regression coefficients for model $\mathcal{M}_t$, $\bar{\gamma} = (\bar{\gamma}_l, \bar{\gamma}_t)$, and their covariance matrix, Σ_γ.

We calculate the χ^2 of our regression coefficients for $\mathcal{M}_t$ estimated from our sample, $\hat{\gamma} = (0.233, 0.029)$, as $\chi_2^2 = (\hat{\gamma} - \bar{\gamma})^T \Sigma_\gamma^{-1} (\hat{\gamma} - \bar{\gamma})$. Here, x^T is the transpose of x. We found a value of $\chi_2^2 = 11.62$; under the null hypothesis that $\gamma_l = 0.134$ and $\gamma_t = 0$, the probability of observing a χ_2^2 this high or higher is $\lesssim 3 \times 10^{-3}$. Similar results were found by calculating the χ^2 of the regression coefficients for $\mathcal{M}_{l+t}$. Therefore, our observed values of γ_t are highly unlikely to be a spurious correlation resulting from variability and measurement error.

In summary, we argue that Malmquist-type biases from measurement error or variability will induce a false correlation of α_{ox} with z in the *opposite* sense of what is observed, and therefore are not causing our finding. We further showed through simulations that measurement errors or variability for objects within the sample are likewise not capable of inducing a false correlation between variables.

6.4.6 Rank Correlation Analysis

An alternative test for evolution of the X-ray emission for a given UV luminosity is Kendall's generalized partial τ (Akritas & Siebert, 1996). Kendall's partial τ has been used by Vignali et al. (2003b), Strateva et al. (2005), and Steffen et al. (2006), where they did not find any evidence for a partial correlation between $\alpha_{\rm ox}$

and z based on it, consistent with their parametric analysis. However, Kendall's partial τ has some undesirable properties that make it difficult to assess the statistical significance of the result (Nelson & Yang, 1988). In particular, conditional independence between variables 1 and 2, given a third variable, does *not* necessarily correspond to a value of $\tau = 0$. Alternatively, conditional dependence of two variables given a third does not necessarily correspond to a value of $\tau \neq 0$. Within the context of this work, this implies that a value of $\tau = 0$ does not necessarily correspond to the null hypothesis that α_{ox} (or L_X) is independent of z at a given L_{UV}; i.e., an expected value of $\tau = 0$ does not necessarily result when $p(L_X|L_{UV}, z) = p(L_X|L_{UV})$.

In order to test the reliability of Kendall's partial τ, we use Monte Carlo simulations to compare the distributions of τ under the assumption that α_{ox} only depends on L_{UV}, and under the assumption that α_{ox} depends on both L_{UV} and z. For each of the simulations, we calculated values of Kendall's partial τ for α_{ox} with z, $\tau_{\alpha z, l}$, and L_X with z, $\tau_{xz, l}$, controlling for the correlation between L_{UV} and z. We did this for two hypotheses. The 'null' hypothesis, H_0, assumed $L_X \propto L_{UV}^{0.65}$, and the alternative (i.e., 'evolution') hypothesis, H_1, assumed $L_X \propto L_{UV}^{0.40} e^{-t(z)/5.5}$. The simulations under both hypothesis were performed in the same manner as described in § 6.4.5. The results are shown in Figure 6.8.

Under the null hypothesis of no evolution, the expected values of τ for both L_X and α_{ox} are indeed non-zero. Using the sample average of the simulations as an estimate of the expectation values, we find that the expected value of $\tau_{\alpha z, l}$ under the assumption that α_{ox} does not depend on z is 0.105. However, under the assumption that α_{ox} does depend on z, the average simulated value of τ is -0.001. Surprisingly, the expected value of τ for the α_{ox}–z partial correlation is approximately equal to zero when α_{ox} depends on z, at least for the simulation
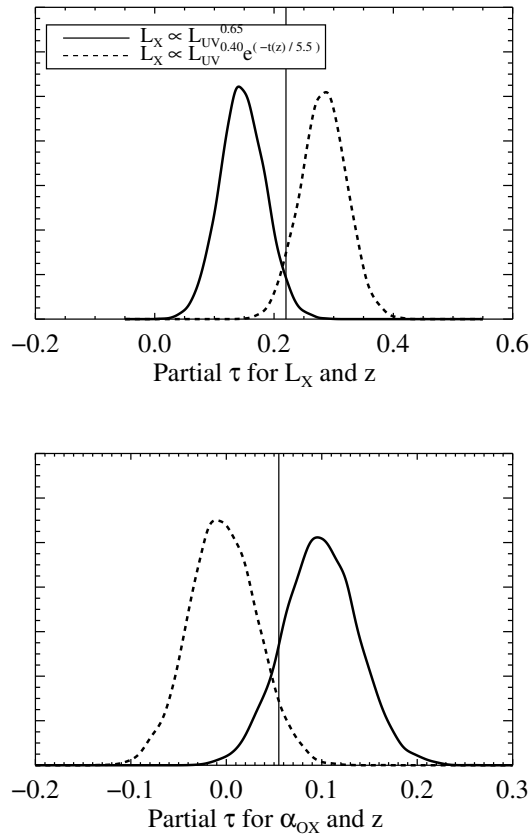
Figure 6.8 Distribution of Kendall's generalized partial τ for L_X and z (top) and α_{ox} and z (bottom) under the no-evolution hypothesis (solid line) and the evolution hypothesis (dashed line). The vertical lines show the observed values of τ for our sample. As can be seen, our value of τ is about as consistent with the evolution model as with the no-evolution model. Also, note that $\tau \neq 0$ under the null hypothesis of no evolution (i.e., statistical independence of α_{ox} and z given L_{UV}), and therefore it is incorrect to calculate significance levels with respect to $\tau = 0$.

performed here. We investigate the behavior of Kendall's partial τ further in Appendix E.

For our quasar sample, we find a value of Kendall's partial τ for L_X and z of $\tau_{xz,l} = 0.212$, and for α_{ox} and z of $\tau_{\alpha z,l} = 0.057$. Our sample has a value of Kendall's regular τ between L_{UV} and z of $\tau_{lz} = 0.686$. As can be seen from the distributions of the simulated τ, both of our observed values of τ are about as equally consistent with evolution of the X-ray emission at a given L_{UV} as with no evolution. In fact, there is considerable overlap between the distributions of τ under both hypotheses, thus making it difficult to distinguish between the two. Unfortunately, we are not able to decide in favor of either hypothesis using Kendall's generalized partial τ.

Based on the simulations, the lack of evidence for a significant correlation between α_{ox} and z based on Kendall's generalized partial τ (Vignali et al., 2003b; Strateva et al., 2005; Steffen et al., 2006) may be the result of an incorrect assumption about the distribution of τ under the null hypothesis. However, it should be noted that the parametric tests performed by Vignali et al. (2003b), Strateva et al. (2005), and Steffen et al. (2006) also did not reveal any evidence for evolution of α_{ox}. In addition, although we have shown that one can both incorrectly reject and accept the null hypothesis based on the partial τ statistic, there has never been a claimed rejection of the null hypothesis of no evolution in α_{ox} in previous studies.

6.5 RESULTS FOR Γ_X

To investigate any dependence of the X-ray photon index, Γ_X, on UV luminosity and redshift, we performed a weighted linear regression of Γ_X on l_{UV} and $\log(1+z)$ using all 157 detected sources. The weights are made up of a combination of

the intrinsic scatter in Γ_X and the measurement errors on Γ_X. The results are

$$\Gamma_X = -3.43(\pm 3.879) + 0.125(\pm 0.087)l_{UV} - 0.678(\pm 0.347)\log(1+z). \qquad (6.19)$$

Based on this regression, there is no significant evidence for a dependence of Γ_X on L_{UV} or z, although the z dependence is marginally significant at $\approx 2\sigma$.

Similar to α_{ox}, we experimented with parameterizing the z dependence using $t(z)$ and z. There was no noticeable difference between the different parameterizations, although the $\log(1+z)$ model gave slightly better results in the sense of minimizing mean squared error. In addition, if any z dependence of Γ_X is due to a systematic hardening of the X-ray spectra at higher energies, then we might expect the z dependence to be best parameterized using $\log(1+z)$.

We show the joint confidence regions of the l_{UV} and $\log(1+z)$ coefficients in Figure 6.9. While there is no significant evidence that Γ_X is related to either L_{UV} or z, we note that the measurement errors on Γ_X are large and contribute signficantly to enlarging the confidence region of the regression coefficients. Because the confidence region of the regression coefficients is large, the possibility that Γ_X is significantly correlated with both UV luminosity and redshift is also consistent with our data.

Previous work by Gallagher et al. (2005) has found evidence for an anti-correlation between Γ_X and α_{UV}. Motivated by their work, we also perform a linear regression of Γ_X on α_{UV}. We only included those $z > 0.5$ sources detected by Chandra, leaving us with 136 sources. The redshift limit was imposed to ensure that an adequate amount of the UV continuum was available for estimating α_{UV}. We used the *FITEXY* procedure (Press et al., 1992) with the Tremaine et al. (2002) modification to account for the measurement errors in both Γ_X and α_{UV}, and the intrinsic scatter about the regression. The result of the regression is:

$$\Gamma_X = 2.21(\pm 0.07) - 0.25(\pm 0.07)\alpha_{uv}. \qquad (6.20)$$
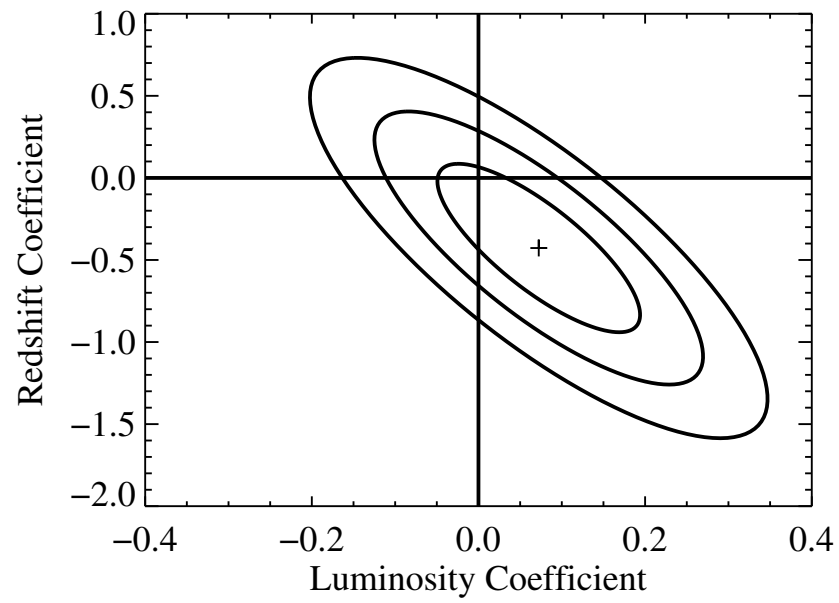
Figure 6.9 Confidence regions for the l_{UV} and $\log(1+z)$ coefficients in the regression of Γ_X on l_{UV} and $\log(1+z)$. The cross denotes the best-fit value, and the contours are the $68\%, 95\%$, and 99.7% joint confidence regions.

The α_{UV} coefficient is significant at 3.5σ. The regression results are consistent with the assumption that Γ_X is linearly related to α_{UV}, and the residuals are approximately normally distributed. We also performed a Spearman and Kendall rank correlation test between the two spectral slopes, and found an anti-correlation of similar significance.

To test if including the UV red quasars affect our limits on the luminosity and redshift dependencies, we also performed the regression after removing all sources with $\alpha_{UV} > 1.2$. Removing these UV red quasars did not significantly change the confidence regions shown in Figure 6.9.

6.6 COMPARISON WITH PREVIOUS STUDIES OF α_{ox}.

The parametric dependence of α_{ox} on L_{UV} and z has been studied previously by several authors (e.g., Avni & Tananbaum, 1982, 1986; Wilkes et al., 1994; Bechtold et al., 2003; Vignali et al., 2003b; Strateva et al., 2005; Steffen et al., 2006). In this analysis, we confirm the anti-correlation between α_{ox} and L_{UV} seen previously, but also find evidence for a correlation between α_{ox} and redshift. Most previous studies have not found any significant evidence that α_{ox} is related to z, with the exception of Bechtold et al. (2003). Yuan et al. (1998a) found evidence for a slight dependence of α_{ox} with z for $z < 0.5$, but with opposite sign as that found here. Using high-quality *Chandra* data, we find that α_{ox} is related to both L_{UV} and z.

We perform a quantitative comparison between our results and those of Avni & Tananbaum (1986), Wilkes et al. (1994), Strateva et al. (2005), and Steffen et al. (2006). These authors have presented their results using a different parameterization for evolution, where they have fit a linear relationship of the form[4]

$$\alpha_{\mathrm{ox}} = A_l(l_{UV} - 30.5 - \log \nu_{UV}) + A_\tau(\tau(z) - 0.5) + A. \qquad (6.21)$$

[4]The term $\log \nu_{UV}$ arises because we define l_{UV} to be the logarithm of $\nu_{UV} L_{UV}$

Here, $\tau(z)$ is the cosmological look-back time in units of the present age of the universe, and ν_{UV} is the frequency corresponding to $2500\mathring{A}$. We fit a relationship of this form and find $A_l = 0.233 \pm 0.026$ and $A_\tau = -0.392 \pm 0.103$, with a correlation of $Corr(A_l, A_\tau) = -0.878$. In Figure 6.10 we show the 95% joint confidence region on our estimate of (A_l, A_τ), as well as the 95% confidence ellipses for AT86, W94, S05, and S06. These authors do not report equations for their confidence regions, so we matched them by eye. We compare with the results of AT86 obtained using their entire sample, i.e., their BQS+BF+HET85 sample, and the results of W94 obtained using only the radio-quiet sources and assuming $\Gamma_X = 2$ as displayed in their Figure 14a. The results for the S05 sample are presented by S06.

Statistically, our results differ from the analysis of AT86 at the $\approx 4\sigma$ level, from W94 at the $\approx 2\sigma$ level, from S05 at the $\approx 2.5\sigma$ level, and from S06 at the $\approx 3\sigma$ level. However, there are a number of systematic differences between our analysis and those of AT86, W94, S05, and S06, that, when taken into account, may introduce an additional systematic component to the errors. The AT86 sample includes both radio-quiet and radio-loud sources, and as noted in § 6.3, the radio-loud sources can have an additional component in their X-ray emission from the jet. To avoid this type of contamination, we have only included RQQs in our sample. In addition, AT86 assumed a value of $\Gamma_X = 1.5$ when calculating the 2 keV flux. As has been found here and in many other studies, a value of $\Gamma_X \approx 2$ is more typical for RQQs. Wilkes et al. (1994) calculated A_τ for both $\Gamma_X = 1.5$ and $\Gamma_X = 2$ and found that assuming $\Gamma_X = 2$ had the effect of shifting A_τ towards more negative values. If the assumption on Γ_X affects the estimated A_τ for AT86 in the same way as for W94, then one would expect assuming $\Gamma_X = 2$ would shift the AT86 confidence ellipse toward our estimate.

The largest systematic difference between our work and that of AT86, W94,
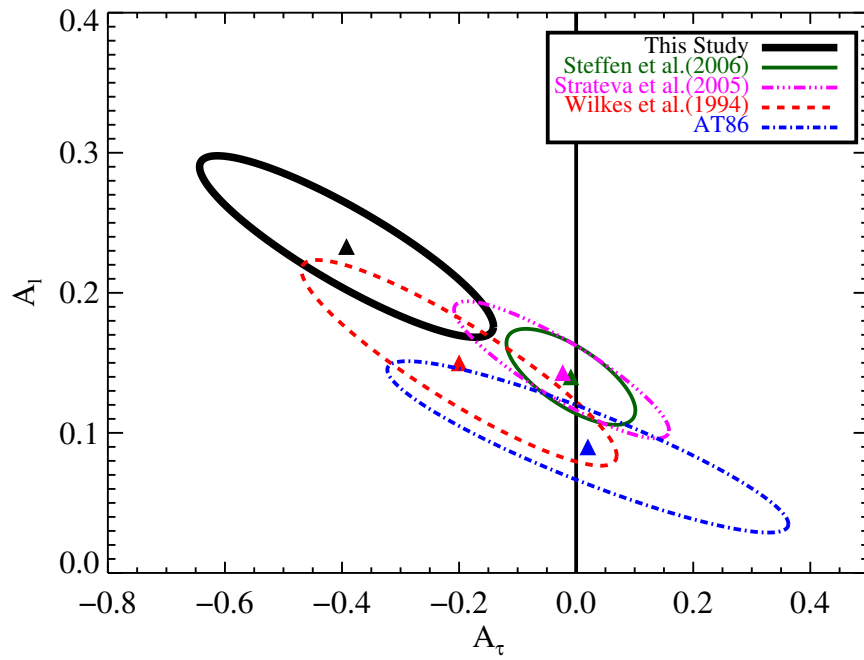
Figure 6.10 Comparison of our values of A_l and A_τ with AT86 (blue, dashed-dotted contour), W94 (red, dashed contour), S05 (magenta, dash-dot-dot-dot contour), and S06 (green, solid thin contour). The estimates of (A_l, A_τ) are denoted by triangles and the ellipses are approximate 95% (2σ) confidence regions.

S05, and S06 is in the differing levels of heterogeneity of the samples and the different instruments used to collect the X-ray data. The AT86 and W94 X-ray observations were done using the *Einstein* Observatory Imaging Proportional Counter (IPC). About two-thirds of the X-ray data for the S06 sample was observed using the *ROSAT* Position Sensitive Proportional Counter (PSPC), with the remaining X-ray data from *Chandra* or *XMM-Newton*. S06 combined sources from the SDSS, COMBO-17 survey (Wolf et al., 2004), Bright Quasar Survey (Schmidt & Green, 1983), a heterogeneous low-z Seyfert 1 sample, and a heterogenous high-z sample similar to our high-z sample. The S06 sample is more heterogeneous than ours, but probes a wider range in luminosity; the ranges in cosmic age probed by S06 and our sample are very similar. The S05 sample is a subset of the S06 sample, and does not contain the COMBO-17 and BQS sources. The main SDSS sample of S05 and S06 consists of 155 radio-quiet quasars that were selected from the SDSS and contained within the inner $19'$ of *ROSAT* PSPC pointings with exposure times > 11 ksec. Thus, the S05 sample is very similar to ours in its heterogeneity, with the exception of the additional low redshift Seyfert 1 sample; S05 perform their analysis both with and without the Seyfert 1 sample. The AT86 and W94 samples are both more heterogeneous than the S06 sample. In addition, the redshift ranges probed by the AT86 and W94 samples are smaller ($z < 3.3$) than that of the S06 sample, the S05 sample, and our sample, probing a slightly smaller range in cosmic time.

Our sample only consists of *Chandra* ACIS observations. Increased sensitivity gives *Chandra* the ability to detect sources with rest-frame 2–10 keV flux down to $f_{HB} \sim 2 \times 10^{-15}$ ergs s^{-1} in a ~ 20 ksec observation. In this work, we estimate the *Chandra* 2 keV flux densities using data over a broader spectral range (0.3–7 keV) than that probed by the *Einstein* IPC (0.4–4 keV) and the *ROSAT* PSPC (0.1–2.5

keV). Our sample is also more homogeneous than those used previously, except for the S05 SDSS + high-z sample, being made up almost entirely of SDSS sources which had serendipitous *Chandra* observations; unfortunately this also results in our sample probing a smaller range in luminosity. Similar to our work, the high-z samples of S06 and S05 both consist of sources with *Chandra* and *XMM* data, and the COMBO-17 sample of S06 also consists of *Chandra* data; both the S05 and S06 samples have slightly lower X-ray detection fractions than our sample.

Avni & Tananbaum (1986) and W94 estimate the $2500\mathring{A}$ flux density from published B- and V-band magnitudes assuming a constant spectral slope of $\alpha = 0.5$. The $2500\mathring{A}$ flux density for many of the S05 and S06 sources were measured directly from the SDSS spectra. However, the $2500\mathring{A}$ flux densities for the 52 COMBO-17 sources of S06 were estimated by interpolation and extrapolation. The $2500\mathring{A}$ flux densities for the 46 BQS sources in S06 were estimated from the $3000\mathring{A}$ flux assuming a constant spectral slope of $\alpha = 0.5$. The dispersion in quasar spectral slopes is large (≈ 0.3, Richards et al., 2001), and this large dispersion can result in a non-negligible error on l_{UV} if one assumes a constant spectra slope, especially if one is extrapolating over a large range in wavelength. These issues are exacerbated when one fits α_{ox} instead of l_X, as the errors on l_{UV} contribute to the errors on α_{ox}, thus not only increasing the scatter about the regression, but also correlating the errors on l_{UV} and α_{ox}. In addition, as noted in § 6.4.5, the X-ray variability can also bias the coefficients for censored regression. In particular, these issues will affect the AT84 and W94 results because of the larger wavelength difference between $2500\mathring{A}$ and the B- and V-bands for many of the sources, and the lower detection fraction ($\sim 60\%$). The analysis of S05 and S06 is unlikely to be significantly affected by these issues due to the high detection fraction and large number of sources with directly measured $2500\mathring{A}$ flux densi-

ties. Furthermore, these authors found consistent results when analyzing different subsamples of their data.

Strateva et al. (2005) report values of Kendall's generalized partial τ for a partial correlation between α_{ox} and z, given L_{UV}, for their main SDSS sample combined with their high-z sample. Because the main + high-z sample of S05 is very similar to ours in distribution of L_{UV} and z and the number of sources, we expect that the distribution of Kendall's partial τ under the no-evolution and evolution hypotheses should also be similar. Strateva et al. (2005) find a value of $\tau = 0.03$, where we have corrected for the sign difference between our definition of α_{ox} and theirs. This value of τ is consistent with our value of $\tau = 0.057$. Because their value of τ is not significantly different than $\tau = 0$, and because their parametric analysis gave similar results, S05 concluded that there is no evidence that α_{ox} changes with redshift. However, as per the discussion in § 6.4.6, the expected value of τ under the null hypothesis of no evolution in α_{ox} is in general not $\tau = 0$, and therefore it is inappropriate to calculate signficance levels with respect to $\tau = 0$. Comparison with Figure 6.8 implies that the value of $\tau = 0.03$ found by S05 is about as equally consistent with the evolution hypothesis as with the no-evolution hypothesis. However, the parametric analysis by S05 still differs from ours at the $\approx 2.5\sigma$ level.

Similar to Bechtold et al. (2003), we find that α_{ox} is correlated with both UV luminosity and redshift. However, in contrast to Bechtold et al. (2003), we find that α_{ox} depends more strongly on L_{UV} than on z. In addition, we find that RQQs are systematically more X-ray loud at higher redshift. Bechtold et al. (2003) found that α_{ox} is larger (more X-ray quiet) for high-z sources, but found an overall trend where α_{ox} becomes more X-ray loud as z increases. Bechtold et al. (2003) did not perform a regression or partial correlation analysis, but we note that their ob-

served marginal distribution of α_{ox}–z is such that α_{ox} becomes more X-ray loud as z increases for $z \lesssim 2$. This is opposite the trend seen in our data, where inspection of Figure 6.3 reveals that α_{ox} is observed to become more X-ray quiet with increasing z, if one does not correct for the L_{UV}–z correlation.

The source of this discrepancy is likely the values of Γ_X used by Bechtold et al. (2003) for their *ROSAT* sources. Bechtold et al. (2003) used Γ_X values taken from Yuan et al. (1998a), which were calculated using the two hardness ratios given by the Standard Analysis Software System (SASS), to estimate the flux density at 2 keV. These values of Γ_X steadily decrease from $\Gamma_X \sim 2.6$ at $z \sim 0$, to $\Gamma_X \sim 2$ at $z \sim 2$. However, values of $\Gamma_X \sim 2.6$ are steeper than is commonly seen in RQQs, as has been found in this work and in other recent studies (e.g., Reeves & Turner, 2000; Piconcelli et al., 2003). Therefore, assuming a power-law and the values of Γ_X obtained by Yuan et al. (1998a) may not provide an accurate estimate of the 2 keV flux density, and thus α_{ox}. If $\Gamma_X \sim 2$, then assuming values of $\Gamma_X \sim 2.6$ will systematically under-predict the 2 keV flux density for a given 0.1–2.4 keV flux, and consequently provide estimates of α_{ox} that are too large. Furthermore, a steady decrease from $\Gamma_X \sim 2.6$ at $z \sim 0$ to $\Gamma_X \sim 2$ at $z \sim 2$ would produce a similar observed decrease in α_{ox} from $z \sim 0$ to $z \sim 2$, thus increasing the magnitude of any α_{ox}–z anti-correlation. Considering that L_{UV} also increases with increasing z due to flux limits, this would also weaken any observed correlation between α_{ox} and l_{UV}, and thus lead Bechtold et al. (2003) to conclude that α_{ox} is a stronger function of redshift. This is what is observed in the Bechtold et al. (2003) data, where α_{ox} is observed to decrease from $z \sim 0$ to $z \sim 2$. At $z \sim 2$, the Yuan et al. (1998a) sources have values of Γ_X that are more typical of RQQs, $\Gamma_X \sim 2$. In addition, after $z \sim 2$, the trend in α_{ox} is observed to change sign, increasing with increasing z, consistent with the data presented here. This

is also the redshift where the Bechtold et al. (2003) sample becomes dominated by *Chandra* sources, and have 2 keV flux densities calculated assuming $\Gamma_X = 2.2$.

6.7   DISCUSSION

In this paper we were able to separate the dependence of α_{ox} on the quasar luminosity, L_{UV}, from that of cosmic epoch, z, and we find that both dependencies are present, though with opposite sign. From this, it follows that RQQs become more X-ray quiet (increasing α_{ox}) with increasing UV luminosity, and become more X-ray loud with increasing redshift. An analysis based on the Kullback-Leibler information finds evidence that α_{ox} may depend nonlinearly on l_{UV} at a given $t(z)$, with α_{ox} increasing more rapidly as l_{UV} increases.

One may be able to find a better parameterization for the redshift dependence than the one adopted here, but that would only strengthen our claims of evidence for a dependence of α_{ox} on z. In addition, as argued in § 6.4.4, it is unlikely that the observed redshift dependence can be explained by nonlinearity in the α_{ox}–l_{UV} relationship. However, expanding the model space to include other parameters such as black hole mass or accretion rate may provide a better fit and be preferred over models which contain only a redshift and L_{UV} dependence. For example, a dependence of α_{ox} on black hole mass, M_{BH}, is predicted by some models for accretion disk and hot corona (e.g., Janiuk & Czerny, 2000; Bechtold et al., 2003). If such a correlation exists, than it may be that the the α_{ox}–$t(z)$ relationship is simply tracing the underlying evolution of the active black hole mass function, which is then projected onto the α_{ox}-$t(z)$ plane via an α_{ox}–M_{BH} relationship. In this case the statistical model that contains M_{BH} would provide the best fit, and there would be no need for an additional redshift dependence. However, in the absence of such information, we find that the model that best describes our data

is given in terms of a quadratic dependence on l_{UV} and a linear dependence on $t(z)$.

6.7.1 K-Corrections and the α_{ox}–z Relationship

It may be suggested that the α_{ox} dependence on z is caused by a varying Γ_X as the observed *Chandra* spectral range shifts to harder rest-frame energies. If Γ_X were to steepen at higher energies, creating a softer X-ray continuum at these energies, then we would be systematically over-estimating the 2 keV flux densities, thus explaining the α_{ox}–z relationship. While we do not find any strong evidence for a change in Γ_X with z, the results from § 6.5 shown in Figure 6.9 suggest that if there is spectral curvature, then Γ_X likely flattens with increasing energy. A flattening of the X-ray continua at higher energies is opposite the trend needed to explain the α_{ox}–z relationship, and thus our result cannot be explained by a systematic steepening of the intrinsic X-ray continuum at harder energies.

Because we do not fit an intrinsic absorber to most of our sources, it may also be suggested that the observed redshift dependence of α_{ox} is caused by red-shifting of soft X-ray absorption out of the observed spectral region. An intrinsic absorber will more strongly absorb the softer X-rays, and therefore will more significantly affect the observed X-ray continuum of lower redshift sources. This could then cause a spurious anti-correlation between L_X and $t(z)$.

To test if the observed dependence of α_{ox} on $t(z)$ is the result of soft X-ray absorption shifting out of the observed band, we used *SHERPA*'s FAKEIT routine to simulate observed X-ray spectra as a function of z. We assumed a power-law continuum with $\Gamma_X = 2$, and an intrinsic neutral absorber with column density $N_H = 10^{21}$ cm^{-2}. We argue in the next paragraph that a column density of $N_H = 10^{21}$ cm^{-2} is greater than the maximum N_H allowed by the Γ_X–z regression, and thus we use $N_H = 10^{21}$ as an upper limit on the effect of unrecognized neutral

intrinsic absorption on the α_{ox}–z relationship. The observed X-ray continuum was simulated for a source at $z = 0, 1, 2, 3$, and 4, and the intrinsic luminosity of the source was kept constant. We then fit each simulated spectrum with only a power-law. This resulted in the inferred X-ray luminosity of $z = 0$ sources being a factor of ~ 2 lower than the $z = 4$ sources. Therefore, based on these simulations, ignoring intrinsic absorption can result in a spurious decline in L_X from $z = 4$ to $z = 0$ by a factor of ~ 2 when $N_H \sim 10^{21}$ cm^{-2}. However, the results of our α_{ox} regression imply that the X-ray luminosity drops by a factor of ~ 8 from $z = 4$ to $z = 0$, and therefore the observed α_{ox}–z relationship cannot be explained as a spurious correlation resulting from unidentified intrinsic neutral absorption.

We can use our simulated spectra to constrain a typical value of N_H for our sources, assuming that N_H remains roughly constant with redshift. The observed photon index of the simulated spectra dropped from $\Gamma_X = 2$ at $z = 4$ to $\Gamma_X \approx 1.4$ at $z = 0$. From Figure 6.9, we note that the 3σ limit on the maximal drop in observed Γ_X between $z = 4$ and $z = 0$ is $\Delta\Gamma_X \approx 0.35$. This is considerable less than the observed drop in Γ_X from the simulations, and thus represents more than the maximal amount of change in Γ_X with redshift that is allowed by our data. Therefore, assuming only neutral absorption, values of $N_H \gtrsim 10^{21}$ cm^{-2} would produce observed values of Γ_X at $z = 0$ that are too flat, and thus $N_H \lesssim 10^{21}$ cm^{-2} for most of our sources.

Because the redshift dependence of α_{ox} cannot be explained by a systematic steepening of Γ_X at higher energies, or by an unidentified intrinsic neutral absorber shifting out of the observed *Chandra* bandpass, we conclude that the α_{ox}–z relationship is likely the result of evolution of the accretion mechanism and environment. However, more complex absorption models, such as an ionized or partial covering absorber cannot be ruled out as causing the observed α_{ox}–z de-

pendence, but investigation of such models is beyond the scope of this work.

6.7.2 Γ_X Relationships

We do not find significant evidence for a correlation between the radio-quiet quasar X-ray spectral photon index, Γ_X, and UV luminosity or redshift. This is consistent with results obtained using *XMM* observations of SDSS RQQs (Risaliti & Elvis, 2005), *ASCA* observations of RQQs (Reeves & Turner, 2000), and fitting of composite spectra of $z > 4$ RQQs (e.g., Vignali et al., 2003c, 2005; Shemmer et al., 2006). However, this is in contrast with the work of Dai et al. (2004) and Bechtold et al. (2003). Dai et al. (2004), found evidence for a correlation between Γ_X and L_X using a small sample of gravitationally-lensed RQQs. Bechtold et al. (2003) used Kendall's generalized τ to assess the 2-dimensional correlations between Γ_X and both luminosity and z, and found evidence that Γ_X is correlated with luminosity and anti-correlated with z. In this work we have used linear regression to control for the artifical correlation between luminosity and redshift, and find that there is no significant evidence that Γ_X varies with l_{UV} and z. However, inspection of Figure 6.9 reveals that if Γ_X does depend on l_{UV} and z, then the directions of these trends are likely in agreement with the correlations seen by Bechtold et al. (2003). In addition, a systematic flattening of Γ_X with increasing z has also been seen in *XMM* data by Page et al. (2003), and Vignali et al. (1999) found some evidence that Γ_X is flatter on average for $z \sim 2$ RQQs than for lower z RQQs..

It is interesting to note that at $z \sim 2$ the observed *ROSAT* band has shifted to ~ 0.3–7.2 keV in the quasar rest-frame, overlapping with our rest-frame Chandra band at $z \sim 0$. Considering that the $z \sim 2$ *ROSAT* sources of Yuan et al. (1998a) have values of Γ_X similar to those observed here with *Chandra* at $z \sim 0$, and noting that the source rest-frame energies are approximately the same in these two observed spectra regions, this implies that the soft X-ray spectra of RQQs

may be more complex than a simple power-law. In particular, the Yuan et al. (1998a) sources may exhibit a soft excess at $\lesssim 0.3$ keV, causing steeper hardness ratios. Soft excesses have been seen in good-quality spectra of other low-z RQQs (e.g., Gierliński & Done, 2004), but the origin of this component is still unclear. If such additional complexity exists, it may be the cause of the steeper values of Γ_X seen in the low-z Yuan et al. (1998a) *ROSAT* sources, and thus the strong anti-correlation between Γ_X and z, and consequently α_{ox} and z, seen by Bechtold et al. (2003).

Although, there is no significant evidence for a Γ_X–z relationship, there is marginally significant evidence (2σ) that Γ_X flattens as z increases. While this may be caused by evolution in Γ_X, it may also represent a systematic flattening of the X-ray continuum at harder energies or the result of a Compton reflection component redshifting into the observed 0.3–7 keV band at higher z. Similarly, a soft excess redshifting out of the observed band may also contribute.

An observed anti-correlation between Γ_X and α_{UV} may result from not fitting an intrinsic absorber to most of the X-ray spectra. The sources with redder α_{UV} may have higher N_H, which would result in a lower value of Γ_X inferred from the power-law spectral fit. To test this, we estimated the increase in N_H between $\alpha_{UV} \sim 0.1$ and $\alpha_{UV} \sim 1.2$ needed to produce the observed decrease in Γ_X. We used *Sherpa*'s FAKEIT command to simulate spectra assuming a value of $\Gamma_X = 2.2$ and negligible intrinsic neutral absorption, $N_H = 10^{20}$ cm^{-2}. A value of $\Gamma_X \sim 2.2$ is typical for the bluer sources, $\alpha_{UV} \sim 0.1$. Based on the simulations, N_H must increase to $\sim 10^{21}$ cm^{-2} at $\alpha_{UV} \sim 1.2$ to produce an observed decrease of Γ_X from $\Gamma_X \sim 2.2$ to $\Gamma_X \sim 1.8$. We fit an absorbed power-law for the ten reddest sources with > 50 counts, and found that these sources typically had 3σ upper limits of $N_H \lesssim 5 \times 10^{21}$ cm^{-2}. Values of $N_H \sim 10^{21}$ cm^{-2} are well within the limits on

N_H at $\alpha_{UV} \sim 1.2$, and therefore we cannot rule out the observed Γ_X–α_{UV} anti-correlation as resulting from unidentified intrinsic absorption.

6.7.3 Expectation of Accretion Models

Sobolewska et al. (2004a) and Sobolewska et al. (2004b) explored the general parameter space available for accreting compact sources, and quasars in particular, in respect to geometry of the disk and X-ray emitting region. They show that α_{ox} is most sensitive to (1) the amount of energy dissipated in the corona or (2) the size of the inner flow or a structure and outflow velocity of the coronal flares. Sobolewska et al. (2004b) suggest that the α_{ox}–L_{UV} anti-correlation can be explained by differences in the structure of the X-ray emitting region. They point out that in the framework of the truncated disk and hot inner flow geometry, L_{UV} increases when the disk extends further towards the last stable orbit, while Γ_X steepens, reducing the 2 keV emission. In the patchy corona geometry, the α_{ox}–L_{UV} relationship can be explained by changes in the fraction of gravitational energy dissipated in the corona, where a lower fraction results in a weaker and softer (higher Γ_X) X-ray continuum.

The redshift dependence may similarly be explained as resulting from evolution in the accretion geometry. This would imply that for a given L_{UV}, the high redshift sources have larger radii of the inner hot flow sphere or they generate more flares with higher outflow velocities. Both of these explanations imply that Γ_X is also correlated with UV luminosity and anti-correlated with redshift, where Γ_X steepens with increasing L_{UV} and flattens with increasing z. We find no statistically significant evidence for a correlation between Γ_X and L_{UV}, and only marginally significant evidence for an anti-correlation between Γ_X and z. However, it should be noted that there is considerable uncertainty in the regression coefficients, and their joint confidence region is large. While values of zero

for the regression coefficients cannot be ruled out, it is interesting to note that the trends of Γ_X with L_{UV} and z implied by the regression are consistent with the model predictions. So long as the accretion disk models do not predict too strong of a relationship between Γ_X and L_{UV}, they may still be consistent with our results.

A dual dependency of α_{ox} on both L_{UV} and z must relate, in current scenarios, to variations in $M_{BH}, \dot{m}$, and chemical abundances, and their effects on the accretion disk and corona. Unfortunately, the models do not yet predict a specific relationship between α_{ox} and the model parameters, and thus quantitative comparison of our results with the models is difficult; this will be the subject of future research. In addition, the discussion in this section has been model-dependent, and hopefully magneto-hydrodynamic simulations will provide further insight (e.g., De Villiers et al., 2003; Krolik et al., 2005).

6.7.4 Improving the α_{ox} Analysis

The main source of statistical uncertainty in γ_l and γ_t is the strong degree of correlation between l_{UV} and $t(z)$. For ordinary least-squares regression, the standard errors in the regression coefficients are inflated upwards by a factor of $1/\sqrt{1 - r^2}$, where r is the correlation between l_{UV} and $t(z)$ (Fox, 1997). For our sample, $r = -0.878$, and therefore the standard errors on the regression coefficients are a factor of ≈ 2 higher than if l_{UV} and $t(z)$ were uncorrelated. Because the anti-correlation between l_{UV} and $t(z)$ is so strong, even a small reduction in this correlation can give a large reduction in the standard errors of γ_l and γ_t. For example, selecting a sample to have $r = -0.7$ will result in a reduction in the standard deviations of γ by about 30%. Future α_{ox} studies should try to select samples that minimize r, as has been done by S06. This, along with the larger sample size and range in luminosity probed by S06, is likely the reason why their confidence

regions are smaller (cf., Fig. 6.10).

Using the fact that the coefficient uncertainties are proportional to $1/\sqrt{1-r^2}$, we investigated whether targeting more faint $z \sim 4$ RQQs with *Chandra* will significantly reduce the standard errors in the regression coefficients. Unfortunately, targeting a reasonable number of additional faint $z \sim 4$ RQQs will not significantly improve the estimates of γ_l and γ_t. Targeting 10 additional RQQs uniformly distribution between $45.5 < \log \nu L_\nu(2500\text{Å}) < 46.5$ and $3.7 < z < 5$ will only result in a reduction in the standard errors of $\sim 8\%$. Including 30 additional faint, high-z RQQs reduces the standard errors by $\sim 18\%$, but about half of this reduction is the result of the increased sample size.

6.8   SUMMARY

- There is a significant relationship between $\alpha_{\rm ox}$, L_{UV}, and $t(z)$, and we did not find any evidence that $\alpha_{\rm ox}$ depends on the UV spectral slope. If we remove the 10 suspected BALs ($z < 1.5$ non-detections), the two best $\alpha_{\rm ox}$ regressions are

$$\bar{\alpha}_{\rm ox} = (-9.273 \pm 1.241) + (0.233 \pm 0.026)l_{UV} + (2.870 \pm 0.750) \times 10^{-2}t(z), \ \sigma = 0.125$$

$$\bar{\alpha}_{\rm ox} = (40.97 \pm 27.34) - (1.949 \pm 1.187)l_{UV} + (2.370 \pm 1.288) \times 10^{-2}l_{UV}^2 + \qquad (6.22)$$
$$(2.263 \pm 0.813) \times 10^{-2}t(z), \ \sigma = 0.124$$

Here, the notation $\bar{\alpha}_{\rm ox}$ denotes the average $\alpha_{\rm ox}$ at a given l_{UV} and $t(z)$, and the intrinsic scatter in $\alpha_{\rm ox}$ at a given l_{UV} and $t(z)$ has a dispersion of $\sigma \approx 0.125$ about $\bar{\alpha}_{\rm ox}$. Although the l_{UV}^2 term is not 'statistically significant' in the classical sense, an analysis based on the Kullback-Leibler information found that this model is best supported by the evidence in our data. The KLI analysis found that both models are preferred over a purely quadratic

dependence of $\alpha_{\rm ox}$ on l_{UV}, and over models that parameterized the redshift dependence as linear in z or $\log(1+z)$.

- We used Monte Carlo simulations to show that the $\alpha_{\rm ox}$–z relationship is not a spurious result caused by variability and measurement error. Based on the simulations, we calculate the χ_2^2 of our regression coefficients for model $\mathcal{M}_t$ and find that the probability of observing a χ_2^2 this high or higher is $\lesssim 3 \times 10^{-3}$, under the assumption of no evolution in $\alpha_{\rm ox}$.

- We used Monte Carlo simulations to show that interpretation of Kendall's generalized partial τ is problematic. In particular, Kendall's partial τ for the $\alpha_{\rm ox}$–z correlation is not necessarily expected to be zero when $\alpha_{\rm ox}$ is unrelated to z, given L_{UV}. Moreover, Kendall's partial τ is not necessarily expected to be non-zero when $\alpha_{\rm ox}$ is correlated with z, given L_{UV}. This can have a significant effect on the power of Kendall's partial τ, and therefore care must be taken when using τ to investigate whether $\alpha_{\rm ox}$ evolves or not. Based on our simulations, we are not able to decide for either evolution or no evolution in $\alpha_{\rm ox}$ using τ.

- The $\alpha_{\rm ox}$–z correlation cannot be explained as a result of a systematic steepening of the X-ray continuum at higher energies, as this is inconsistent with the regression of Γ_X on z. Furthermore, the $\alpha_{\rm ox}$–z relationship is not the result of soft X-ray neutral absorption shifting out of the observed band. The observed factor of ~ 8 drop in L_X from $z = 4$ to $z = 0$ is higher than the factor of ~ 2 drop in L_X that would result for an unidentified intrinsic neutral absorber with $N_H \sim 10^{21}$ cm^{-2}. Higher values of N_H are inconsistent with the Γ_X–z results.

- We do not find any evidence for a dependence of Γ_X on UV luminosity, and

only marginally significant evidence (2σ) for a dependence of Γ_X on redshift. The Γ_X–z relationship may be caused by a systematic flattening of the X-ray continuum at higher energies, by redshifting of a Compton reflection component into the observed 0.3–7 keV band, and/or by redshifting of a soft excess out of the observable band.

- We find evidence for an anti-correlation (3.5σ) between Γ_X and the UV spectral slope, where the X-ray continuum hardens as the UV continuum softens. This may be the result of unidentified intrinsic absorption, with the UV redder sources having higher intrinsic N_H, thus causing a flatter inferred X-ray continuum.

CHAPTER 7

OBSERVATIONAL CONSTRAINTS ON THE DEPENDENCE OF RADIO-QUIET

QUASAR X-RAY EMISSION ON BLACK HOLE MASS AND ACCRETION RATE

7.1 CHAPTER ABSTRACT

In this work we use a sample of 318 radio-quiet quasars (RQQ) to investigate the dependence of the ratio of optical/UV flux to X-ray flux, α_{ox}, and the X-ray photon index, Γ_X, on black hole mass, UV luminosity relative to Eddington, and X-ray luminosity relative to Eddington. Our sample is drawn from the literature, with X-ray data from *ROSAT* and *Chandra*, and optical data mostly from the SDSS; 153 of these sources have estimates of Γ_X from *Chandra*. We estimate M_{BH} using standard estimates derived from the Hβ, Mg II, and C IV broad emission lines. Our sample spans a broad range in black hole mass ($10^6 \lesssim M_{BH}/M_\odot \lesssim 10^{10}$), redshift ($0 < z < 4.8$), and luminosity ($10^{43} \lesssim \lambda L_\lambda(2500\text{Å})[\mathrm{erg\ s}^{-1}] \lesssim 10^{48}$). We find that α_{ox} increases with increasing M_{BH} and L_{UV}/L_{Edd}, and decreases with increasing L_X/L_{Edd}. In addition, we confirm the correlation seen in previous studies between Γ_X and M_{BH} and both L_{UV}/L_{Edd} and L_X/L_{Edd}; however, we also find evidence that the dependence of Γ_X of these quantities is not monotonic, changing sign at $M_{BH} \sim 3 \times 10^8 M_\odot$. We argue that the α_{ox} correlations imply that the fraction of bolometric luminosity emitted by the accretion disk, as compared to the corona, increases with increasing accretion rate relative to Eddington, $\dot{m}$. In addition, we argue that the Γ_X trends are caused by a dependence of X-ray spectral index on $\dot{m}$. We discuss our results within the context of accretion models with comptonizing corona, and discuss the implications of the α_{ox} correlations for quasar feedback. To date, this is the largest study of the dependence of RQQ X-ray parameters on black hole mass and related quantities, and

the first to attempt to correct for the large statistical uncertainty in the broad line mass estimates.

7.2   CHAPTER INTRODUCTION

The extraordinary activity associated with quasars involves accretion onto a supermassive black hole (SMBH), with the UV/optical emission arising from a geometrically thin, optically thick cold accretion disk (Shakura & Syunyaev, 1973), and the X-ray continuum arising from a hot, optically thin corona that Compton upscatters the disk UV photons (e.g., Haardt & Maraschi, 1991). In highly accreting objects, like quasars ($0.01 \lesssim L_{bol}/L_{Edd} \lesssim 1$, e.g., Woo & Urry, 2002; Vestergaard, 2004; McLure & Dunlop, 2004; Kollmeier et al., 2006), the X-ray plasma geometry is expected to be that of a hot, possibly patchy, ionized 'skin' that sandwiches the cold disk (e.g., Bisnovatyi-Kogan & Blinnikov, 1977; Liang & Price, 1977; Nayakshin, 2000). However, the evidence for this is not conclusive, and relies on data from X-ray binaries and low-z sources (e.g., see the dicussion by Czerny et al., 2003). Other geometries are possible, including an accretion disk that evaporates into a hot inner flow (e.g., Shapiro et al., 1976; Zdziarski et al., 1999), or a combination of a hot inner flow and a corona that sandwiches the disk (e.g., Poutanen et al., 1997; Sobolewska et al., 2004a). Furthermore, radiation pressure can drive an outflow from the disk into the corona if the two are cospatial, thus altering the physics of the corona (Proga, 2005). Investigations of how quasar X-ray parameters depend on black hole mass, M_{BH}, and accretion rate relative to Eddington, $\dot{m}$, offer important constraints on models of the disk/corona system.

There have been attempts to link the evolution of SMBHs to analytic and semi-analytic models of structure formation (e.g., Kauffmann & Haehnelt, 2000;

Hatziminaoglou et al., 2003; Bromley et al., 2004), where black holes grow by accreting gas funneled towards the center during a galaxy merger until feedback energy from the SMBH expels gas and shuts off the accretion process (e.g., Silk & Rees, 1998; Fabian, 1999; Wyithe & Loeb, 2003; Begelman & Nath, 2005). This 'self-regulated' growth of black holes has recently been successfully applied in smoothed particle hydrodynamics simulations (Di Matteo et al., 2005; Springel et al., 2005). Within this framework, the AGN or quasar phase occurs during the episode of significant accretion that follows the galaxy merger, persisting until feedback from the black hole 'blows' the gas away (e.g., Hopkins et al., 2006a). Hydrodynamic calculations have shown that line pressure is more efficient than thermal pressure at driving an outflow (Proga, 2007), and therefore, the efficiency of AGN feedback depends on the fraction of energy emitted through the UV/disk component as compared to the X-ray/corona component. If the fraction of energy emitted in the UV as compared to the X-ray depends on M_{BH} or $\dot{m}$, then it follows that the efficiency of AGN feedback will also depend on M_{BH} and $\dot{m}$. This has important consequences for models of SMBH growth, as the SMBH may become more or less efficient at driving an outflow depending on its mass and accretion rate. Studies of the dependence of quasar X-ray/UV emission on black hole mass and accretion rate are therefore important as they allow us to constrain a M_{BH}- or $\dot{m}$-dependent feedback efficiency.

Numerous previous studies have searched for a luminosity and redshift dependence of $\alpha_{ox} = -0.384 \log L_X/L_{UV}$, the ratio of X-ray to UV/optical flux (e.g., Avni & Tananbaum, 1982; Wilkes et al., 1994; Yuan et al., 1998a; Vignali et al., 2003b; Strateva et al., 2005; Steffen et al., 2006; Kelly et al., 2007), and Γ_X, the X-ray spectra slope (e.g., Reeves & Turner, 2000; Bechtold et al., 2003; Dai et al., 2004; Risaliti & Elvis, 2005; Grupe et al., 2006). Most studies have found a corre-

lation between α_{ox} and UV luminosity, L_{UV}, while the existence of a correlation between α_{ox} and z is still a matter of debate (e.g., Bechtold et al., 2003; Vignali et al., 2003b; Steffen et al., 2006; Just et al., 2007; Kelly et al., 2007). In addition, studies of Γ_X have produced mixed results. Some authors have claimed a correlation between Γ_X and luminosity (e.g., Bechtold et al., 2003; Dai et al., 2004) or redshift (e.g., Reeves et al., 1997; Vignali et al., 1999; Page et al., 2003), while, others find no evidence for a correlation between Γ_X and L_{UV} or z (e.g., Vignali et al., 2005; Risaliti & Elvis, 2005; Kelly et al., 2007).

A correlation between Γ_X and the $FWHM$ of the Hβ line has also been found (e.g., Boller et al., 1996; Brandt et al., 1997), suggesting a correlation between Γ_X and black hole mass or Eddington ratio (e.g., Laor et al., 1997; Brandt & Boller, 1998). Recently, it has become possible to obtain estimates of M_{BH} for broad line AGN by calibrating results from reverberation mapping (Peterson et al., 2004; Kaspi et al., 2005) for use on single-epoch spectra (Wandel et al., 1999; Vestergaard, 2002; McLure & Jarvis, 2002; Vestergaard & Peterson, 2006; Kelly & Bechtold, 2007). This has enabled some authors to confirm a correlation between Γ_X and either M_{BH} or L_{bol}/L_{Edd} (e.g., Lu & Yu, 1999; Gierliński & Done, 2004; Porquet et al., 2004; Piconcelli et al., 2005; Shemmer et al., 2006), where the X-ray continuum hardens with increasing M_{BH} or softens with increasing L_{bol}/L_{Edd}. In addition, previous work has also found evidence for quasars becoming more X-ray quiet as M_{BH} or L_{bol}/L_{Edd} increase (Brunner et al., 1997; Wang et al., 2004); however, studies involving the dependence of α_{ox} on M_{BH} or L_{bol}/L_{Edd} have remained rare compared to studies of Γ_X. It is important to note that the correlations inferred in previous work generally employ broad line mass estimates in combination with a constant bolometric correction. Therefore, most of the correlations found in previous work are, strictly speaking, between Γ_X or α_{ox} and the

estimates $M_{BH} \propto L_\lambda^\gamma FWHM^2$ and $L_{bol}/L_{Edd} \propto L_\lambda^{1-\gamma} FWHM^{-2}$, where $\gamma \sim 0.5$.

In this work, we investigate the dependence of $\alpha_{\rm ox}$ and Γ_X on black hole mass, optical/UV luminosity relative to Eddington, and X-ray luminosity relative to Eddington. We combine the main Sloan Digital Sky Survey (SDSS) sample of Strateva et al. (2005) with the sample of Kelly et al. (2007), creating a sample of 318 radio-quiet quasars (RQQ) with X-ray data from *ROSAT* and *Chandra*, and optical spectra mostly from the SDSS; 153 of these sources have estimates of Γ_X from *Chandra*. Because the X-ray emission in radio-loud sources can have an additional component from the jet (e.g., Zamorani et al., 1981; Wilkes & Elvis, 1987), we focus our analysis on the radio-quiet majority. Our sample has a detection fraction of 87% and spans a broad range in black hole mass ($10^6 \lesssim M_{BH}/M_\odot \lesssim 10^{10}$), redshift ($0 < z < 4.8$), and luminosity ($10^{43} \lesssim \lambda L_\lambda(2500\text{Å})[\text{erg s}^{-1}] \lesssim 10^{48}$), enabling us to effectively look for trends regarding $\alpha_{\rm ox}$ and Γ_X.

The outline of this paper is as follows. In § 7.3 we describe the construction of our sample, and in § 7.4 we describe the procedure we used to fit the optical continuum and emission lines. In § 7.5 we describe how we obtain broad line mass estimates, our bolometric correction, and argue that a constant bolometric correction provides a poor estimate of the bolometric luminosity. In § 7.6 we describe the results from a regression analysis of $\alpha_{\rm ox}$ on $M_{BH}, L_{UV}/L_{Edd}$, and L_X/L_{Edd}, and in § 7.7 we report evidence for a non-monotonic dependence of Γ_X on either $M_{BH}, L_{UV}/L_{Edd}$, and L_X/L_{Edd}. In § 7.8 we discuss our results within the context of AGN disk/corona models, and we discuss the implications for a dependence of quasar feedback efficiency on black hole mass or accretion rate. In § 7.9 we summarize our main results.

We adopt a cosmology based on the the WMAP best-fit parameters ($h = 0.71, \Omega_m = 0.27, \Omega_\Lambda = 0.73$, Spergel et al., 2003). For ease of notation, we de-

fine $L_{UV}\nu \equiv L_\nu(2500\text{Å}), L_X \equiv \nu L_\nu(2 \text{ keV}), l_{UV} \equiv \log \nu L_\nu(2500\text{Å})$, and $m_{BH} \equiv \log M_{BH}/M_\odot$.

7.3 SAMPLE CONSTRUCTION

In this analysis we combine 169 RQQs from Kelly et al. (2007, hereafter K07) with 149 RQQs from the main SDSS sample of Strateva et al. (2005, hereafter S05) to create a sample of 318 RQQs. Out of these 318 sources, 276 (86.8%) are detected in the X-rays. The $z \lesssim 4$ sources from the K07 sample were selected by cross-correlating the SDSS DR3 quasar catalogue (Schneider et al., 2005) with the *Chandra* public archive as of 2005 February 22. The $z \gtrsim 4$ sources from the K07 sample consist of targeted *Chandra* RQQs taken from the literature (Bechtold et al., 2003; Vignali et al., 2001, 2003a), and new observations reported by K07. The sources taken from S05 were selected from the SDSS to be contained within the inner $19'$ of *ROSAT* PSPC pointings with exposure times > 11 ksec. The X-ray data for both samples are as reported by S05 and K07.

Both the S05 and K07 samples consist only of radio-quiet quasars. We focus our analysis on the radio-quiet majority because the radio-loud sources have an additional component of X-ray emission arising from the jet (e.g., Zamorani et al., 1981; Wilkes & Elvis, 1987; Worrall et al., 1987). In addition, both S05 and K07 omitted BAL QSOs when possible. It is necessary to remove the BAL QSOs because their high column density gives them the appearance of being X-ray weak (e.g., Green et al., 2001; Gallagher et al., 2002, 2006), potentially biasing our analysis. However, neither S05 nor K07 were able to remove the high-ionization BAL quasars for $z < 1.5$, as their identification requires observations of the C IV line. In addition, low-ionization BALs can be identified at $0.45 < z < 2.25$ based on Mg II absorption. Reichard et al. (2003) found the fraction of BALs in the SDSS

to be $\sim 14\%$, and therefore we expect there to be 25 ± 5 BALs in our sample at $z < 1.5$. This number may be higher if one relaxes the definition of a BAL quasar (Trump et al., 2006).

We exclude five sources from the S05 sample due to significant intrinsic narrow UV absorption or obvious host galaxy contamination: Source SDSS J103747.4-001643.9 ($z = 1.500$) had significant C IV absorption, and sources SDSS J124520.7-002128.1 ($z = 2.354$) and SDSS J103709.8+000235.2 ($z = 2.679$) had significant absorption in both C IV and Lα. These three sources were omitted because the absorption prohibits obtaining an accurate line width measurement, necessary for broad line mass estimates, and to ensure that the X-ray emission under study is not effected by the absorption. Sources SDSS J230440.6-082220.8 ($z = 0.201$) and SDSS J023306.0+003856.4 ($z = 0.244$) have a significant host-galaxy component in their spectra. In addition, we exclude source SDSS J144340.8+585653.2 ($z = 4.278$) from the K07 sample because it has significant UV absorption. We removed source SDSS J142414.1+421400.1 ($z = 1.608$) from the K07 sample and source SDSS J170441.4+604430.5 (PG 1704+608, $z = 0.372$) from the S05 sample, as both sources are radio-loud.

We could not estimate black hole masses for sources SDSS J083206.0+524359.3 ($z = 1.573$), SDSS J144231.7+011055.3 ($z = 4.507$), and PC 0910+5625 ($z = 4.035$). All three of these sources are from the K07 sample. The region containing the Mg II emission line for SDSS J083206.0+524359.3 was missing from the SDSS spectrum, the emission lines are too weak for SDSS J144231.7+011055.3, and an optical spectrum was not available for PC 0910+5625.

7.4 OPTICAL/UV SPECTRAL FITS

Optical spectra were obtained for most sources from the SDSS. We also obtained spectra for some of the high redshift quasars from Anderson et al. (2001), Péroux et al. (2001), and Constantin et al. (2002). The values of L_{UV} and α, $L_\nu \propto \nu^{-\alpha}$, for the K07 sources are taken from K07. We processed the optical spectra for the S05 sources in the same manner as for the K07 sources. We do this for consistency and because S05 did not correct for quasar iron emission.

7.4.1 Continuum Fitting

As described by K07, we corrected the optical spectra for Galactic absorption using the $E(B-V)$ values taken from Schlegel et al. (1998), as listed in the NASA/IPAC Extragalactic Database (NED), and the extinction curve of Cardelli et al. (1989), assuming a value of $A_V/E(B-V) = 3.1$. We model the continuum as a power law of the form $f_\nu \propto \nu^{-\alpha}$, and the Fe emission as a scaled and broadened iron template extracted from I Zw I. The optical iron template was extracted by Véron-Cetty et al. (2004), and the UV iron template was extracted by Vestergaard & Wilkes (2001). The continuum and iron emission were fit simultaneously using the Levenberg-Marquardt method for nonlinear χ^2-minimization. Continuum flux densities were then estimated using the power law parameters.

We were not able to use a power-law fit to calculate L_{UV} for the $z \lesssim 0.4$ sources, as the SDSS spectral range for these sources does not contain the rest-frame UV continuum. Instead, we use the luminosity of the broad component of the Hβ emission line, $L_{H\beta}$, as a proxy for L_{UV}. It is preferable to use the broad Hβ emission line luminosity over, say, the optical continuum luminosity as a proxy for L_{UV} because the broad Hβ emission line is not contaminated by emission from the host galaxy, and thus should provide an approximately unbiased estimate of

L_{UV}. Host-galaxy contamination is likely negligible for all $z > 0.4$ sources, as $\nu L_\nu^* \sim 10^{44}$ ergs s^{-1} at 2500Å for galaxies (Budavári et al., 2005). We fit a power-law relationship between L_{UV} and $L_{\mathrm{H}\beta}$ using the 44 sources at $0.4 < z < 0.9$ for which a measurement of both quantities if available. We used the linear regression method of Kelly (2007), which allows for measurement errors in both variables, and find

$$\frac{\lambda L_\lambda(2500\text{Å})}{10^{44} \text{ ergs s}^{-1}} = (1.556 \pm 0.282) \left(\frac{L_{\mathrm{H}\beta}}{10^{42} \text{ ergs s}^{-1}}\right)^{0.768 \pm 0.086}. \tag{7.1}$$

The intrinsic scatter about this relationship is ≈ 0.179 dex, implying a potential uncertainty in l_{UV} inferred from this relationship of the same magnitude. There was no trend in the residuals with either z or $L_{\mathrm{H}\beta}$, implying that Equation (7.1) should give unbiased estimates of l_{UV} for the $z < 0.4$ sources. Values of L_{UV} were estimated using Equation (7.1) for both the K07 and S05 $z < 0.4$ sources, a total of 42 sources.

The distributions of L_{UV} and L_X as a function of redshift are shown in Figure 7.1. We calculate the ratio of optical to X-ray flux (Tananbaum et al., 1979) as

$$\alpha_{\mathrm{ox}} = -0.384 \log(f_{2\mathrm{keV}}/f_{2500}), \tag{7.2}$$

where $f_{2\mathrm{keV}}$ and f_{2500} are the rest-frame flux densities at 2 keV and 2500Å, respectively. If the flux density from 2500Å to 2 keV is a simple power law, then α_{ox} is the spectral slope of this continuum, and thus α_{ox} may be thought of as a crude estimate of the shape of the ionizing continuum. The parameter α_{ox} is an important parameter for model comparison, as it summarizes the amount of energy emitted in the X-ray region (most likely a Comptonized component), compared with that emitted in the optical-UV (accretion disk component). The distribution of α_{ox} as a function of L_{UV} and z are also shown in Figure 7.1.
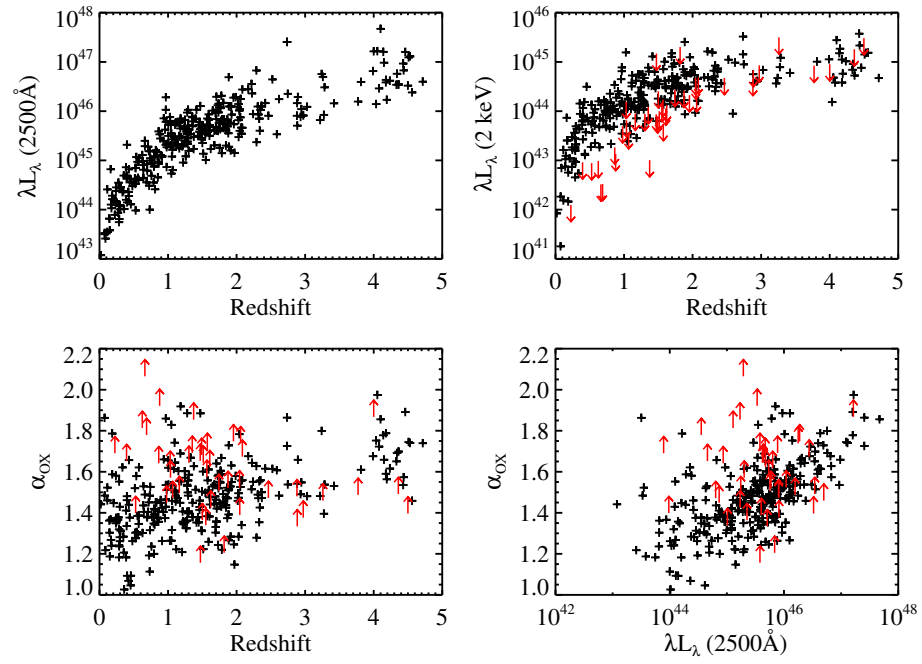
Figure 7.1 The (L_{UV}, L_X, z) distribution of our sample. Non-detections are denoted by red arrows.

The error on α_{ox} is the result of measurement errors on the UV and X-ray flux, as well as error caused by quasar variability over the different epochs for the UV and X-ray observations. In general, the measurement errors on l_{UV} are negligible compared to the error on l_X. S05 estimates a error on the X-ray flux of ~ 0.23 dex, including both the contributions from measurement error and variability. Typical long-term X-ray variability for Seyfert 1s is 20%–40% with no obvious trend with luminosity (Grupe et al., 2001; Uttley et al., 2002; Markowitz et al., 2003). The measurement errors in l_X for the K07 sample are typically ~ 0.07 dex. Assuming X-ray variability amplitudes of 30%, this implies typical uncertainties in the X-ray luminosity of ~ 0.15 dex. Therefore, we estimate the uncertainty on α_{ox} to be ~ 0.06 for the K07 sources and ~ 0.09 for the S05 sources.

7.4.2   Line Profile Extraction and Fitting

We extracted the Hβ, Mg II, and C IV emission lines in order to use their widths in our black hole mass estimates (e.g., Vestergaard, 2002; McLure & Dunlop, 2002; Vestergaard & Peterson, 2006). These line were extracted by first subtracting the continuum and Fe emission, interpolating over any narrow absorption features, and modelling all lines within the extraction region as a sum of Gaussian functions. Any nearby lines were then subtracted, leaving only the broad emission line profile. In all cases the line profile extraction was done interactively and every line fit was inspected visually.

For Hβ, we extracted the region within $\pm 2\times 10^4\,\mathrm{km\,s^{-1}}$ of 4861Å, where we use the standard convention that negative velocities are blueward of a given wavelength. The Hβ profile was modeled as a sum of 2–3 Gaussian functions. The [O III] λ4959Å and [O III] λ5007Å lines were modeled as a sum of 1–2 Gaussian functions, depending on the signal-to-noise of the lines. A sum of two Gaussian functions was used for the higher S/N lines because the [O III] line profiles are

not exactly a Gaussian function; the individual Gaussian components are not considered to be physically distinct components. The widths of the narrow Gaussian functions for Hβ and [O III] lines were fixed to be equal to eachother. The [O III] lines and the narrow component of the Hβ line were then subtracted, leaving the broad component of Hβ.

For Mg II, we extracted the region within $\pm 2 \times 10^4$ km s^{-1} of 2800Å. There are no nearby non-iron emission lines that Mg II is blended with, so the extraction is trivial after removing the Fe and continuum emission.

For C IV, we extracted the region within -2×10^4 km s^{-1} and 3×10^4 km s^{-1} of 1549Å. The C IV line was modeled as a sum of 2–3 Gaussian functions, and He II λ1640 and O III] λ1665 were modeled as a sum of 1–2 Gaussian functions each. After obtaining estimates of the He II and O III] profiles, we subtracted these components. We did not model the N IV] λ1486 emission line as this line is typically weak and lost in the C IV wings.

In order to estimate M_{BH}, it is necessary to measure the $FWHM$ of the emission lines. After extracting the line profiles, we estimate the $FWHM$ for the Hβ, Mg II, and C IV emission lines by fitting them to a sum of 1–5 Gaussian functions, enabling us to obtain a smooth representation of each line. In contrast to our profile extraction technique, we choose the number of Gaussian functions to minimize the Bayesian Information Criterion (BIC, Schwartz, 1979). The BIC is a common criterion to use for selecting the number of parameters in a model (e.g., see Hastie, Tibshirani, & Friedman, 2001); the model that minimizes the BIC is approximately the model that is most supported by the data. For Gaussian errors, as assumed in this work, the BIC is simply a modification to the standard χ^2 statistic:

$$BIC = \chi^2 + 3K \ln n, \tag{7.3}$$

where K is the number of Gaussian functions used, $3K$ is the number of free parameters, and n is the number of data points used in the fit. Using the BIC to 'fit' the number of Gaussian functions thus allows us more flexibility in obtaining a smooth representation of the line profile, as we are not choosing the number of Gaussian functions arbitrarily. Once a smooth representation is obtained, we automatically measure the $FWHM$ directly from the best fit line profile.

The standard errors on $FWHM$ are estimated using a bootstrap method. We simulated 100 'observed' emission lines by adding random Gaussian noise to the best fit line profile with standard deviation equal to the noise level of the spectrum, including the propagated errors from the continuum and iron emission fitting. We then fit each of the simulated emission lines, keeping the number of Gaussian functions fixed at the number found from fitting the original profile, and measured the $FWHM$ for each simulated line. The standard error on $FWHM$ was then estimated as the standard deviation of the $FWHM$ values measured from the simulated line profiles.

7.5   ESTIMATING M_{BH}

Recently, reverberation mapping studies of broad line AGN (e.g., Peterson et al., 2004) established a correlation between the broad line region (BLR) size, R, and the continuum luminosity (the R–L relationship, e.g., Kaspi et al., 2005; Bentz et al., 2006). This has made it possible to estimate black hole virial mass $M_{BH} = fv^2R/G$ for individual sources, where the BLR velocity v is estimated from the width of an emission line (e.g., Wandel et al., 1999; Vestergaard, 2002; McLure & Dunlop, 2002; Vestergaard & Peterson, 2006). We choose the proportionality constant to give broad line mass estimates consistent with the M_{BH}–σ relationship (Gebhardt et al., 2000a; Merritt & Ferrarese, 2001; Tremaine et al., 2002),

$f = 1.4 \pm 0.45$ (Onken et al., 2004). An estimate of the Eddington luminosity can be computed as $L_{Edd} = 1.3 \times 10^{38} M_{BH}/M_\odot$ erg s^{-1}.

7.5.1 Black Hole Mass Estimates from Hβ, Mg II, and C IV

In this work we estimate M_{BH} from the Hβ, Mg II, and C IV emission lines. We use the relationship of Vestergaard & Peterson (2006) to estimate M_{BH} from the Hβ and C IV emission lines, and the relationship of Vestergaard et al. (2008, in preparation) to estimate M_{BH} from the Mg II emission lines. These relationship are:

$$m_{\mathrm{H}\beta} = 6.67 + 2\log\left[\frac{FWHM(\mathrm{H}\beta)}{1000 \text{ km s}^{-1}}\right] + 0.63\log\left[\frac{L(\mathrm{H}\beta)}{10^{42} \text{ erg s}^{1}}\right] \quad (7.4)$$

$$m_{\mathrm{MgII}} = 6.86 + 2\log\left[\frac{FWHM(\mathrm{MgII})}{1000 \text{ km s}^{-1}}\right] + 0.50\log\left[\frac{\lambda L_\lambda(3000\text{Å})}{10^{44} \text{ erg s}^{1}}\right] \quad (7.5)$$

$$m_{\mathrm{CIV}} = 6.66 + 2\log\left[\frac{FWHM(\mathrm{CIV})}{1000 \text{ km s}^{-1}}\right] + 0.53\log\left[\frac{\lambda L_\lambda(1350\text{Å})}{10^{44} \text{ erg s}^{1}}\right], \quad (7.6)$$

where we have used the notation that $m_{\mathrm{H}\beta}$ is the Hβ-based estimate of $\log M_{BH}/M_\odot$, and likewise for m_{MgII} and m_{CIV}. The calibration for the Mg II mass estimates was calculated to ensure that they are consistent with the mass estimates based on Hβ and C IV. We have 49 sources with both Hβ and Mg II mass estimates, and 73 sources with both C IV and Mg II mass estimates. Both samples show consistent mass estimates between the different emission lines, within the intrinsic uncertainty in the broad line mass estimates (~ 0.4 dex).

We will denote the broad line mass estimates as $\hat{M}_{BL}$, and $\hat{m}_{BL} \equiv \log \hat{M}_{BL}/M_\odot$. It is important to distinguish between $\hat{M}_{BL}$ and M_{BH}, as $\hat{M}_{BL} \propto L^\gamma FWHM^2, \gamma \sim 0.5$, is an estimate of M_{BH} derived from reverberation mapping, and thus in general $\hat{M}_{BL} \neq M_{BH}$. The statistical uncertainty needs to be taken into account when analyzing correlations involving derived quantities like $\hat{M}_{BL} \propto L^\gamma FWHM^2$, as they can bias the results (Kelly & Bechtold, 2007; Kelly, 2007).

The uncertainty in f increases the formal statistical uncertainty in the broad line estimates of M_{BH} to ~ 0.46 dex. Our adopted formal uncertainty of ~ 0.46 dex is merely statistical, and additional systematic uncertainties in reverberation mapping may contribute (Krolik, 2001; Collin et al., 2006). Because our adopted uncertainty of ~ 0.46 describes the scatter in broad line mass estimates about the reverberation mapping estimates, as calibrated via the M_{BH}–σ relationship, the regression results found in this work should be understood as results that could have been obtained if we had reverberation-based M_{BH} for the sources in this work. However, instead of reverberation-based mass estimates, we have broad line mass estimates with 'measurement error' equal to ~ 0.46 dex with respect to the reverberation-based mass estimates, thus increasing the uncertainty from the regression analysis.

For most sources, measurement errors on $FWHM$ and L_λ did not significantly contribute to the uncertainty on M_{BH}. If there were two emission lines in the same spectrum we averaged the two mass estimates, where the average was weighted by the uncertainties in the two estimates. The distribution of M_{BH} as a function of z for our sample is shown in Figure 7.2. The broad line masss estimates for the sources in our sample are reported in Table 7.1.
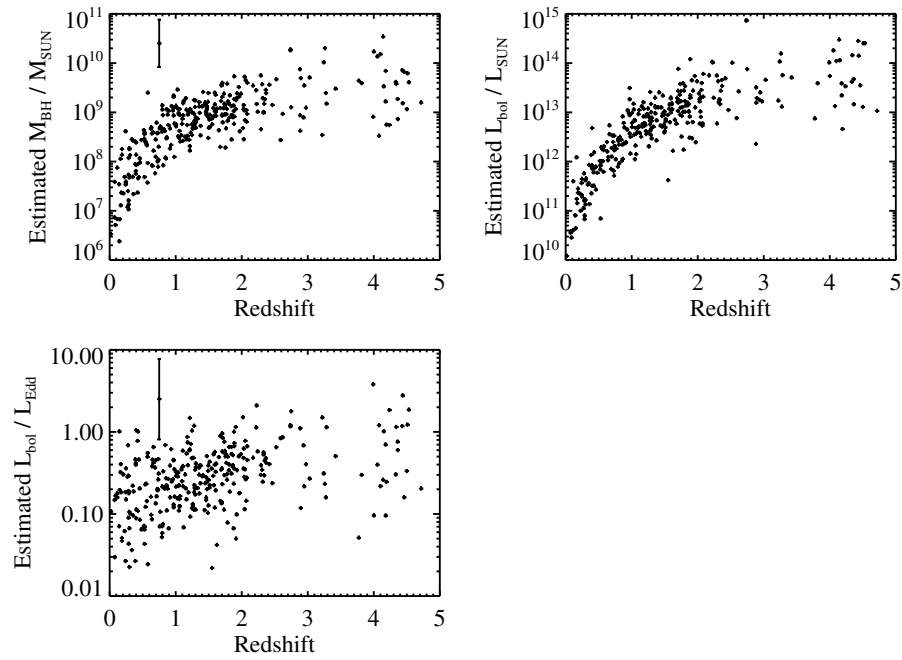
Figure 7.2 The distributions of estimated M_{BH}, L_{bol}, and L_{bol}/L_{Edd}, as a function of z for our sample. The data points with error bars in the left two plots are fictitious data points illustrating the typical error in $\hat{M}_{BL}$ and $\hat{L}_{bol}/\hat{L}_{Edd}$.

Table 7.1.   Black Hole Parameters of the Sample

α (J2000)	δ (J2000)	$\log \hat{M}_{BL}/M_\odot$	$\log L_X/\hat{L}_{Edd}$[a]	$\log L_{UV}/\hat{L}_{Edd}$[b]	
00 02 30.7	+00 49 59.0	1.352	9.20 ± 0.45	-2.43 ± 0.45	-1.44 ± 0.45
00 06 54.1	-00 15 33.4	1.725	9.08 ± 0.29	-2.58 ± 0.30	-1.07 ± 0.29
00 22 10.0	+00 16 29.3	0.574	7.96 ± 0.32	-2.73 ± 0.32	-1.11 ± 0.32
00 27 52.4	+00 26 15.7	0.205	7.35 ± 0.45	-2.42 ± 0.46	-1.26 ± 0.49
00 31 31.4	+00 34 20.2	1.735	9.17 ± 0.29	-2.66 ± 0.29	-1.31 ± 0.29
00 50 06.3	-00 53 19.0	4.331	9.61 ± 0.41	-2.89 ± 0.42	-1.19 ± 0.41
00 57 01.1	+14 50 03.0	0.623	8.66 ± 0.32	-3.79 ± 0.32	-1.66 ± 0.32
00 59 22.8	+00 03 01.0	4.178	9.21 ± 0.38	-2.28 ± 0.41	-0.74 ± 0.38
01 06 19.2	+00 48 22.0	4.437	9.18 ± 0.38	-1.96 ± 0.39	-0.46 ± 0.38
01 13 05.7	+15 31 46.5	0.576	9.39 ± 0.32	-3.47 ± 0.32	-2.46 ± 0.32
01 13 09.1	+15 35 53.6	1.806	9.19 ± 0.29	-2.32 ± 0.29	-1.13 ± 0.29
01 15 37.7	+00 20 28.7	1.275	9.25 ± 0.45	-2.75 ± 0.45	-1.81 ± 0.45
01 26 02.2	-00 19 24.1	1.765	8.96 ± 0.29	-2.58 ± 0.29	-1.01 ± 0.29
01 33 40.4	+04 00 59.0	4.150	9.52 ± 0.38	-2.37 ± 0.39	-0.65 ± 0.38
01 34 21.5	+33 07 56.6	4.530	9.60 ± 0.39	-2.53 ± 0.41	-0.61 ± 0.39
01 48 12.2	+00 01 53.3	1.704	9.59 ± 0.29	-3.08 ± 0.30	-1.41 ± 0.29
01 48 21.0	-00 02 25.8	0.930	8.33 ± 0.45	-2.10 ± 0.45	-1.05 ± 0.45
01 52 58.7	+01 05 07.4	0.647	7.90 ± 0.33	-2.17 ± 0.33	-1.11 ± 0.33
01 53 09.1	+00 52 50.1	1.161	9.05 ± 0.45	-3.11 ± 0.52	-1.67 ± 0.45
01 56 50.3	+00 53 08.5	1.652	8.63 ± 0.30	-2.11 ± 0.31	-0.73 ± 0.30

Table 7.1—Continued

α (J2000)	δ (J2000)	$\log \hat{M}_{BL}/M_\odot$	$\log L_X/\hat{L}_{Edd}$[a]	$\log L_{UV}/\hat{L}_{Edd}$[b]	
01 59 50.2	+00 23 40.8	0.162	8.05 ± 0.45	-2.30 ± 0.45	-1.34 ± 0.48
02 01 18.7	-09 19 35.8	0.660	8.58 ± 0.32	-2.58 ± 0.32	-1.30 ± 0.32
02 08 45.5	+00 22 36.1	1.885	9.72 ± 0.29	-3.05 ± 0.30	-1.15 ± 0.29
02 09 44.7	+05 17 14.0	4.140	10.5 ± 0.38	-3.47 ± 0.39	-1.44 ± 0.38
02 10 00.2	-10 03 54.3	1.960	9.32 ± 0.39	-2.33 ± 0.39	-1.69 ± 0.39
02 22 25.5	-09 02 58.5	0.224	7.68 ± 0.45	-3.70 ± 0.45	-1.90 ± 0.48
02 22 26.1	-08 57 01.3	0.166	7.10 ± 0.45	-1.81 ± 0.45	-1.12 ± 0.48
02 23 56.3	-08 57 07.8	1.576	8.92 ± 0.29	-2.90 ± 0.29	-1.45 ± 0.29
02 24 35.9	-09 00 01.3	1.611	8.93 ± 0.30	-2.57 ± 0.30	-1.43 ± 0.30
02 32 17.7	-07 31 19.9	1.163	8.44 ± 0.46	-2.61 ± 0.46	-1.32 ± 0.46
02 33 06.3	+00 46 14.5	2.290	8.65 ± 0.40	-2.12 ± 0.40	-1.39 ± 0.40
02 33 25.3	+00 29 14.9	2.017	8.67 ± 0.29	-1.89 ± 0.29	-0.70 ± 0.29
02 33 33.2	+01 03 33.1	2.058	9.64 ± 0.29	-3.43 ± 0.29	-1.51 ± 0.29
02 33 59.7	+00 49 38.6	2.522	9.67 ± 0.38	-2.73 ± 0.38	-1.13 ± 0.38
02 34 14.6	+00 57 07.9	0.269	7.07 ± 0.45	-2.20 ± 0.45	-1.25 ± 0.48
02 41 05.8	-08 11 53.2	0.978	8.09 ± 0.47	-2.51 ± 0.47	-1.34 ± 0.47
02 41 10.0	+00 23 01.4	0.790	8.65 ± 0.45	-3.12 ± 0.46	-2.21 ± 0.45
02 44 01.9	-01 34 03.0	4.053	10.1 ± 0.38	-3.56 ± 0.40	-1.02 ± 0.38
02 48 54.3	+18 02 49.9	4.430	9.85 ± 0.38	-2.39 ± 0.39	-1.00 ± 0.38
02 53 56.1	+00 10 57.5	1.699	8.95 ± 0.30	-2.60 ± 0.30	-1.39 ± 0.30

Table 7.1—Continued

α (J2000)	δ (J2000)	$\log \hat{M}_{BL}/M_\odot$	$\log L_X/\hat{L}_{Edd}$[a]	$\log L_{UV}/\hat{L}_{Edd}$[b]	
02 54 22.9	+00 07 58.8	1.213	8.31 ± 0.45	-2.16 ± 0.45	-1.10 ± 0.45
02 54 38.4	+00 21 32.8	2.463	9.14 ± 0.39	-2.58 ± 0.39	-1.35 ± 0.39
02 55 01.2	+00 17 45.5	0.359	7.33 ± 0.52	-1.50 ± 0.52	-1.43 ± 0.55
02 55 05.7	+00 25 23.2	0.353	8.42 ± 0.45	-2.38 ± 0.45	-1.84 ± 0.48
02 55 10.6	-00 07 12.9	1.686	9.22 ± 0.29	-2.97 ± 0.29	-1.72 ± 0.29
02 55 13.0	+00 06 39.5	1.882	8.76 ± 0.29	-2.14 ± 0.29	-1.12 ± 0.29
02 55 25.7	+00 21 09.9	0.526	7.68 ± 0.61	-2.85 ± 0.61	-1.81 ± 0.62
02 55 28.9	-00 02 19.1	1.435	9.16 ± 0.45	-2.75 ± 0.45	-1.77 ± 0.45
02 55 56.0	+00 14 57.4	0.863	8.48 ± 0.45	-3.33 ± 0.45	-1.66 ± 0.45
02 59 59.7	+00 48 13.6	0.892	9.03 ± 0.45	-3.24 ± 0.45	-2.03 ± 0.45
03 11 15.2	-17 22 47.3	4.000	10.2 ± 0.38	-3.39 ± 0.38	-1.13 ± 0.38
03 14 27.5	-01 11 52.3	0.387	7.94 ± 0.32	-2.26 ± 0.32	-1.19 ± 0.36
04 03 56.6	-17 03 24.1	4.236	8.74 ± 0.38	-1.84 ± 0.40	-0.23 ± 0.38
04 19 50.9	-57 16 13.1	4.460	9.82 ± 0.38	-3.05 ± 0.41	-0.73 ± 0.38
07 34 05.3	+32 03 15.4	2.082	8.96 ± 0.29	-2.87 ± 0.29	-1.41 ± 0.29
07 55 02.1	+22 03 46.9	0.399	8.38 ± 0.32	-3.07 ± 0.32	-1.87 ± 0.37
07 55 35.6	+40 58 03.0	2.417	9.55 ± 0.38	-3.11 ± 0.40	-1.50 ± 0.38
07 55 40.0	+41 11 19.1	0.967	9.33 ± 0.45	-3.90 ± 0.47	-1.66 ± 0.45
07 55 45.6	+40 56 43.6	2.348	9.26 ± 0.38	-2.80 ± 0.40	-1.35 ± 0.38
07 57 51.7	+37 45 54.3	1.242	8.75 ± 0.46	-2.74 ± 0.46	-1.17 ± 0.46

Table 7.1—Continued

α (J2000)	δ (J2000)	$\log \hat{M}_{BL}/M_\odot$	$\log L_X/\hat{L}_{Edd}$[a]	$\log L_{UV}/\hat{L}_{Edd}$[b]	
07 58 21.0	+37 55 33.2	0.859	8.47 ± 0.33	-2.54 ± 0.33	-1.61 ± 0.33
07 58 25.9	+37 46 28.8	1.500	8.68 ± 0.30	-2.39 ± 0.30	-1.44 ± 0.30
08 13 03.9	+45 57 13.4	0.297	8.25 ± 0.45	-3.04 ± 0.45	-2.35 ± 0.49
08 13 39.4	+46 03 41.3	1.986	9.37 ± 0.29	-2.71 ± 0.29	-1.20 ± 0.29
08 19 51.4	+36 49 50.8	0.736	8.03 ± 0.33	-1.97 ± 0.33	-1.15 ± 0.33
08 45 26.6	+34 31 02.0	2.046	8.83 ± 0.42	-2.36 ± 0.42	-1.35 ± 0.42
08 47 16.0	+37 32 18.1	0.453	8.06 ± 0.32	-1.68 ± 0.32	-1.56 ± 0.32
08 47 45.9	+37 53 59.1	1.908	9.06 ± 0.29	-2.46 ± 0.29	-1.08 ± 0.29
08 48 18.9	+37 40 09.1	0.307	8.12 ± 0.45	-2.59 ± 0.45	-1.78 ± 0.48
08 48 24.7	+37 31 14.1	1.567	9.08 ± 0.29	-2.96 ± 0.29	-1.46 ± 0.29
08 48 34.3	+37 23 34.1	1.741	8.95 ± 0.29	-2.63 ± 0.29	-1.31 ± 0.29
08 48 38.6	+37 45 41.7	2.618	8.96 ± 0.45	-2.09 ± 0.45	-0.83 ± 0.45
08 49 05.1	+44 57 14.8	1.259	8.68 ± 0.45	-2.56 ± 0.45	-1.71 ± 0.45
08 49 05.1	+44 57 14.8	1.259	8.66 ± 0.45	-2.63 ± 0.45	-1.69 ± 0.45
08 49 43.7	+45 00 24.3	1.592	9.01 ± 0.29	-2.52 ± 0.29	-1.19 ± 0.29
09 09 11.6	+54 09 52.6	0.142	7.53 ± 0.45	-2.77 ± 0.45	-1.83 ± 0.48
09 09 26.7	+42 42 28.5	0.240	7.53 ± 0.45	-2.38 ± 0.45	-1.47 ± 0.49
09 10 06.3	+43 08 59.3	0.740	8.29 ± 0.32	-2.73 ± 0.32	-1.36 ± 0.32
09 10 23.2	+42 45 54.9	1.303	8.68 ± 0.47	-2.69 ± 0.47	-1.16 ± 0.47
09 10 29.0	+54 27 19.0	0.525	8.55 ± 0.32	-2.54 ± 0.32	-1.82 ± 0.32

Table 7.1—Continued

α (J2000)	δ (J2000)	$\log \hat{M}_{BL}/M_\odot$	$\log L_X/\hat{L}_{Edd}$[a]	$\log L_{UV}/\hat{L}_{Edd}$[b]	
09 10 29.0	+54 27 19.0	0.525	8.58 ± 0.32	-2.57 ± 0.32	-1.85 ± 0.32
09 12 10.3	+05 47 42.1	3.240	10.0 ± 0.38	-3.45 ± 0.41	-1.37 ± 0.38
09 18 28.6	+51 39 32.1	0.185	8.25 ± 0.45	-3.61 ± 0.46	-2.08 ± 0.48
09 18 47.5	+06 47 04.7	0.821	8.65 ± 0.33	-2.73 ± 0.33	-1.55 ± 0.33
09 32 59.0	+55 04 33.6	1.512	8.69 ± 0.29	-2.25 ± 0.29	-1.67 ± 0.29
09 33 59.3	+55 15 50.7	1.863	8.89 ± 0.30	-2.54 ± 0.30	-1.18 ± 0.30
09 33 59.3	+55 15 50.8	1.863	8.91 ± 0.30	-2.18 ± 0.30	-1.20 ± 0.30
09 35 30.1	+61 36 50.0	1.292	9.10 ± 0.45	-2.64 ± 0.45	-1.19 ± 0.45
09 35 31.2	+61 29 02.1	1.865	8.92 ± 0.29	-2.57 ± 0.29	-1.34 ± 0.29
09 36 39.1	+61 19 18.1	1.329	8.86 ± 0.45	-3.18 ± 0.45	-1.44 ± 0.45
09 41 33.7	+59 48 11.3	0.967	9.31 ± 0.45	-2.53 ± 0.45	-1.14 ± 0.45
09 42 31.9	+47 11 21.4	1.788	9.21 ± 0.29	-3.02 ± 0.29	-1.49 ± 0.29
09 43 45.2	+46 51 54.7	1.246	8.53 ± 0.45	-2.31 ± 0.45	-0.82 ± 0.45
09 50 24.0	+56 19 46.7	1.912	9.24 ± 0.39	-2.90 ± 0.40	-2.16 ± 0.39
09 51 30.2	+59 40 37.1	1.056	9.06 ± 0.45	-2.94 ± 0.45	-1.74 ± 0.45
09 51 51.6	+59 44 30.0	2.338	9.17 ± 0.38	-2.79 ± 0.39	-1.68 ± 0.38
09 52 40.2	+51 52 50.0	0.553	8.36 ± 0.32	-2.50 ± 0.32	-1.50 ± 0.32
09 52 43.0	+51 51 21.1	0.861	9.05 ± 0.45	-2.64 ± 0.45	-1.58 ± 0.45
09 55 05.6	+59 35 17.6	0.912	8.26 ± 0.45	-2.40 ± 0.45	-1.30 ± 0.45
09 55 11.3	+59 40 32.2	4.340	9.57 ± 0.39	-2.73 ± 0.41	-0.93 ± 0.39

Table 7.1—Continued

α (J2000)	δ (J2000)	$\log \hat{M}_{BL}/M_\odot$	$\log L_X/\hat{L}_{Edd}$[a]	$\log L_{UV}/\hat{L}_{Edd}$[b]	
09 55 42.1	+41 16 55.3	3.420	9.47 ± 0.39	-2.81 ± 0.40	-1.43 ± 0.39
09 55 48.1	+41 09 55.3	2.307	9.56 ± 0.38	-2.86 ± 0.39	-1.58 ± 0.38
09 56 40.4	+41 10 43.5	1.887	8.91 ± 0.32	-3.08 ± 0.33	-1.74 ± 0.32
09 58 20.5	+07 34 36.1	1.885	9.26 ± 0.31	-1.97 ± 0.33	-1.28 ± 0.31
09 58 22.6	+07 47 47.7	3.218	8.53 ± 0.39	-2.08 ± 0.40	-0.83 ± 0.39
09 58 36.6	+07 45 56.3	1.487	9.21 ± 0.45	-2.68 ± 0.45	-1.86 ± 0.45
10 02 05.4	+55 42 57.9	1.151	9.28 ± 0.45	-3.20 ± 0.45	-1.63 ± 0.45
10 03 52.8	+47 36 53.4	2.934	8.89 ± 0.39	-2.40 ± 0.40	-1.12 ± 0.39
10 09 09.2	+53 51 02.4	1.761	8.67 ± 0.29	-2.58 ± 0.29	-1.15 ± 0.29
10 09 21.9	+53 49 25.7	0.387	8.40 ± 0.32	-2.79 ± 0.32	-2.31 ± 0.36
10 09 26.7	+53 34 24.4	1.730	9.37 ± 0.38	-2.52 ± 0.38	-1.89 ± 0.38
10 09 46.0	+52 34 41.2	0.174	7.36 ± 0.45	-2.07 ± 0.45	-1.45 ± 0.48
10 10 37.5	+52 37 22.5	1.245	8.99 ± 0.45	-2.62 ± 0.45	-1.70 ± 0.45
10 13 14.9	-00 52 33.6	0.275	7.39 ± 0.45	-1.95 ± 0.45	-0.96 ± 0.48
10 16 44.3	+59 21 34.6	1.118	9.22 ± 0.45	-2.61 ± 0.45	-2.03 ± 0.45
10 17 30.6	+59 30 42.2	1.047	8.48 ± 0.45	-2.55 ± 0.45	-1.86 ± 0.45
10 17 37.7	+59 22 27.7	0.887	8.74 ± 0.45	-2.58 ± 0.45	-1.41 ± 0.45
10 17 50.2	+59 28 03.2	2.316	9.28 ± 0.38	-2.69 ± 0.38	-1.46 ± 0.38
10 17 57.5	+59 21 08.8	1.676	9.08 ± 0.29	-3.05 ± 0.29	-1.89 ± 0.29
10 18 11.8	+59 39 28.5	1.300	8.89 ± 0.45	-2.87 ± 0.45	-1.81 ± 0.45

Table 7.1—Continued

α (J2000)	δ (J2000)	$\log \hat{M}_{BL}/M_\odot$	$\log L_X/\hat{L}_{Edd}$[a]	$\log L_{UV}/\hat{L}_{Edd}$[b]	
10 19 02.0	+47 37 14.6	2.944	9.54 ± 0.38	-3.04 ± 0.39	-1.61 ± 0.38
10 19 39.2	+52 46 27.9	2.169	9.54 ± 0.29	-2.97 ± 0.29	-1.20 ± 0.29
10 19 43.7	+52 57 21.3	0.841	8.07 ± 0.32	-2.08 ± 0.32	-1.10 ± 0.32
10 20 23.3	+52 45 10.0	1.543	8.86 ± 0.31	-2.74 ± 0.31	-1.45 ± 0.32
10 23 50.9	+04 15 42.0	1.809	8.88 ± 0.29	-2.50 ± 0.29	-1.34 ± 0.29
10 30 31.6	+05 24 54.9	1.182	8.95 ± 0.45	-3.61 ± 0.46	-1.21 ± 0.45
10 32 27.9	+57 38 22.5	1.968	8.61 ± 0.29	-1.95 ± 0.29	-1.57 ± 0.29
10 32 36.2	+58 00 34.0	0.686	7.76 ± 0.32	-3.35 ± 0.32	-1.32 ± 0.32
10 36 23.8	-03 43 20.0	4.509	9.79 ± 0.38	-2.75 ± 0.40	-0.81 ± 0.38
10 37 03.1	-00 18 54.8	0.287	7.12 ± 0.53	-2.04 ± 0.53	-1.43 ± 0.56
10 38 08.7	+47 27 34.9	1.047	8.95 ± 0.45	-3.68 ± 0.46	-1.52 ± 0.45
10 42 30.7	+01 00 01.6	1.400	9.11 ± 0.45	-2.71 ± 0.46	-1.38 ± 0.45
10 44 54.9	+59 21 34.1	1.291	8.83 ± 0.45	-2.76 ± 0.45	-1.50 ± 0.45
10 47 23.5	+54 04 06.9	1.508	8.84 ± 0.29	-2.82 ± 0.29	-1.26 ± 0.29
10 49 21.5	+57 50 36.6	1.106	8.41 ± 0.45	-2.43 ± 0.46	-1.11 ± 0.45
10 50 15.6	+57 02 55.7	3.273	9.39 ± 0.38	-2.61 ± 0.39	-1.58 ± 0.38
10 50 50.1	+57 38 20.0	1.281	8.75 ± 0.45	-2.54 ± 0.45	-1.48 ± 0.45
10 52 39.6	+57 24 31.5	1.111	8.93 ± 0.45	-2.54 ± 0.45	-1.28 ± 0.45
10 52 39.6	+57 24 31.4	1.111	9.18 ± 0.45	-2.71 ± 0.45	-1.53 ± 0.45
10 53 16.8	+57 35 50.8	1.204	8.44 ± 0.45	-1.81 ± 0.49	-1.26 ± 0.45

Table 7.1—Continued

α (J2000)	δ (J2000)	$\log \hat{M}_{BL}/M_\odot$	$\log L_X/\hat{L}_{Edd}$[a]	$\log L_{UV}/\hat{L}_{Edd}$[b]	
10 54 04.1	+57 40 19.8	1.100	8.86 ± 0.45	-2.93 ± 0.46	-1.25 ± 0.45
10 54 22.6	+57 20 31.0	2.972	9.09 ± 0.38	-2.27 ± 0.38	-1.29 ± 0.39
10 55 18.1	+57 04 23.6	0.695	8.09 ± 0.32	-2.27 ± 0.32	-1.28 ± 0.32
10 56 44.5	+57 22 33.5	0.286	8.07 ± 0.46	-3.52 ± 0.46	-2.09 ± 0.49
10 57 56.4	+45 55 52.0	4.100	10.1 ± 0.38	-2.85 ± 0.39	-0.62 ± 0.38
11 02 05.7	+60 38 36.7	0.638	8.56 ± 0.32	-3.00 ± 0.32	-1.87 ± 0.32
11 02 50.1	+60 56 25.8	2.363	9.45 ± 0.38	-2.76 ± 0.38	-1.46 ± 0.38
11 03 07.8	+60 52 17.8	1.067	8.40 ± 0.45	-2.94 ± 0.45	-1.71 ± 0.45
11 03 08.8	+60 47 46.6	0.809	7.92 ± 0.45	-2.70 ± 0.45	-1.62 ± 0.45
11 03 49.8	+61 04 12.6	0.851	8.06 ± 0.45	-2.49 ± 0.45	-1.52 ± 0.45
11 09 05.3	+09 00 48.7	1.674	8.29 ± 0.30	-2.20 ± 0.31	-0.81 ± 0.30
11 11 32.1	+55 32 40.3	1.004	8.75 ± 0.45	-2.76 ± 0.45	-1.36 ± 0.45
11 14 52.8	+53 15 31.7	1.213	8.22 ± 0.45	-1.95 ± 0.50	-1.04 ± 0.45
11 15 20.7	+53 09 22.1	0.877	8.95 ± 0.45	-3.93 ± 0.45	-1.53 ± 0.45
11 29 43.9	-01 37 52.3	1.294	9.24 ± 0.45	-2.68 ± 0.45	-1.64 ± 0.45
11 29 51.2	-01 50 37.3	1.784	8.82 ± 0.29	-2.36 ± 0.30	-1.78 ± 0.29
11 32 19.7	+63 53 46.9	1.752	8.28 ± 0.29	-2.30 ± 0.29	-1.14 ± 0.29
11 36 21.2	+01 59 27.9	0.766	8.71 ± 0.32	-2.86 ± 0.32	-1.87 ± 0.32
11 36 31.9	+01 58 01.1	1.470	9.11 ± 0.45	-3.42 ± 0.47	-1.12 ± 0.45
11 36 33.1	+02 07 47.7	0.239	8.61 ± 0.45	-3.86 ± 0.45	-2.34 ± 0.48

Table 7.1—Continued

α (J2000)	δ (J2000)	$\log \hat{M}_{BL}/M_{\odot}$	$\log L_X/\hat{L}_{Edd}$[a]	$\log L_{UV}/\hat{L}_{Edd}$[b]	
11 36 34.9	+61 06 20.4	2.029	8.45 ± 0.29	-1.82 ± 0.29	-1.09 ± 0.29
11 38 29.1	+61 10 21.2	1.489	8.90 ± 0.47	-3.12 ± 0.47	-1.34 ± 0.47
12 01 14.4	-03 40 41.0	0.019	6.52 ± 0.45	-2.71 ± 0.45	-1.56 ± 0.48
12 02 26.8	-01 29 15.3	0.150	6.82 ± 0.54	-1.84 ± 0.54	-1.36 ± 0.57
12 03 29.6	-03 44 10.7	1.191	8.99 ± 0.45	-3.08 ± 0.45	-1.64 ± 0.45
12 03 53.9	-03 33 02.4	1.591	9.15 ± 0.30	-3.03 ± 0.30	-1.44 ± 0.30
12 04 15.2	-03 34 29.3	3.814	9.59 ± 0.38	-2.70 ± 0.38	-1.11 ± 0.38
12 04 31.4	-03 31 15.6	0.966	8.93 ± 0.45	-2.85 ± 0.45	-1.73 ± 0.45
12 04 36.6	+01 50 25.6	1.927	9.38 ± 0.29	-3.00 ± 0.29	-1.56 ± 0.29
12 04 50.6	+02 21 18.5	2.884	8.62 ± 0.38	-1.90 ± 0.38	-1.03 ± 0.38
12 04 52.9	-03 26 26.0	0.296	8.26 ± 0.45	-2.68 ± 0.45	-1.95 ± 0.48
12 05 23.7	-03 36 18.2	1.494	9.38 ± 0.45	-3.24 ± 0.45	-1.93 ± 0.45
12 05 43.0	+01 48 27.2	1.590	9.39 ± 0.29	-2.50 ± 0.29	-1.83 ± 0.29
12 06 30.4	+01 58 08.4	1.551	8.75 ± 0.45	-2.74 ± 0.45	-1.86 ± 0.45
12 08 29.6	+00 16 42.7	1.063	8.94 ± 0.45	-2.91 ± 0.45	-1.84 ± 0.45
12 13 43.0	+02 52 48.9	0.641	7.92 ± 0.32	-2.77 ± 0.32	-1.39 ± 0.32
12 14 15.2	+00 55 11.5	0.395	7.80 ± 0.32	-2.10 ± 0.32	-1.23 ± 0.36
12 15 40.5	-00 34 33.8	0.757	8.67 ± 0.33	-2.71 ± 0.34	-2.10 ± 0.33
12 18 36.1	+05 46 28.1	0.795	9.05 ± 0.32	-2.81 ± 0.32	-2.29 ± 0.32
12 18 36.1	+05 46 28.1	0.795	9.14 ± 0.32	-3.39 ± 0.32	-2.37 ± 0.32

Table 7.1—Continued

α (J2000)	δ (J2000)	$\log \hat{M}_{BL}/M_\odot$	$\log L_X/\hat{L}_{Edd}$[a]	$\log L_{UV}/\hat{L}_{Edd}$[b]	
12 19 25.8	+04 10 12.9	0.912	8.67 ± 0.45	-2.37 ± 0.45	-1.69 ± 0.45
12 19 42.5	+03 45 30.3	0.794	8.60 ± 0.45	-2.72 ± 0.45	-1.63 ± 0.45
12 19 48.9	+05 45 31.7	0.113	6.83 ± 0.45	-2.76 ± 0.45	-1.40 ± 0.48
12 19 55.3	+03 41 04.0	1.492	8.47 ± 0.45	-1.97 ± 0.45	-1.16 ± 0.45
12 20 04.4	-00 25 39.1	0.421	7.36 ± 0.32	-1.29 ± 0.32	-1.12 ± 0.32
12 23 07.5	+10 34 48.2	2.747	9.10 ± 0.42	-2.26 ± 0.42	-0.86 ± 0.42
12 26 52.0	-00 11 59.6	1.175	8.74 ± 0.45	-2.23 ± 0.45	-1.06 ± 0.45
12 26 56.5	+01 31 24.4	0.731	7.89 ± 0.47	-2.30 ± 0.47	-2.00 ± 0.47
12 27 23.7	+01 48 06.0	2.880	9.87 ± 0.38	-3.33 ± 0.38	-2.08 ± 0.38
12 27 24.6	+01 39 30.5	0.334	8.44 ± 0.46	-3.82 ± 0.46	-2.20 ± 0.50
12 27 43.0	+01 34 38.4	1.279	9.24 ± 0.45	-3.21 ± 0.45	-1.83 ± 0.45
12 28 05.6	+01 41 19.2	1.777	9.58 ± 0.29	-3.12 ± 0.29	-1.38 ± 0.29
12 28 18.0	+44 13 02.0	0.662	8.21 ± 0.32	-3.81 ± 0.32	-1.04 ± 0.32
12 28 53.7	+44 11 52.9	1.276	8.97 ± 0.45	-2.87 ± 0.45	-1.47 ± 0.45
12 30 05.8	+03 02 04.2	1.604	9.35 ± 0.29	-2.98 ± 0.30	-1.74 ± 0.29
12 30 25.9	+03 05 35.4	1.055	8.61 ± 0.45	-2.47 ± 0.45	-1.65 ± 0.45
12 30 27.4	+03 06 27.5	0.628	8.69 ± 0.32	-2.60 ± 0.32	-1.82 ± 0.32
12 30 39.9	+03 08 57.3	1.843	9.12 ± 0.30	-2.84 ± 0.33	-1.60 ± 0.30
12 30 54.7	+03 05 37.2	0.427	7.94 ± 0.32	-2.92 ± 0.33	-1.78 ± 0.32
12 35 14.9	+00 47 40.7	1.874	9.13 ± 0.29	-2.83 ± 0.29	-1.47 ± 0.29

Table 7.1—Continued

α (J2000)	δ (J2000)	$\log \hat{M}_{BL}/M_\odot$	$\log L_X/\hat{L}_{Edd}$[a]	$\log L_{UV}/\hat{L}_{Edd}$[b]	
12 35 52.7	+00 50 29.8	2.240	9.00 ± 0.39	-2.61 ± 0.39	-1.18 ± 0.39
12 36 22.9	+62 15 26.6	2.587	8.43 ± 0.41	-2.60 ± 0.41	-1.01 ± 0.41
12 36 59.9	+00 42 12.2	2.368	9.38 ± 0.38	-2.63 ± 0.38	-1.40 ± 0.38
12 37 16.0	+62 03 23.4	2.068	9.04 ± 0.30	-2.94 ± 0.30	-1.49 ± 0.30
12 37 29.8	+00 54 33.5	3.282	9.17 ± 0.39	-2.20 ± 0.39	-0.80 ± 0.39
12 42 14.1	+02 42 24.2	1.603	9.01 ± 0.29	-2.62 ± 0.29	-1.62 ± 0.29
12 42 19.2	+02 31 18.3	0.479	8.43 ± 0.32	-2.75 ± 0.32	-2.07 ± 0.32
12 42 20.1	+02 32 57.6	2.224	9.16 ± 0.29	-2.59 ± 0.29	-0.92 ± 0.29
12 42 44.6	+02 29 54.4	0.937	9.12 ± 0.45	-3.10 ± 0.45	-1.38 ± 0.45
12 42 55.3	+02 49 57.0	1.458	9.02 ± 0.45	-2.63 ± 0.45	-1.55 ± 0.45
12 42 55.3	+02 49 57.0	1.458	8.99 ± 0.45	-2.63 ± 0.46	-1.52 ± 0.45
12 43 20.0	+02 52 56.2	0.086	6.71 ± 0.52	-1.97 ± 0.52	-1.41 ± 0.55
12 45 24.6	-00 09 38.0	2.084	9.59 ± 0.29	-3.02 ± 0.29	-1.27 ± 0.29
12 45 41.0	-00 27 44.8	1.693	9.56 ± 0.29	-2.50 ± 0.29	-1.71 ± 0.29
12 45 41.0	-00 27 44.9	1.693	9.54 ± 0.29	-2.55 ± 0.30	-1.69 ± 0.29
12 45 55.1	-00 37 35.3	1.042	8.98 ± 0.45	-2.77 ± 0.45	-1.37 ± 0.45
12 55 35.1	+56 52 39.6	1.803	9.17 ± 0.29	-2.82 ± 0.29	-1.54 ± 0.29
12 55 36.2	+56 50 00.1	1.373	8.52 ± 0.45	-3.63 ± 0.45	-1.41 ± 0.45
12 58 49.8	-01 43 03.3	0.967	9.08 ± 0.45	-2.39 ± 0.45	-1.17 ± 0.45
12 59 43.6	+01 02 55.0	0.394	7.75 ± 0.33	-2.90 ± 0.33	-1.20 ± 0.38

Table 7.1—Continued

α (J2000)	δ (J2000)	$\log \hat{M}_{BL}/M_\odot$	$\log L_X/\hat{L}_{Edd}$[a]	$\log L_{UV}/\hat{L}_{Edd}$[b]	
13 10 55.6	-01 27 24.5	1.002	8.71 ± 0.45	-3.09 ± 0.45	-1.47 ± 0.45
13 11 08.5	+00 31 51.7	0.429	7.65 ± 0.32	-2.66 ± 0.34	-0.95 ± 0.32
13 11 31.0	-01 03 32.4	1.303	8.79 ± 0.45	-2.96 ± 0.45	-1.28 ± 0.45
13 12 08.7	-01 07 10.2	0.804	8.94 ± 0.32	-2.88 ± 0.32	-1.99 ± 0.32
13 17 43.2	+35 31 31.1	4.360	9.26 ± 0.42	-2.12 ± 0.42	-0.84 ± 0.42
13 21 10.8	+00 38 22.0	4.716	9.19 ± 0.39	-2.63 ± 0.41	-0.71 ± 0.40
13 29 48.5	-02 00 28.4	1.356	9.20 ± 0.45	-2.79 ± 0.45	-1.62 ± 0.45
13 30 27.5	-01 42 03.2	1.268	8.68 ± 0.45	-2.05 ± 0.45	-1.21 ± 0.45
13 31 02.5	-01 50 14.5	0.561	8.32 ± 0.32	-2.56 ± 0.32	-1.79 ± 0.32
13 31 11.3	-02 06 12.3	2.026	9.14 ± 0.30	-2.67 ± 0.30	-1.44 ± 0.30
13 31 35.2	-01 44 30.3	1.461	8.95 ± 0.46	-2.48 ± 0.46	-1.28 ± 0.46
13 31 39.3	-02 06 02.2	2.043	9.72 ± 0.29	-3.15 ± 0.29	-1.79 ± 0.29
13 31 41.0	-01 52 12.4	0.145	6.38 ± 0.45	-1.46 ± 0.45	-0.67 ± 0.49
13 42 33.7	-00 11 48.1	0.515	8.63 ± 0.32	-3.47 ± 0.32	-2.12 ± 0.32
13 43 25.8	-00 16 12.2	1.513	8.94 ± 0.45	-3.09 ± 0.45	-1.41 ± 0.45
13 43 51.1	+00 04 34.8	0.073	6.86 ± 0.50	-3.72 ± 0.50	-1.47 ± 0.53
13 44 25.9	-00 00 56.2	1.095	8.76 ± 0.46	-3.34 ± 0.47	-1.41 ± 0.46
13 44 26.0	-00 00 56.1	1.095	8.76 ± 0.45	-2.71 ± 0.45	-1.41 ± 0.45
13 44 59.5	-00 15 59.5	0.244	7.56 ± 0.45	-2.20 ± 0.45	-1.23 ± 0.48
13 58 54.5	+62 39 13.3	1.228	8.58 ± 0.46	-1.95 ± 0.46	-1.06 ± 0.46

Table 7.1—Continued

α (J2000)	δ (J2000)	$\log \hat{M}_{BL}/M_\odot$	$\log L_X/\hat{L}_{Edd}$[a]	$\log L_{UV}/\hat{L}_{Edd}$[b]	
13 59 32.3	+62 35 46.4	0.522	8.46 ± 0.32	-2.84 ± 0.32	-2.05 ± 0.32
14 00 37.1	+62 21 32.9	0.075	7.58 ± 0.45	-3.43 ± 0.45	-2.17 ± 0.48
14 00 41.1	+62 25 16.2	1.877	9.02 ± 0.30	-2.91 ± 0.30	-1.58 ± 0.30
14 00 45.0	+62 18 14.6	0.542	8.39 ± 0.32	-2.70 ± 0.32	-1.90 ± 0.32
14 11 04.1	+52 17 55.6	2.882	8.93 ± 0.42	-2.30 ± 0.42	-0.88 ± 0.42
14 11 04.9	+52 05 16.8	1.083	8.53 ± 0.46	-2.94 ± 0.50	-1.34 ± 0.46
14 11 51.5	+59 23 49.1	1.179	8.51 ± 0.45	-2.77 ± 0.45	-0.98 ± 0.45
14 12 18.5	+58 59 18.0	2.026	9.06 ± 0.45	-2.32 ± 0.45	-1.03 ± 0.45
14 12 39.7	+59 21 31.4	0.980	8.76 ± 0.45	-2.94 ± 0.45	-1.61 ± 0.45
14 12 41.5	+58 52 56.5	1.608	9.04 ± 0.30	-3.10 ± 0.30	-1.48 ± 0.30
14 13 15.3	+00 00 32.0	4.078	8.52 ± 0.38	-2.05 ± 0.40	-0.17 ± 0.38
14 17 00.8	+44 56 06.4	0.113	7.87 ± 0.46	-2.77 ± 0.46	-1.57 ± 0.49
14 19 51.9	+47 09 01.4	2.288	9.74 ± 0.38	-2.67 ± 0.38	-1.03 ± 0.38
14 24 36.0	+42 10 30.5	2.217	9.44 ± 0.30	-2.35 ± 0.30	-1.00 ± 0.30
14 32 44.4	-00 59 15.2	1.026	9.45 ± 0.45	-2.38 ± 0.45	-1.69 ± 0.45
14 33 35.3	+02 27 18.3	2.072	8.80 ± 0.38	-2.26 ± 0.40	-1.37 ± 0.38
14 34 07.5	+02 27 04.6	1.710	8.79 ± 0.29	-2.40 ± 0.30	-1.30 ± 0.29
14 38 42.0	+03 41 10.4	1.737	9.09 ± 0.29	-2.58 ± 0.29	-1.05 ± 0.29
14 38 59.1	+03 35 47.5	0.733	8.58 ± 0.32	-2.77 ± 0.32	-1.51 ± 0.32
14 44 28.7	-01 23 44.0	4.179	9.83 ± 0.38	-3.37 ± 0.40	-1.31 ± 0.38

Table 7.1—Continued

α (J2000)	δ (J2000)	$\log \hat{M}_{BL}/M_\odot$	$\log L_X/\hat{L}_{Edd}$[a]	$\log L_{UV}/\hat{L}_{Edd}$[b]	
14 45 54.8	+01 29 03.3	1.845	8.94 ± 0.29	-2.55 ± 0.30	-1.44 ± 0.29
14 48 53.4	+47 38 21.3	2.894	9.69 ± 0.38	-3.23 ± 0.40	-1.51 ± 0.38
14 48 56.7	+00 15 10.3	0.832	8.61 ± 0.45	-2.73 ± 0.46	-1.55 ± 0.45
14 49 13.5	+00 24 06.9	0.440	7.68 ± 0.32	-2.22 ± 0.32	-1.24 ± 0.32
14 52 15.6	+43 04 48.7	0.296	7.67 ± 0.51	-2.81 ± 0.51	-1.70 ± 0.54
14 52 40.9	+43 08 14.4	1.704	9.30 ± 0.29	-2.71 ± 0.30	-1.79 ± 0.29
14 53 00.6	+33 52 06.0	4.191	8.74 ± 0.38	-2.09 ± 0.39	-0.63 ± 0.38
15 05 43.9	+55 49 36.2	0.708	8.78 ± 0.32	-2.67 ± 0.32	-1.62 ± 0.32
15 07 30.6	+55 37 10.9	4.499	9.06 ± 0.38	-1.69 ± 0.38	-0.65 ± 0.45
15 11 26.5	+56 59 34.8	1.031	9.07 ± 0.45	-2.99 ± 0.45	-1.38 ± 0.45
15 14 53.9	+56 10 32.0	1.286	8.69 ± 0.45	-2.77 ± 0.45	-1.47 ± 0.45
15 15 04.9	+55 21 07.3	1.844	8.44 ± 0.29	-2.40 ± 0.34	-1.26 ± 0.29
15 15 10.1	+56 28 34.8	0.723	8.00 ± 0.35	-2.13 ± 0.35	-1.63 ± 0.35
15 15 30.7	+56 01 32.5	0.571	7.57 ± 0.32	-1.98 ± 0.32	-1.07 ± 0.32
15 39 47.6	+43 13 41.6	0.347	7.68 ± 0.45	-2.16 ± 0.45	-1.42 ± 0.48
15 43 16.4	+54 05 26.1	0.245	7.59 ± 0.45	-2.48 ± 0.45	-1.55 ± 0.48
15 45 30.2	+48 46 09.1	0.399	8.13 ± 0.33	-2.18 ± 0.33	-0.93 ± 0.37
16 05 17.8	-01 09 55.5	1.572	8.91 ± 0.30	-2.17 ± 0.30	-1.17 ± 0.30
16 18 34.0	+34 56 25.6	1.922	9.28 ± 0.29	-2.94 ± 0.31	-1.29 ± 0.29
16 32 57.0	+41 58 29.0	1.469	9.05 ± 0.45	-3.23 ± 0.45	-1.53 ± 0.45

Table 7.1—Continued

α (J2000)	δ (J2000)	$\log \hat{M}_{BL}/M_\odot$	$\log L_X/\hat{L}_{Edd}$[a]	$\log L_{UV}/\hat{L}_{Edd}$[b]	
16 33 37.5	+42 06 52.8	0.402	7.92 ± 0.32	-2.13 ± 0.32	-1.89 ± 0.32
16 40 25.0	+46 44 49.1	0.537	8.40 ± 0.32	-2.96 ± 0.32	-1.45 ± 0.32
16 41 10.6	+46 49 11.9	0.695	8.44 ± 0.32	-2.75 ± 0.32	-1.76 ± 0.32
16 41 54.2	+40 00 33.1	1.002	9.13 ± 0.45	-3.22 ± 0.45	-1.52 ± 0.45
16 56 27.3	+62 32 26.7	0.184	8.13 ± 0.45	-4.08 ± 0.45	-2.03 ± 0.48
16 57 13.2	+35 24 39.4	2.328	9.00 ± 0.38	-2.19 ± 0.38	-1.27 ± 0.38
16 57 13.2	+35 24 39.4	2.328	8.98 ± 0.38	-2.69 ± 0.40	-1.25 ± 0.38
16 57 39.8	+35 04 26.9	1.953	9.51 ± 0.29	-3.30 ± 0.29	-1.35 ± 0.29
16 59 03.6	+35 24 23.6	1.218	9.06 ± 0.45	-3.15 ± 0.45	-2.22 ± 0.45
17 01 00.6	+64 12 09.1	2.735	10.2 ± 0.38	-2.87 ± 0.38	-0.98 ± 0.38
17 01 00.6	+64 12 09.0	2.735	10.2 ± 0.38	-3.21 ± 0.38	-0.96 ± 0.38
17 02 24.5	+34 05 39.0	2.038	9.04 ± 0.30	-3.18 ± 0.30	-1.14 ± 0.30
17 03 55.8	+60 45 11.8	0.284	7.21 ± 0.45	-2.08 ± 0.45	-1.25 ± 0.49
17 03 55.8	+60 45 11.7	0.284	7.02 ± 0.48	-1.87 ± 0.48	-1.39 ± 0.51
17 04 06.1	+60 47 53.9	1.361	8.98 ± 0.45	-2.29 ± 0.45	-1.30 ± 0.45
17 08 17.9	+61 54 48.6	1.414	9.41 ± 0.45	-2.41 ± 0.45	-1.49 ± 0.45
17 12 24.5	+33 17 21.1	1.622	9.07 ± 0.29	-3.06 ± 0.29	-1.95 ± 0.29
17 19 27.3	+27 32 46.8	1.446	9.20 ± 0.45	-2.51 ± 0.45	-1.59 ± 0.45
17 20 26.5	+26 38 16.0	1.141	8.50 ± 0.45	-2.51 ± 0.47	-1.28 ± 0.45
17 35 51.9	+53 55 15.7	0.955	9.03 ± 0.45	-2.49 ± 0.45	-1.52 ± 0.45

7.5.2 Eddington Ratio Estimates

A constant bolometric correction has been used in most previous studies involving the AGN Eddington ratio. However, recent work by Vasudevan & Fabian (2007) has suggested that bolometric corrections show a large spread with no obvious dependence on luminosity. Furthermore, these authors found evidence that the bolometric correction depends on the Eddington ratio. This implies that the error in the bolometric correction is correlated with Eddington ratio, which therefore implies that the error in the estimated Eddington ratio is correlated with the actual Eddington ratio. An Eddington ratio-dependent error in the bolometric correction my cause problems when using the estimated Eddington ratios to infer correlations.

Further difficulties with a constant bolometric correction are illustrated with Figure 7.3. In 7.3 we plot α_{ox} as a function of $L_{UV}/\hat{L}_{Edd}$ and $L_X/\hat{L}_{Edd}$, where we estimate the Eddington luminosity from the broad line mass estimates as $\hat{L}_{Edd} = 1.3 \times 10^{38} \hat{M}_{BL}/M_\odot$ erg s^{-1}. As with $\hat{M}_{BL}$, we use the notation $\hat{L}_{Edd}$ to emphasize that $\hat{L}_{Edd}$ is an estimate of the true L_{Edd} based on the broad line mass estimates, and therefore $\hat{L}_{Edd} \propto L^\gamma FWHM^2$. Constant bolometric corrections are often applied to either the optical/UV or X-ray luminosity. If a constant bolometric correction was valid for both L_{UV} and L_X, then we would expect that $L_{UV}/\hat{L}_{Edd} \propto L_X/\hat{L}_{Edd} \propto L_{bol}/L_{Edd}$. However, while a correlation between α_{ox} and both $L_{UV}/\hat{L}_{Edd}$ and $L_X/\hat{L}_{Edd}$ is apparent, they are of opposite sign. Because the correlations are of opposite sign, it cannot be true that both $L_{UV}/\hat{L}_{Edd}$ and $L_X/\hat{L}_{Edd}$ are proportional to the Eddington ratio, L_{bol}/L_{Edd}.

Because of the current significant uncertainty regarding RQQ bolometric corrections, we take the conservative approach and merely compare α_{ox} and Γ_X with L_{UV}/L_{Edd} and L_X/L_{Edd}. The estimated values of L_{UV}/L_{Edd} and L_X/L_{Edd}

Table 7.1—Continued

α (J2000)	δ (J2000)	$\log \hat{M}_{BL}/M_\odot$	$\log L_X/\hat{L}_{Edd}$[a]	$\log L_{UV}/\hat{L}_{Edd}$[b]	
17 37 16.6	+58 28 39.5	1.775	9.06 ± 0.30	-2.49 ± 0.31	-1.37 ± 0.30
17 38 36.2	+58 37 48.6	1.279	8.81 ± 0.45	-3.26 ± 0.47	-0.95 ± 0.45
21 56 21.8	+01 03 30.6	1.353	9.11 ± 0.45	-3.14 ± 0.45	-1.34 ± 0.45
21 56 47.1	+01 05 02.7	1.036	8.51 ± 0.45	-2.81 ± 0.45	-1.32 ± 0.45
22 15 27.1	-16 11 33.0	3.990	8.90 ± 0.38	-1.96 ± 0.40	-0.20 ± 0.38
22 38 19.8	-09 21 06.0	3.259	10.3 ± 0.38	-2.91 ± 0.38	-1.72 ± 0.38
22 38 54.7	-09 37 36.2	1.472	9.12 ± 0.45	-2.06 ± 0.45	-1.66 ± 0.45
22 39 17.3	-09 33 40.9	1.817	9.34 ± 0.29	-2.15 ± 0.29	-1.62 ± 0.29
22 49 03.3	-08 08 41.7	0.457	7.97 ± 0.32	-2.22 ± 0.32	-1.98 ± 0.32
23 13 04.3	+13 53 03.5	3.026	9.70 ± 0.38	-2.98 ± 0.38	-1.73 ± 0.38
23 14 07.9	+14 07 36.6	3.772	9.64 ± 0.38	-2.82 ± 0.38	-1.55 ± 0.39
23 37 18.1	+00 25 50.7	2.053	9.03 ± 0.29	-2.64 ± 0.31	-1.27 ± 0.29
23 37 22.0	+00 22 38.9	1.376	8.72 ± 0.45	-2.99 ± 0.47	-1.35 ± 0.45
23 37 39.1	+00 26 56.2	1.703	8.93 ± 0.31	-2.14 ± 0.32	-1.20 ± 0.31
23 48 40.1	+01 07 53.5	0.718	8.57 ± 0.32	-2.67 ± 0.32	-1.49 ± 0.32
23 54 19.7	-10 07 18.5	1.745	9.14 ± 0.29	-2.61 ± 0.29	-1.17 ± 0.29
23 55 00.9	-10 31 01.3	0.284	7.72 ± 0.48	-2.39 ± 0.48	-1.77 ± 0.51
23 57 18.3	+00 43 50.0	4.362	8.86 ± 0.40	-2.25 ± 0.41	-0.43 ± 0.40

[a]Logarithm of the ratio of νL_ν [2 keV] to $\hat{L}_{Edd}$, where $\hat{L}_{Edd}$ is calculated from the broad emission line estimate of M_{BH}, $\hat{M}_{BL}$.

[b]Logarithm of the ratio of νL_ν [2500Å] to $\hat{L}_{Edd}$, where $\hat{L}_{Edd}$ is calculated from the broad emission line estimate of M_{BH}, $\hat{M}_{BL}$.

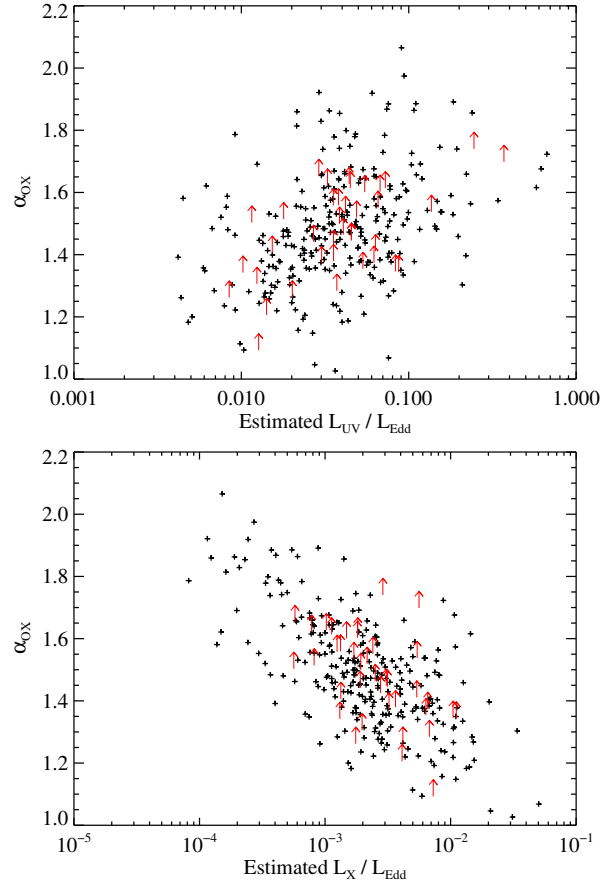Figure 7.3 The distribution of α_{ox} as a function of L_{UV}/L_{Edd} (top) and L_X/L_{Edd} (bottom). The opposite dependence of α_{ox} on L_{UV}/L_{Edd} and L_X/L_{Edd} suggests that at least one of these quantities is not proportional to L_{bol}/L_{Edd}. As such, we do not employ bolometric corrections in this work, and instead compare directly with L_{UV}/L_{Edd} and L_X/L_{Edd}.

for the sources in our sample are reported in Table 7.1. We can write $L_{UV}/L_{Edd} \propto$ $f_{UV} L_{bol}/L_{Edd}$ and $L_X/L_{Edd} \propto f_X L_{bol}/L_{Edd}$, where f_{UV} and f_X are the inverses of the bolometric corrections for L_{UV} and L_X, respectively. The quantities f_{UV} and f_X are proportional to the fraction of the bolometric luminosity emitted at 2500Å and 2 keV. Then, correlations between either α_{ox} or Γ_X and L_{UV}/L_{Edd} will result if α_{ox} or Γ_X is correlated with f_{UV}, the Eddington ratio, or both, and likewise for f_X.

While we do not use an estimate of the Eddington ratio in our analysis, it is helpful to estimate the distribution of Eddington ratios probed by our sample. We assume the bolometric correction described in Hopkins et al. (2007) for the $z < 1.5$ sources, and constant bolometric described in Vestergaard (2004) of $L_{bol} = 4.62\lambda L_\lambda (1350\text{Å})$ at $z > 1.5$. In Figure 7.2 we also show the distribution of estimated Eddington ratios as a function of z. Because the distribution of estimated L_{bol}/L_{Edd} is the true distribution of L_{bol}/L_{Edd} broadened by the distribution of errors in the estimates, our sample likely probes a smaller range in Eddington ratio than that inferred from Figure 7.2. Therefore, at most our sample probes RQQs with Eddington ratios $0.03 \lesssim L_{bol}/L_{Edd} \lesssim 2$.

7.6   DEPENDENCE OF α_{ox} ON $M_{BH}, L_{UV}/L_{Edd}$, and L_X/L_{Edd}

We used our sample of 318 sources with estimates of M_{BH} to investigate the dependence of α_{ox} at a given black hole mass, L_{UV}/L_{Edd}, and L_X/L_{Edd}. We use linear regression analysis in order to understand how α_{ox} varies with respect to these parameters. We use the method of Kelly (2007) to estimate the regression parameters. The method of Kelly (2007) accounts for measurement errors, non-detections, and intrinsic scatter. In addition, Kelly (2007) adopts a Bayesian approach, computing the posterior probability distribution of the parameters, given

the observed data. Thus the uncertainties on the regression coefficients have a straight-forward interpretion, and do not rely on large-sample approximations. Many other methods, such as traditional maximum-likelihood, assume that the errors in the regression parameters follow a Gaussian distribution, which is valid as the sample size approaches infinity. However, this assumption is not necessarily valid for our finite sample size, especially in the presence of censoring (i.e., presence of upper/lower limits) and significant measurement error. The method of Kelly (2007) directly estimates the probability distribution of the regression parameters, and is therefore preferred.

We assess the simple 2-dimensional correlations between $\alpha_{\rm ox}$ and $M_{BH}, L_{UV}/L_{Edd}$, and L_X/L_{Edd}, and compare with the $\alpha_{\rm ox}$–L_{UV} correlation. The results from the regressions are

$$\alpha_{\rm ox} = -3.91^{+1.04}_{-1.01} + (0.12^{+0.02}_{-0.02}) \log L_{UV}, \quad \sigma_{\alpha_{\rm ox}} = 0.14^{+0.02}_{-0.01}, \quad \rho = 0.57^{+0.09}_{-0.10} \quad (7.7)$$

$$\alpha_{\rm ox} = 0.05^{+0.39}_{-0.42} + (0.17^{+0.05}_{-0.04}) \log M_{BH}, \quad \sigma_{\alpha_{\rm ox}} = 0.14^{+0.02}_{-0.02}, \quad \rho = 0.53^{+0.12}_{-0.13} \quad (7.8)$$

$$\alpha_{\rm ox} = 2.90^{+0.71}_{-0.43} + (0.99^{+0.50}_{-0.31}) \log L_{UV}/L_{Edd}, \quad \sigma_{\alpha_{\rm ox}} = 0.05^{+0.05}_{-0.03}, \quad \rho = 0.95^{+0.04}_{-0.16} (7.9)$$

$$\alpha_{\rm ox} = -0.03^{+0.25}_{-0.34} - (0.57^{+0.09}_{-0.12}) \log L_X/L_{Edd}, \quad \sigma_{\alpha_{\rm ox}} = 0.03^{+0.04}_{-0.03},$$

$$\rho = -0.98^{+0.06}_{-0.02} \quad (7.10)$$

where $\sigma_{\alpha_{\rm ox}}$ is the intrinsic dispersion in $\alpha_{\rm ox}$ at a given $L_{UV}, M_{BH}, L_{UV}/L_{Edd}$, or L_X/L_{Edd}, ρ is the linear correlation coefficient for $\alpha_{\rm ox}$ and the respective independent variables, and the errors are quoted at the 95% (2σ) level. All four relationships are significant, with RQQs becoming more X-ray quiet as L_{UV}, M_{BH}, or L_{UV}/L_{Edd} increases, and more X-ray loud as L_X/L_{Edd} increases. Because we have attempted to correct for the intrinsic statistical scatter in the broad line mass estimates, Equations (7.7)–(7.10) refer to the intrinsic relationships involving M_{BH}, barring any systematic errors in reverberation mapping, and are not simply cor-
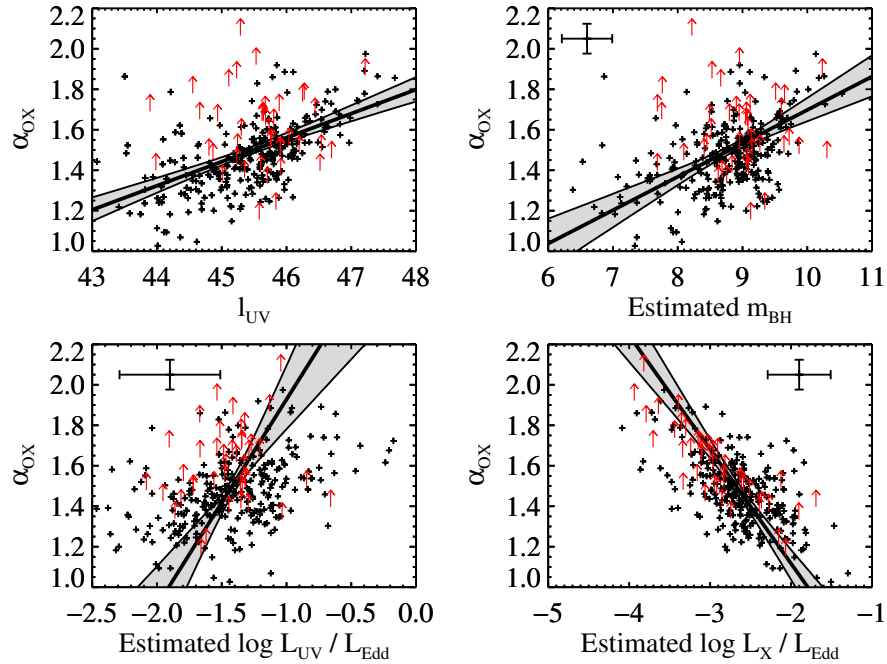
Figure 7.4 Ratio of optical/UV to X-ray flux as a function of $L_{UV}, M_{BH}, L_{UV}/L_{Edd}$, and L_X/L_{Edd}. The solid lines denote the best fit, and the shaded regions contain 95% (2σ) of the probability on the regression line. The data points with error bars in the plots of α_{ox} as a function of $M_{BH}, L_{UV}/L_{Edd}$, and L_X/L_{Edd} are fictitious and illustrate the typical errors in each direction.

relations between α_{ox} and the broad line mass estimates. The estimated distributions of α_{ox} as a function of $L_{UV}, M_{BH}, L_{UV}/L_{Edd}$, and L_X/L_{Edd} are shown in Figure 7.4, along with the regression results.

The intrinsic dispersion in α_{ox} quantifies the magnitude of scatter in α_{ox} at a given $L_{UV}, M_{BH}, L_{UV}/L_{Edd}$, or L_X/L_{Edd}. Because we have attempted to account for contribution to the scatter in α_{ox} resulting from measurement error and variability, $\sigma_{\alpha_{ox}}$ represents the dispersion in the real physical scatter in α_{ox} over the population of RQQs. This 'residual' scatter represents the amount of variation in

α_{ox} that is not 'explained' by variations in $L_{UV}, M_{BH}, L_{UV}/L_{Edd}$, or L_X/L_{Edd}, respectively. This intrinsic scatter in α_{ox} may be due to variations in accretion rate, viscosity, column density, and other quantities not included in our regression.

In § 7.3 we estimate that there are 25 ± 5 BAL quasars in our sample at $z < 1.5$. Because these objects have the appearance of being X-ray weak, and because redshift is artificially correlated with luminosity and M_{BH} in a flux limited sample, we expect that the presence of unidentified BALs at $z < 1.5$ will produce an excess of X-ray weak objects at low L_{UV} and M_{BH}, thus flattening the inferred slopes. Inspection of the plot of α_{ox} and z in Figure 7.1 suggests an excess of X-ray weak objects at $z < 1.5$ and $\alpha_{\mathrm{ox}} \gtrsim 1.8$, implying these objects are BAL quasars. We removed these 10 objects and refit the regressions. Omission of these objects resulted in a steepening of the slopes for the L_{UV} and M_{BH} regression, and a flattening of the slope for the L_{UV}/L_{Edd} regression. In addition, the intrinsic dispersion in α_{ox} decreased for the L_{UV} and M_{BH} regressions, while it remained the same for the L_{UV}/L_{Edd} regression. These changes were small ($\sim 10\%$) and have no effect on our conclusions. There was no difference in the results for the L_X/L_{Edd} regression.

Once can use Equation (7.2) to express the regression results (Eq. [7.7]–[7.10]) in the alternate form

$$\frac{\nu L_\nu(2500\text{Å})}{\nu L_\nu(2\text{ keV})} = 1.17^{+0.08}_{-0.07} \times 10^4 \left(\frac{\nu L_\nu(2500\text{Å})}{10^{46}\text{ergs}^{-1}}\right)^{0.31\pm0.03}, \tag{7.11}$$

$$\frac{\nu L_\nu(2500\text{Å})}{\nu L_\nu(2\text{ keV})} = 9.81^{+0.65}_{-0.63} \times 10^3 \left(\frac{M_{BH}}{10^9 M_\odot}\right)^{0.43\pm0.06}, \tag{7.12}$$

$$\frac{\nu L_\nu(2500\text{Å})}{\nu L_\nu(2\text{ keV})} = 3.51^{+15.6}_{-2.58} \times 10^7 \left(\frac{\nu L_\nu(2500\text{Å})}{L_{Edd}}\right)^{2.57\pm0.45}, \tag{7.13}$$

$$\frac{\nu L_\nu(2500\text{Å})}{\nu L_\nu(2\text{ keV})} = 0.85^{+1.02}_{-0.52} \left(\frac{\nu L_\nu(2\text{ keV})}{L_{Edd}}\right)^{-1.48\pm0.14}, \tag{7.14}$$

where the intrinsic dispersion in $\log L_{UV}/L_X$ at a given $L_{UV}, M_{BH}, L_{UV}/L_{Edd}$, and

L_X/L_{Edd} is $\sim 0.356, 0.375, 0.133$, and 0.089 dex, respectively. In contrast to Equations (7.7)–(7.10), we quote the 68% (1σ) uncertainties on the constants of proportionality, and the posterior standard deviations on the exponents. Equations (7.11)–(7.14) may be more physically interpretable and allow easier comparison with models.

7.7 NONMONOTONIC DEPENDENCE OF Γ_X, L_{UV}/L_{Edd}, and L_X/L_{bol}

Recent work has suggested a correlation between quasar X-ray spectral slope, $\alpha_X = \Gamma_X - 1$, $f_\nu \propto \nu^{-\alpha_X}$, and quasar Eddington ratio as inferred from broad line mass estimates based on the Hβ emission line (e.g., Porquet et al., 2004; Piconcelli et al., 2005; Shemmer et al., 2006). The Kelly et al. (2007) sample contains measurements of Γ_X for 157 sources, and we were able to estimate black hole masses for 153 of them. In this section we use these 153 RQQs to investigate the dependence of Γ_X on M_{BH}, L_{UV}/L_{Edd}, and L_X/L_{Edd}.

7.7.1 Regression Analysis

The distributions of Γ_X as a function of estimated black hole mass, L_{UV}/L_{Edd}, and L_X/L_{Edd} are shown in Figure 7.5 for the entire sample, and in Figure 7.6 seperately for each emission line. While there does not appear to be a monotonic trend between Γ_X and M_{BH}, L_{UV}/L_{Edd}, or L_X/L_{Edd} when using the entire sample, there is evidence for a trend between Γ_X and these quantities when using the Hβ line, and an opposite trend between Γ_X and these quantities when using the C IV line.

We performed a linear regression of Γ_X on $\log M_{BH}$, $\log L_{UV}/L_{Edd}$, and $\log L_X/L_{Edd}$ seperately for each emission line. As before, we used the method of Kelly (2007) when performing the regression in order to correct for the intrinsic statistical un-
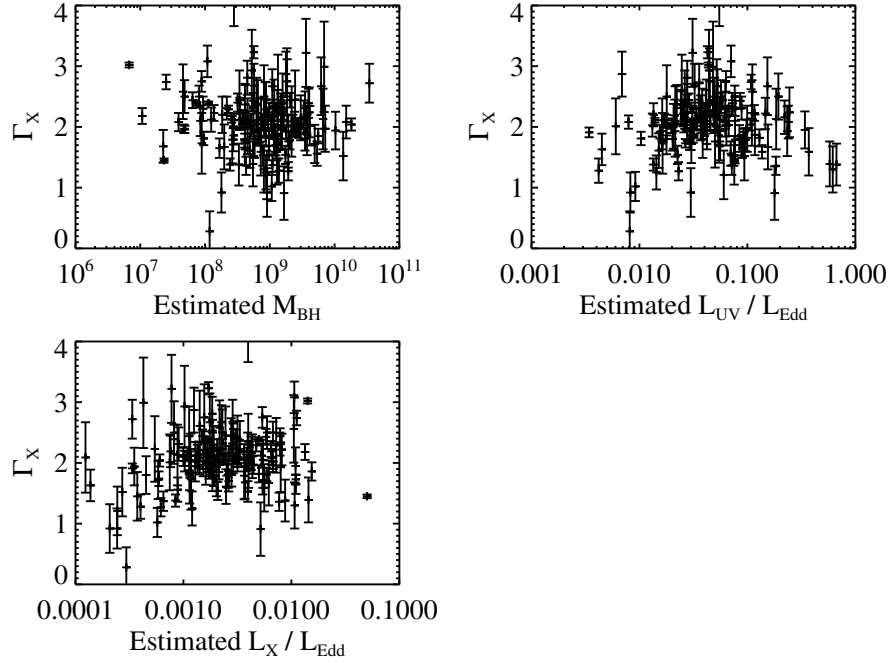
Figure 7.5 Distribution of the X-ray photon index as a function of estimated M_{BH}, L_{UV}/L_{Edd}, and L_X/L_{Edd}. For clarity, error bars are only shown on Γ_X, and we cut-off the one data point with estimated $\Gamma_X > 4$. While no obvious trends between Γ_X and M_{BH}, L_{UV}/L_{Edd}, or L_X/L_{Edd} exist for the whole sample, there is evidence of opposite trends in Γ_X for the Hβ and C IV samples.
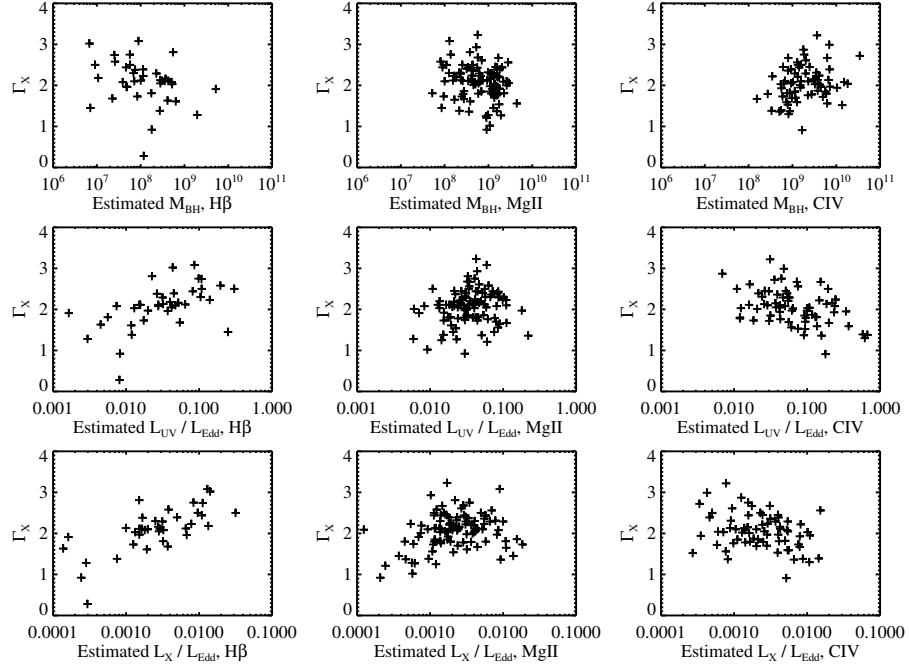
.

Figure 7.6 Distribution of the X-ray photon index as a function of estimated M_{BH}, L_{UV}/L_{Edd}, and L_X/L_{Edd} for the individual emission lines. For clarity, error bars have been omitted, and we omit the one data point with estimated $\Gamma_X > 4$. While no obvious trends exist for the whole sample, there is evidence of opposite trends for the Hβ and C IV samples.

certainty in the broad line estimates of M_{BH}. The results for M_{BH} are

$$\Gamma_X = 5.69^{+7.32}_{-4.29} - \left(0.44^{+0.53}_{-0.91}\right) \log M_{BH}, \quad \sigma = 0.43^{+0.17}_{-0.17},$$

$$\rho = -0.45^{+0.53}_{-0.44}, \ (\mathrm{H}\beta) \tag{7.15}$$

$$\Gamma_X = -19.0^{+46.0}_{-68.0} - \left(1.92^{+7.73}_{-5.18}\right) \log M_{BH}, \quad \sigma = 0.27^{+0.12}_{-0.20},$$

$$\rho = -0.62^{+1.23}_{-0.36}, \ (\mathrm{MgII}) \tag{7.16}$$

$$\Gamma_X = -2.79^{+5.22}_{-10.8} + \left(0.52^{+1.17}_{-0.56}\right) \log M_{BH}, \quad \sigma = 0.22^{+0.11}_{-0.14},$$

$$\rho = 0.53^{+0.42}_{-0.58}, (\mathrm{CIV}) \tag{7.17}$$

the results for L_{UV}/L_{Edd} are

$$\Gamma_X = 3.96^{+2.28}_{-1.15} + \left(1.23^{+1.48}_{-0.75}\right) \log L_{UV}/L_{Edd}, \quad \sigma = 0.26^{+0.23}_{-0.21},$$

$$\rho = 0.87^{+0.13}_{-0.50}, \ (\mathrm{H}\beta) \tag{7.18}$$

$$\Gamma_X = 5.13^{+10.8}_{-14.4} + \left(2.14^{+7.28}_{-9.94}\right) \log L_{UV}/L_{Edd}, \quad \sigma = 0.28^{+0.12}_{-0.22},$$

$$\rho = 0.54^{+0.44}_{-1.40}, \ (\mathrm{MgII}) \tag{7.19}$$

$$\Gamma_X = 0.85^{+0.78}_{-2.00} - \left(0.95^{+0.65}_{-1.69}\right) \log L_{UV}/L_{Edd}, \quad \sigma = 0.17^{+0.13}_{-0.13},$$

$$\rho = -0.81^{+0.53}_{-0.18}, (\mathrm{CIV}) \tag{7.20}$$

and the results for L_X/L_{Edd} are

$$\Gamma_X = 4.24^{+2.58}_{-1.43} + \left(0.85^{+1.01}_{-0.56}\right) \log L_X/L_{Edd}, \quad \sigma = 0.33^{+0.19}_{-0.21},$$

$$\rho = 0.76^{+0.22}_{-0.46}, \ (\mathrm{H}\beta) \tag{7.21}$$

$$\Gamma_X = 7.26^{+16.2}_{-7.13} + \left(1.97^{+5.96}_{-2.71}\right) \log L_X/L_{Edd}, \quad \sigma = 0.23^{+0.15}_{-0.18},$$

$$\rho = 0.77^{+0.22}_{-0.90}, \ (\mathrm{MgII}) \tag{7.22}$$

$$\Gamma_X = -0.54^{+4.05}_{-5.69} - \left(0.96^{+1.52}_{-2.17}\right) \log L_X/L_{Edd}, \quad \sigma = 0.21^{+0.13}_{-0.15},$$

$$\rho = -0.65^{+0.90}_{-0.34}(\mathrm{CIV}). \tag{7.23}$$

In these equations we have quoted the errors at 95% (2σ) confidence. The probability distributions of the slope and intrinsic dispersion are shown in Figure 7.7.
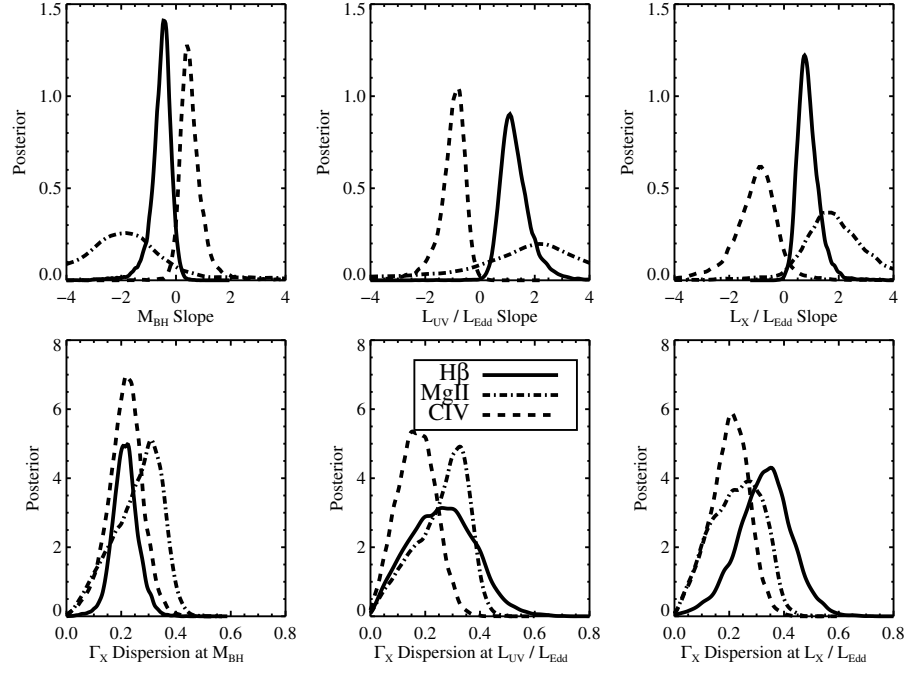
Figure 7.7 Posterior probability distributions of the slope (top) and intrinsic dispersion (bottom) from a linear regression of Γ_X on $\log M_{BH}, \log L_{UV}/L_{Edd}$, and $\log L_X/L_{Edd}$. The solid lines mark the posterior for the regression using the Hβ sample, the dashed-dotted lines mark the posterior for the regression using the Mg II sample, and the dashed lines mark the posterior for the regression using the C IV sample.

The larger uncertainty in the results for the Mg II sample is likely caused by the more narrow range in L_{UV}, L_X, and M_{BH} probed.

There are formally no significant linear correlations for the Γ_X–M_{BH} relationship. However, there is a statistically significant difference between the Hβ and C IV slopes, with $\approx 99.3\%$ of the posterior probability at $\beta_m^{\mathrm{CIV}} > \beta_m^{\mathrm{H}\beta}$, where β_m denotes the Γ_X–$\log M_{BH}$ regression slope. The probability distribution for the difference in slopes from the M_{BH} regression is shown in Figure 7.8. The significant

difference in the slope for the Hβ and C IV sample implies a nonlinear relationship between Γ_X and m_{BH}, in spite of the fact that the Hβ and C IV correlations themselves are not 'statistically significant'. Results similar to the Γ_X–$\log M_{BH}$ regressions were found for the Γ_X–L_{UV}/L_{Edd} and Γ_X–L_X/L_{UV} regressions, but with opposite sign and higher statistical significance.

We performed monte carlo simulations as a consistency check on our inferred non-monotonicity of the Γ_X relationships. While we have attempted to account for the significant statistical uncertainty on the broad line mass estimates, we employ these monte carlo simulations to ensure that the observed non-monotonic behavior is not a spurious result caused by the uncertainty on M_{BH}. We performed 10^5 simulations under two null hypotheses: (1) that Γ_X is independent of L_{UV}/L_{Edd}, and (2) that Γ_X depends linearly on $\log L_{UV}/L_{Edd}$. For both cases we simulated black hole mass estimates derived from Hβ and C IV seperately. We first simulated 'true' values of M_{BH} for each emission line from a normal distribution with means equal to the observed mean of the two respective subsamples, and variances equal to the difference between the observed variance of the subsamples and the average intrinsic variance in the broad line mass estimates. To simulate the uncertainty in the mass estimates, we added random Gaussian errors to these 'true' values of M_{BH} with standard deviation equal to the uncertainty in the mass estimates, ~ 0.4 dex. For the case where Γ_X was assumed to be independent of L_{UV}/L_{Edd}, we simulated values of Γ_X from a normal distribution with mean equal to the sample mean of Γ_X and variance equal to the difference between the observed variance in Γ_X and the average of the variance in the measurement errors. For the case where Γ_X was assumed to depend linearly on $\log L_{UV}/L_{Edd}$, we simulated values of Γ_X according to our best fit relationship to the Hβ subsample, given by Equation (7.18). Finally, for both cases we added

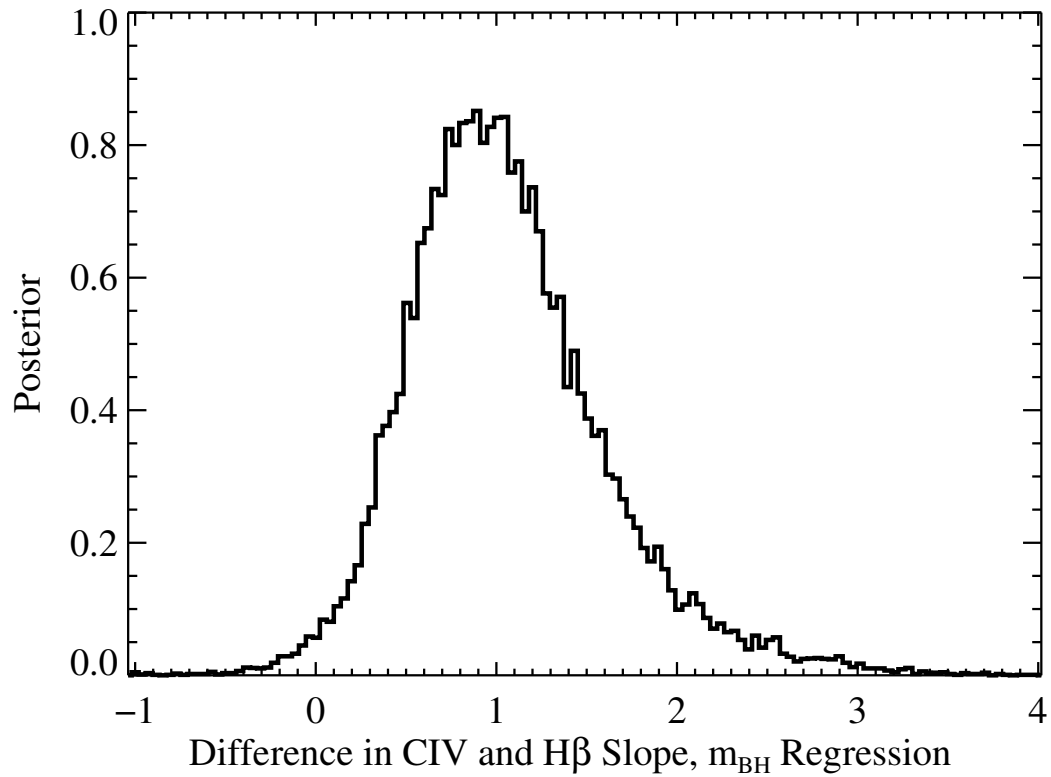Figure 7.8 Posterior distribution for the difference in slopes between the C IV and Hβ regressions of Γ_X on m_{BH}. While there is no significant evidence that either the Hβ or C IV regression slope is different from zero, there *is* significant evidence that they are not the same, implying a nonmonotonic trend between Γ_X and M_{BH}.

random Gaussian errors to the simulated values of Γ_X by randomly reshuffling the dispersions in the measurement errors in Γ_X.

For each of the 10^5 simulated samples, we selected those samples that displayed a non-monotonic trend, i.e., those sample where the slope for the Hβ-based regression had a different sign from the slope of the C IV-based regression. Under the hypothesis that Γ_X is independent of L_{UV}/L_{Edd}, only 3 of the 10^5 simulated samples had both a non-monotonic trend and an absolute value of the difference in slopes between the Hβ and C IV regression that were larger than that observed for our actual sample. Under the hypothesis that Γ_X depends linearly on $\log L_{UV}/L_{Edd}$, none of the 10^5 simulated samples exhibited a non-monotonic trend. Therefore, the observed non-monotonic trend in Γ_X with Eddington ratio is not a spurious result caused by the statistical uncertainty in the broad line mass estimates, in agreement with our Bayesian regression results.

In order to investigate whether the non-monotonicity in the dependence of Γ_X on Eddington ratio depends on M_{BH}, we performed a linear regression of Γ_X simultaneously on $\log L_{UV}/L_{Edd}$ and $\log M_{BH}$. This also allows us to quantify whether the Eddington ratio is the driver behind the Γ_X–M_{BH} relationship. In particular, the Γ_X–M_{BH} relationship is weak compared to the Eddington ratio relationships, and therefore it is reasonable to conclude that Eddington ratio is the primary driver in these relationships. We applied the multiple regression technique of Kelly (2007) seperately to both the Hβ and C IV subsamples. The results are:

$$\Gamma_X = 2.58^{+4.81}_{-3.80} + \left(0.18^{+0.60}_{-0.61}\right) \log M_{BH} + \left(1.32^{+1.45}_{-0.80}\right) \log L_{UV}/L_{Edd},$$

$$\sigma = 0.28^{+0.22}_{-0.22}, \ (\text{H}\beta) \tag{7.24}$$

$$\Gamma_X = -4.26^{+6.45}_{-15.6} + \left(0.56^{+1.44}_{-0.61}\right) \log M_{BH} - \left(0.84^{+0.75}_{-2.16}\right) \log L_{UV}/L_{Edd},$$

$$\sigma = 0.17^{+0.14}_{-0.14}, \ (\text{CIV}) \tag{7.25}$$

Here, we have quoted the errors at 95% significance. There is no statistically significant evidence that Γ_X depends on M_{BH} at a given L_{UV}/L_{Edd}, and therefore we conclude that the primary driver in the Γ_X relationships is Eddington ratio. However, this does not rule out the possibility that the non-monotonic trends with Eddington ratio are the result of a discontinuous change in the slopes at a 'critical' M_{BH}, as discussed in the next two sections.

7.7.2 Is the Sign Change in the Correlations Caused by the Different Emission Lines Used to Estimate M_{BH}?

The opposite correlations for Hβ and C IV are intriguing but may represent problems with the broad line mass estimates. In particular, it is possible that the error in the broad line mass estimates is correlated with Γ_X, but in opposite ways for Hβ and C IV. The most likely source of such a spurious correlation would be a correlation between Γ_X and the scatter about the R–L relationship for Hβ and C IV, respectively. For example, if one were to systematically overestimate R with increasing Γ_X for the C IV emitting region, then one would infer a larger M_{BH} from C IV, and thus one would infer a spurious correlation between M_{BH} and Γ_X. However, the Hβ line is only available at $z \lesssim 0.8$ and the C IV line is only available at $z \gtrsim 1.6$, and thus the change in sign for the Γ_X–M_{BH} correlations could be due to different spectral components shifting into or out of the observable X-ray spectral region (0.3–7.0 keV). In Figure 7.9, we show the distribution of M_{BH} as a function of z for the Hβ and C IV samples. As is clear from Figure 7.9, the C IV line is probing sources with $M_{BH} \gtrsim 3 \times 10^8 M_\odot$, while the H$\beta$ line is probing sources with $M_{BH} \lesssim 3 \times 10^8 M_\odot$. Therefore, the change in sign for the Γ_X correlations could also be due to something more physically interesting, such as a change in the structure of the corona that occurs at some critical black hole mass or accretion rate.
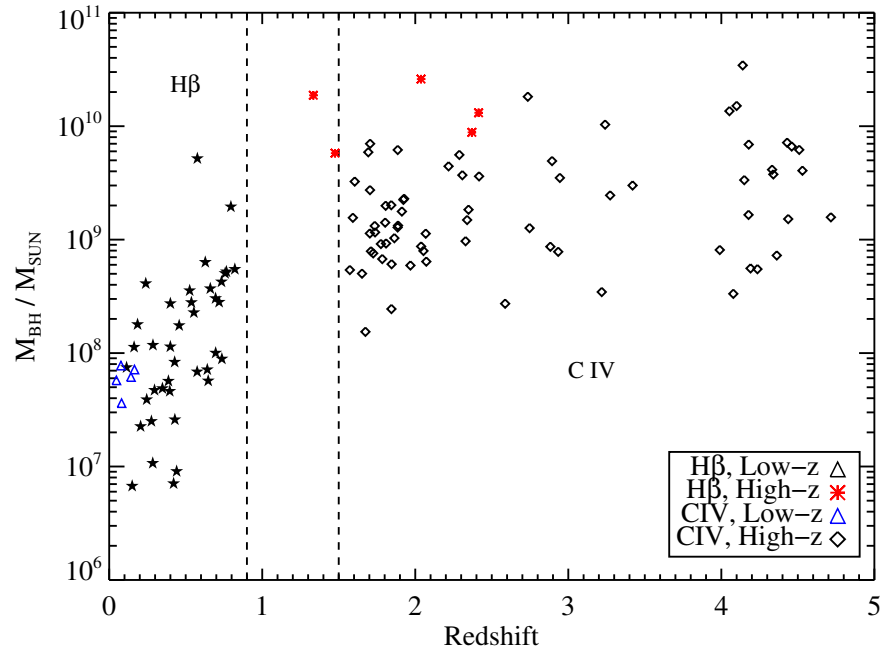
Figure 7.9 Distribution of estimated M_{BH} as a function of z for the Hβ sample (stars) and the C IV sample (open diamonds). The Hβ sample probes sources with lower M_{BH} and z, while the C IV sample probes sources with higher M_{BH} and z. To break the degeneracy between emission line, M_{BH}, and z, we have collected a sample of Hβ test sources (red asterisks) at high M_{BH} and z, and a sample of C IV test sources (open blue triangles) at low M_{BH} and z.

We can test if the change in sign for the Γ_X correlations is the result of problems with the mass estimates for either Hβ or C IV, or if it is the result of the differences in z and M_{BH} probed by the two lines. While the Γ_X–M_{BH} relationship is weak compared to the Eddington ratio dependencies, we can use the Γ_X–M_{BH} relationship to test whether the non-monotonic dependency of Γ_X on Eddington ratio is a spurious result caused by systematic difference in the broad line mass estimates. This is because the Eddington ratios are inferred from the broad line mass estimates, and any systematic differences between the Eddington ratio inferred from Hβ as compared to C IV should also manifest themselves in the weaker Γ_X–M_{BH} relationship. This is true irregardless of whether the difference in slopes between the Hβ and C IV Γ_X–M_{BH} relationships is 'statistically significant' or not.

We compiled five sources from the literature at $z > 1.3$ with Hβ-based mass estimates of $M_{BH} > 5 \times 10^9 M_\odot$. In addition, we compiled five more sources from the literature at $z < 0.2$ and C IV-based mass estimates of $M_{BH} < 10^8 M_\odot$. This 'test sample' of 10 sources is listed in Table 7.2. Then, we test whether the Hβ test sources are better described by the Hβ regression or by the C IV regression, and likewise for the C IV test sources. If the change in sign for the Γ_X correlations is due to problems with the broad line mass estimates, then we would expect the Hβ-based mass estimates to be better described by the Hβ regression. However, if the change in sign is due to the difference in redshift and M_{BH} probed by the two regressions, then we would expect the Hβ test sources to be better described by the C IV regression, as the Hβ test sources are at high-z and have high-M_{BH}. A similar argument applies to the C IV test sources, since they are at low-z and have low black hole masses.

Figure 7.10 compares Γ_X and M_{BH} for Hβ and C IV for both the sources in our

Table 7.2. Sources with Hβ and C IV Used for Testing the Γ_X–M_{BH}

Correlations

Quasar Name	α (J2000)	δ (J2000)	z	Line	$\log \hat{M}_{BL}$ $M_\odot$	$\hat{L}_{bol}/\hat{L}_{Edd}$	Opt. Ref.[a]	Γ_X	X-ray Ref.[b]
PG 0026+129	00 29 13.7	+13 16 03.8	0.142	C IV	7.789	1.768	1	1.96[c]	5
Fairall 9	01 23 45.7	-58 48 21.8	0.046	C IV	7.760	0.265	1	1.83 ± 0.06	6
PG 1202+281	12 04 42.2	+27 54 12.0	0.165	C IV	7.855	0.359	1	1.76 ± 0.07	7
PG 1211+143	12 14 17.7	+14 03 12.3	0.080	C IV	7.559	1.380	1	2.06 ± 0.05	8
PG 1247+267	12 50 05.7	+26 31 07.7	2.038	Hβ	10.41	0.379	2	2.23 ± 0.10	9
Q1346-036	13 48 44.1	-03 53 25.0	2.370	Hβ	9.946	0.609	3	2.02 ± 0.17	3
MRK 478	14 42 07.5	+35 26 22.9	0.077	C IV	7.890	0.498	1	2.41 ± 0.07	7
PG 1630+377	16 32 01.1	+37 37 50.0	1.476	Hβ	9.762	0.569	4	2.20 ± 0.30	10
PG 1634+706	16 34 28.9	+70 31 33.0	1.334	Hβ	10.27	0.734	4	2.19 ± 0.05	10
HE 2217-2818	22 20 06.8	-28 03 23.9	2.414	Hβ	10.12	0.807	3	1.97 ± 0.06	3

[a]Reference for the rest frame optical/UV data.

[b]Reference for the X-ray data.

[c]O'Neill et al. (2005) do not report an error on Γ_X.

References. — (1) Kelly & Bechtold (2007) (2) McIntosh et al. (1999) (3) Shemmer et al. (2006) (4) Nishihara et al. (1997) (5) O'Neill et al. (2005) (6) Nandra et al. (1997) (7) Brocksopp et al. (2006) (8) Reeves & Turner (2000) (9) Page et al. (2004) (10) Piconcelli et al. (2005)

main sample and the test sources, as well as the best fit regression lines for the Hβ and C IV samples, respectively. The high-z, high-M_{BH} Hβ test sources appear to be better described by the high-z, high-M_{BH} C IV-based regression, and likewise the low-z, low-M_{BH} C IV test sources appear to be better described by the low-z, low-M_{BH} Hβ-based regression.

We can quantify this result by calculating the probability that the Hβ test sources 'belong' to the Hβ-based regression, as compared to the probability that the Hβ test sources 'belong' to the C IV-based regression. Assuming that the test sources are as equally likely to belong to either regression *a priori*, this ratio of probabilities is simply the ratio of the likelihood functions of the test sources for each regression relationship, where the likelihood functions are given by Equation (24) in Kelly (2007); this ratio is called the 'Bayes Factor' (e.g., Congdon, 2006). In order to incorporate our uncertainty in the regression parameters, we use the value of the likelihood function averaged over the probability distribution of the regression parameters. We find that the Hβ test sources are ≈ 250 times more likely to 'belong' to the C IV-based regression, and that the C IV test sources are ≈ 140 times more likely to 'belong' to the Hβ-based regression. Because the test sources are independent, it follows that the test sources are $\gtrsim 10^4$ times more likely to be described by the regression fit using the opposite emission line sample. This is further evidence that Γ_X and M_{BH} are not statistically independent; if Γ_X and M_{BH} were independent, then the test sources would not show a strong preference for either regression. Based on this analysis, we conclude that the change in sign of the Γ_X correlations is not due to problems associated with the use of the Hβ and C IV emission lines, but rather due to the different range of z and M_{BH} probed by the two subsamples.
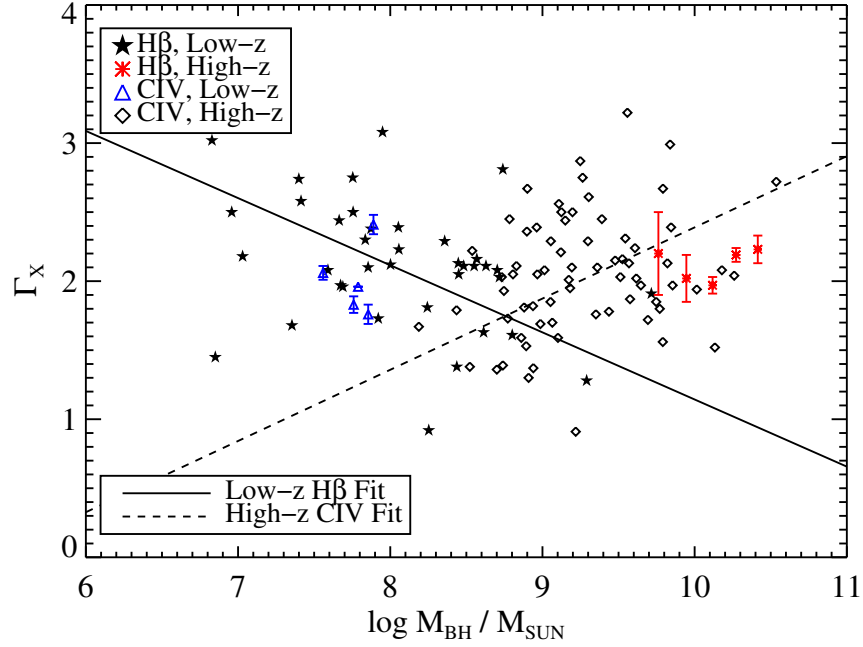
Figure 7.10 X-ray photon index as a function of estimated M_{BH} for low-z sources with M_{BH} derived from the Hβ line, high-z test sources with M_{BH} derived from the Hβ line, low-z test sources with M_{BH} derived from the C IV line, and high-z sources with M_{BH} derived from the C IV line. The symbols are the same as in Figure 7.9. Also shown is the best fit regression using the Hβ sample (solid line) and the C IV sample (dashed line). The high-z Hβ sources are better described by the high-z C IV regression, and the low-z C IV sources are better described by the low-z Hβ regression, implying that the difference in slopes between the Hβ and C IV samples is not due to systematic differences in mass estimates derived from either line.

7.7.3 Is the Sign Change in the Correlations Caused by the Different Redshift Ranges Probed?

While it appears that Γ_X has a nonmonotonic dependence on $M_{BH}, L_{UV}/L_{Edd}$, and L_X/L_{Edd}, it is unclear as to whether the sign change in the correlations is dependent on z or M_{BH}. However, we can test this in the same manner as was used to test if the sign change is due to problems with the Hβ- or C IV-based mass estimates. In this case, we need a sample of high-z, low-M_{BH} sources in order to break the degeneracy between M_{BH} and z. If the sign change in the correlation is redshift dependent, then we would expect the test sources to be better described by the high-z, high-M_{BH} C IV-based regression; but, if the sign change is black hole mass dependent, then we would expect the test sources to be better described by the low-z, low-M_{BH} Hβ-based regression.

Our test sample consists of nine $z > 1$, $\hat{M}_{BL} < 3 \times 10^8 M_\odot$ quasars from the Cosmic Evolution Survey (COSMOS, Scoville et al., 2007), with optical spectra from Magellan (Trump et al., 2007) and X-ray spectra from XMM-Newton (Mainieri et al., 2007). Black hole masses for these objects (Trump et al., 2008, in preparation) were estimated from the Mg II and C IV emission lines in the same manner as above. The objects are summarized in Table 7.3, and their location in the M_{BH}–z plane are shown in Figure 7.11. The COSMOS sources break the degeneracy between M_{BH} and z present in our SDSS sample, and are therefore adequate to test for a redshift dependence in the slope of the Γ_X–M_{BH} relationship. In Figure 7.12 we compare the COSMOS test sources with the Hβ- and C IV-based regressions. As can be seen, the high-z, low-M_{BH} COSMOS sources are better described by the low-z, low-M_{BH} regression. We can quantify this in the same manner as described in § 7.7.2 by averaging the likelihood function of the test sources over the posterior probability distribution. We find that the COSMOS test sources are

Table 7.3.   Test Sources from COSMOS

α (J2000)	δ (J2000)	z	Line	$\log \hat{M}_{BL}$ $M_\odot$	$\hat{L}_{bol}/\hat{L}_{Edd}$	Γ_X
09 58 48.8	+02 34 42.3	1.551	C IV	8.276	0.131	2.01 ± 0.11
09 59 02.6	+02 25 11.8	1.105	Mg II	7.612	0.025	2.17 ± 0.28
09 59 49.4	+02 01 41.1	1.758	C IV	8.108	0.719	2.51 ± 0.16
10 00 50.0	+02 05 00.0	1.235	Mg II	7.692	0.501	2.50 ± 0.13
10 00 51.6	+02 12 15.8	1.829	Mg II	7.807	0.131	2.14 ± 0.17
10 00 58.9	+01 53 59.5	1.559	C IV	8.346	0.172	2.04 ± 0.17
10 02 19.6	+01 55 36.9	1.509	C IV	8.333	0.177	2.19 ± 0.23
10 02 34.4	+01 50 11.5	1.506	C IV	7.991	0.941	2.25 ± 0.12
10 02 43.9	+02 05 02.0	1.234	Mg II	7.817	0.303	1.97 ± 0.29

$\gtrsim 10^5$ times more likely to be better described by the Hβ-Γ_X regression, and thus the sign change in the Γ_X correlations is not due to the difference in redshifts probed by the Hβ and C IV samples.

7.8   DISCUSSION

Previous work has found evidence for a correlation between $\alpha_{\rm ox}$ and both M_{BH} (Brunner et al., 1997) and L_{bol}/L_{Edd} (Wang et al., 2004), and for a correlation between Γ_X and both M_{BH} (e.g., Porquet et al., 2004; Piconcelli et al., 2005) and L_{bol}/L_{Edd} (e.g., Lu & Yu, 1999; Gierliński & Done, 2004; Shemmer et al., 2006), in aggreement with the results found in this work. However, our study differs from previous work in that we study a large sample of RQQs (318 sources with $\alpha_{\rm ox}$, 153 with Γ_X) over a broad range in black hole mass ($10^6 \lesssim M_{BH}/M_\odot \lesssim 10^{10}$) and redshift $0 < z < 4.8$; to date, this is the largest study of the dependence of the X-ray properties of RQQs on $M_{BH}, L_{UV}/L_{Edd}$, and L_X/L_{Edd}. In addition, this the first study of its kind to correct for the intrinsic statistical uncertainty in broad line
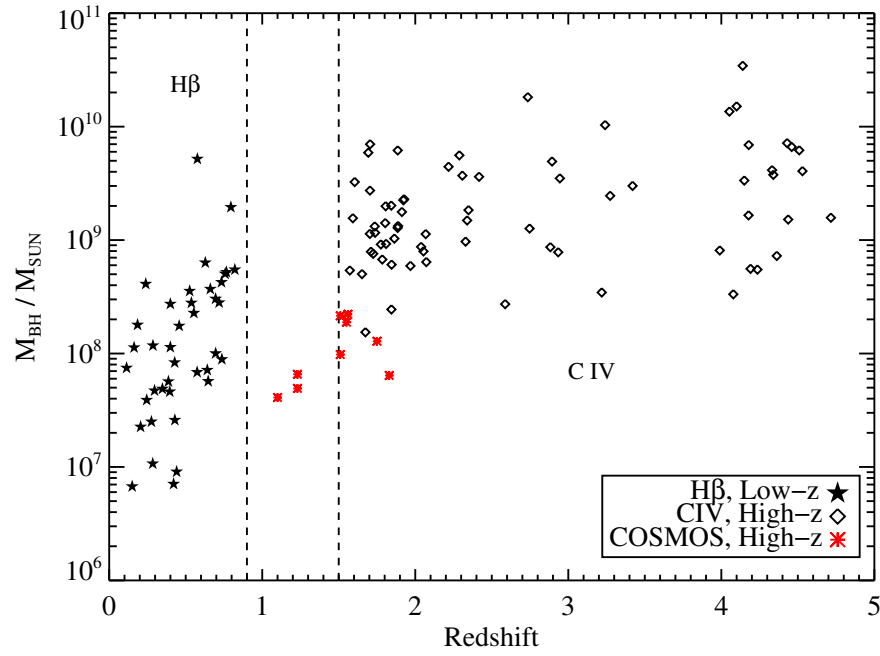
Figure 7.11 Same as Figure 7.9, but for the lower M_{BH} and higher z test sources from COSMOS. These test sources help break the degeneracy between the M_{BH} and z present in our main sample.

Figure 7.12 Same as Figure 7.10, but for a sample of high-z, low-M_{BH} test sources from *COSMOS* (red asterisks with error bars). The *COSMOS* sources are better described by the low-z, low-M_{BH} regression, implying that the difference in slopes between the Hβ and C IV samples is due to the difference in M_{BH} probed by the two samples, and not due to the redshift differences.

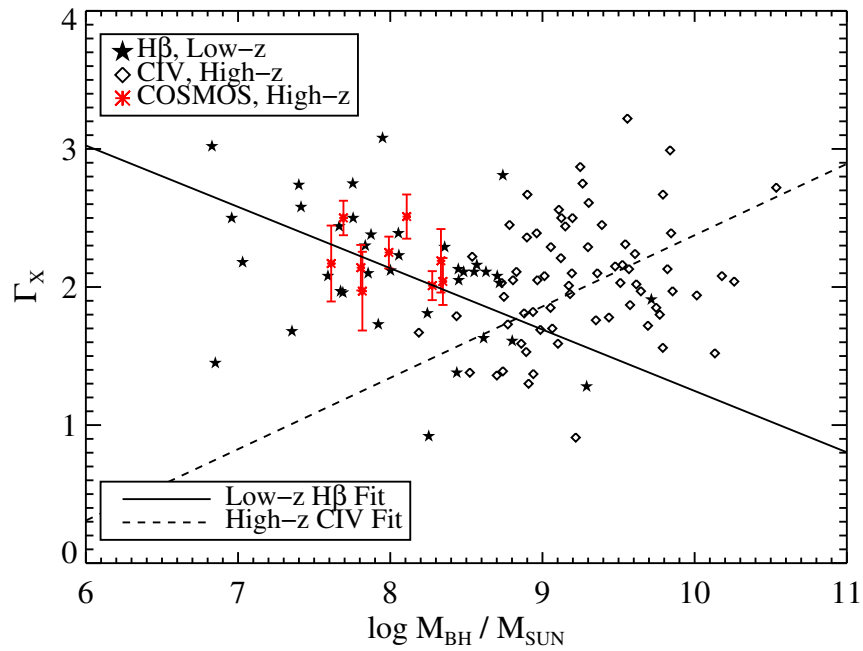mass estimates when quantifying the *intrinsic* trends between the X-ray emission and $M_{BH}, L_{UV}/L_{Edd}$, and L_X/L_{Edd}.

Currently, there are two main types of geometries being considered for the comptonizing corona. The first of these is that of a 'slab'-type geometry, possibly patchy, that sandwiches the disk (e.g., Bisnovatyi-Kogan & Blinnikov, 1977; Galeev et al., 1979; Nayakshin, 2000; Sobolewska et al., 2004b), and the second is that of a hot spherical inner advection dominated flow (e.g., Shapiro et al., 1976; Zdziarski et al., 1999); hybrids between the two geometries have also been considered (e.g., Poutanen et al., 1997; Sobolewska et al., 2004a). There is a growing body of evidence that the advection dominated hot inner flow does not exist in objects with Eddington ratios $L_{bol}/L_{Edd} \gtrsim 0.01$, as inferred from the existence of a relativistically broadened iron line (e.g., Mineo et al., 2000; Lee et al., 2002; Fabian et al., 2002), relativistically broadened reflection of ionized material (Janiuk et al., 2001), and by analogy with galactic black holes (e.g., Esin et al., 1997; Nowak et al., 2002). The range in Eddington ratios probed by our study is at most $0.03 \lesssim L_{bol}/L_{Edd} \lesssim 2$, with a mean of $L_{bol}/L_{Edd} \sim 0.25$. Therefore, the RQQs in our study are likely to have disks that extend approximately down to the last marginally stable orbit, and thus should only have the 'slab' type geometries.

7.8.1 Dependence of α_{ox} on M_{BH}

In this work we have found that RQQs become more X-ray quiet as M_{BH} increases, and confirmed the well-established relationship between α_{ox} and L_{UV}. Because L_{UV} increases with M_{BH} and the accretion rate relative to Eddington, $\dot{m}$, the well-known α_{ox}–L_{UV} correlation is likely driven by the α_{ox}–M_{BH} and α_{ox}–$\dot{m}$ correlations. A correlation between α_{ox} and M_{BH} is expected even if the fraction of the bolometric luminosity emitted by the disk is independent of M_{BH}, as the effective temperature of the disk depends on M_{BH}. As M_{BH} increases, the effec-

tive temperature of the disk decreases, thus shifting the peak of the disk emission toward longer wavelengths. Because the flux density at 2500Å lies redward of the peak in the disk SED over most of the range M_{BH} probed by our study, this shift in the disk SED toward longer wavelengths produces an increase in L_{UV} relative to L_X.

We can use the standard thin disk solution to assess the evidence that the fraction of energy emitted by the corona depends on M_{BH}. We assume a simple model where the spectrum for the disk emission is that expected for an extended thin accretion disk, and the spectrum for the corona emission is a simple power-law with exponential cutoffs at the low and high energy end. According to Wandel (2000), the spectrum from a radially extended thin accretion disk can be approximated as

$$f_\nu^D \approx A_D \left(\frac{\nu}{\nu_{co}}\right)^{-1/3} e^{-\nu/\nu_{co}}, \tag{7.26}$$

where A_D is the normalization and ν_{co} is the cut-off frequency. In this work, we choose the normalization to ensure that Equation (7.26) integrates to unity, and therefore f_ν^D gives the shape of the disk emission. For a Kerr black hole, Malkan (1991) finds that the cut-off frequency is related to M_{BH} as

$$h\nu_{co} = (6\text{eV})\dot{m}^{1/4}(M_{BH}/10^8 M_\odot)^{-1/4}. \tag{7.27}$$

We assume that the X-ray emission from the corona can be described by a simple power law with an exponential cutoff at the high and low end:

$$f_\nu^C = A_C \nu^{-(\Gamma_X - 1)} e^{-\nu/\nu_{high}} e^{-\nu_{low}/\nu}. \tag{7.28}$$

Here, A_C is the corona spectrum normalization, ν_{high} is the high energy cutoff, and ν_{low} is the low energy cutoff. We choose the low energy cutoff to be $\nu_{low} = 20\text{ eV}$, and we choose the high energy cutoff to be $\nu_{high} = 200\text{ keV}$ (e.g., Gilli et al.,

2007). As with Equation (7.26), we choose the normalization in Equation (7.28) to be equal to unity.

Denoting f_D to be the fraction of bolometric luminosity emitted by the disk, our model RQQ spectrum is then

$$L_\nu \approx L_{bol} \left[f_D f_\nu^D + (1 - f_D) f_\nu^C \right].$$ (7.29)

We computed Equation (7.29) assuming a value of $\dot{m} = 0.2$ and $f_D = 0.85$. We chose the value of the $\dot{m} = 0.2$ because it is representative of the RQQs in our sample, and we chose the value $f_D = 0.85$ because it gives values of α_{ox} typical of the RQQs in our sample. We vary M_{BH} but keep f_D and $\dot{m}$ constant because we are interested in investigating whether there is evidence that assuming independence between M_{BH} and both f_D and $\dot{m}$ is inconsistent with our α_{ox} results.

We compute Equation (7.29) for two forms of the dependence of Γ_X on M_{BH}. For the first model, we assume a constant value of $\Gamma_X = 2$. For the second model, we assume that Γ_X depends on M_{BH} according to our best fit regression results, where Γ_X depends on M_{BH} according to Equation (7.15) for $M_{BH} \lesssim 3 \times 10^8 M_\odot$, and Γ_X depends on M_{BH} according to Equation (7.17) for $M_{BH} \gtrsim 3 \times 10^8 M_\odot$. We ignore the intrinsic dispersion in Γ_X. In Figure 7.13 we show the spectra computed from Equation (7.29) for RQQs with $M_{BH}/M_\odot = 10^7, 10^8, 10^9$, and 10^{10}. The dependence of the location of the peak in the disk emission on M_{BH} is clearly illustrated.

In Figure 7.14 we compare the α_{ox}–M_{BH} regression results with the dependence of α_{ox} on M_{BH} expected from Equation (7.29) for both Γ_X–M_{BH} models. Under the thin disk approximation, a correlation is expected between α_{ox} and M_{BH}, even if the fraction of bolometric luminosity emitted by the disk is independent of M_{BH}. However, our data are inconsistent with the assumption that f_D and M_{BH} are independent, given the thin disk approximation. Under the as-
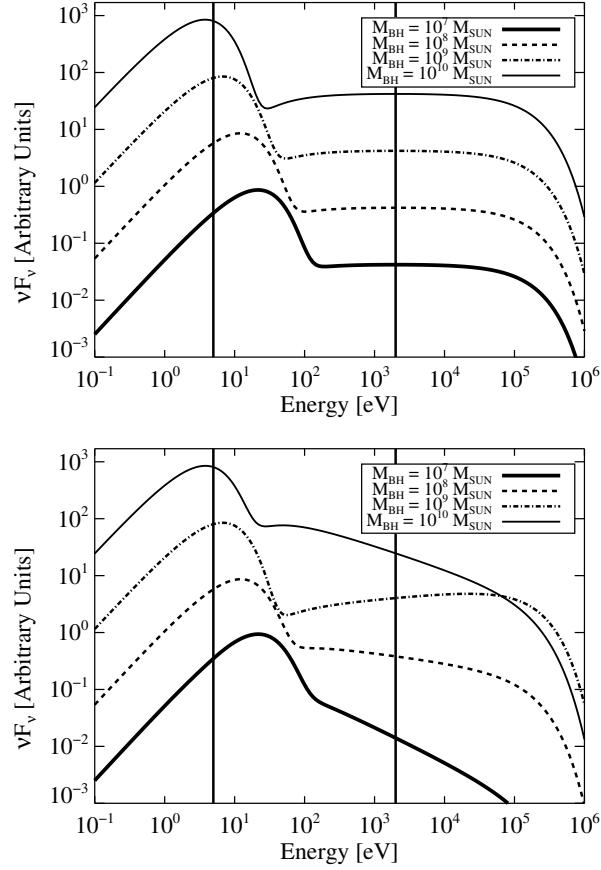
Figure 7.13 Model RQQ spectra computed from Equations (7.26)–(7.29), assuming $\Gamma_X = 2$ (top) and a varying Γ_X with M_{BH} (bottom). The spectra are computed for a RQQ with $M_{BH}/M_\odot = 10^7$ (thick solid line), 10^8 (dashed line), 10^9 (dashed-dotted line), and 10^{10} (thin solid line). In all cases we assume $\dot{m} = 0.2$ and that $f_D = 85\%$ of the bolometric luminosity is emitted by the disk. The vertical lines mark the locations of 2500Å and 2 keV. The dependence of the location of the peak in the disk emission on M_{BH} is apparent, producing a correlation between α_{ox} and M_{BH} even if the fraction of bolometric luminosity emitted by the disk is independent of M_{BH}.

.

sumption that f_D and M_{BH} are independent, the α_{ox}–M_{BH} correlation is too flat, and a increase in the fraction of bolometric luminosity emitted by the disk with increasing M_{BH} is needed to match the steeper observed dependence of α_{ox} on M_{BH}. Alternatively, if f_D increases with increasing $\dot{m}$, as we argue in § 7.8.2, then a steeper α_{ox}–M_{BH} correlation also results if M_{BH} and $\dot{m}$ are correlated. In this case, if f_D increases with increasing $\dot{m}$, and if M_{BH} increases with increasing $\dot{m}$, then f_D will also increase with increasing M_{BH}, thus producing a steeper observed α_{ox}–M_{BH} correlation.

To the extant that Equations (7.26)–(7.29) accurately approximate the spectral shape of RQQs, our data imply that either the fraction of bolometric luminosity emitted by the disk increases with increasing M_{BH}, that M_{BH} and $\dot{m}$ are correlated, or both. Some theoretical models have suggested that the fraction of bolometric luminosity emitted by the disk should depend on $\dot{m}$, but be relatively insensitive to M_{BH} (e.g., Czerny et al., 2003; Liu et al., 2003). Therefore, while a significant dependence of f_D on M_{BH} is not predicted by these disk/corona models, these models are still consistent with the interpretation that a M_{BH}–$\dot{m}$ correlation is driving the steeper α_{ox}–M_{BH} correlation. Unfortunately, without accurate estimates of $\dot{m}$ we are unable to distinguish between these two possibilities.

A shift in the peak of the disk SED with M_{BH} may also explain the dependence of α_{ox} on redshift observed by K07. K07 speculated that the observed hardening of α_{ox} with increasing z at a given L_{UV} may be due to a correlation between α_{ox} and M_{BH}, manifested through a M_{BH}–z correlation. At a given L_{UV}, an increase in M_{BH} will result in an increase in L_X relative to L_{UV}, assuming that $\dot{m}$ is not strongly correlated with M_{BH}. This is because an increase in M_{BH} decreases the temperature of the disk, shifting the peak in the disk SED toward the red, and
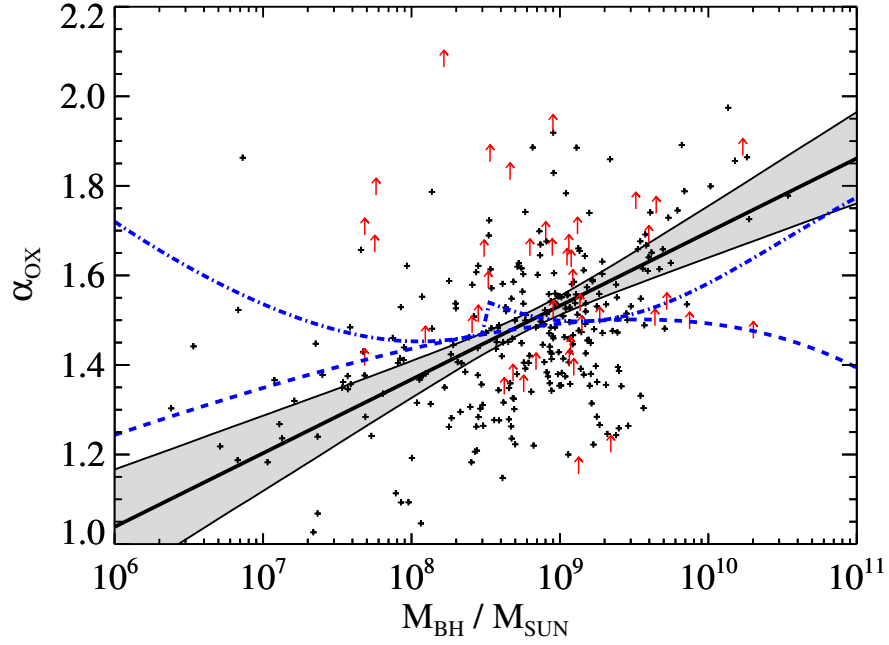
Figure 7.14 Dependence of α_{ox} on M_{BH} computed from Equations (7.26)–(7.29), assuming $\Gamma_X = 2$ (blue dashed line) and a varying Γ_X with M_{BH} (blue dot-dashed line). As with Figure 7.13, we compute Equations (7.26)–(7.29) assuming $\dot{m} = 0.2$ and $f_D = 0.85$. The predictions from the model RQQ spectra are compared with our observed data and the regression results, where the symbols and lines have the same meaning as in Figure 7.4. The α_{ox}–M_{BH} relationships predicted from assuming that f_D is independent of M_{BH} are inconsistent with the observed α_{ox}–M_{BH} relationship.

thus increasing the luminosity at 2500Å. However, since K07 investigated the dependence of α_{ox} on z at a given L_{UV}, the luminosity at 2500Å is held constant. Therefore, the overall disk emission must decrease in order to keep the luminosity at 2500Å constant despite the increase in M_{BH}. As a result, an increase in M_{BH} at a given L_{UV} will result in an increase in the X-ray luminosity relative to the luminosity at 2500Å. Because M_{BH} and z are correlated in our flux limited sample (e.g., see Figure 7.2), an increase in z will probe RQQs with higher M_{BH}. As a result, RQQs will become more X-ray loud with increasing z, at a given 2500Å luminosity. Consequently, deeper surveys that probe a greater range of M_{BH} should not see as strong of a correlation between M_{BH} and z, thereby reducing the magnitude of a α_{ox}–z correlation. Indeed, investigations based on samples that span a greater range in luminosity do not find evidence for a correlation between α_{ox} and z (e.g., Steffen et al., 2006; Just et al., 2007), qualitatively consistent with our interpretation of a α_{ox}–z correlation.

7.8.2   Dependence of α_{ox} on $\dot{m}$

We have found that α_{ox} increases with increasing L_{UV}/L_{Edd}, and decreases with increasing L_X/L_{Edd}. The mere existence of these correlations is not particularly interesting, as we would expect that the ratio of optical/UV luminosity to X-ray luminosity would increase as the fraction of optical/UV luminosity relative to Eddington increases, and vice versa for an increase in L_X/L_{Edd}. However, the relative magnitude of these dependencies carries some information regarding the dependence of α_{ox} on $\dot{m}$. A correlation between α_{ox} and L_{UV}/L_{Edd} implies that L_{UV}/L_X increases as the quantity $f_{UV}\dot{m}$ increases, where f_{UV} is the fraction of bolometric luminosity emitted at 2500Å. Likewise, an anti-correlation between α_{ox} and L_X/L_{Edd} implies that L_{UV}/L_X decreases as the quantity $f_X\dot{m}$ increases, where f_X is the fraction of bolometric luminosity emitted at 2 keV. If the fraction

of the bolometric luminosity emitted by the disk, as compared to the corona, increases with increasing $\dot{m}$, then we would expect a strong increase in L_{UV}/L_X with the product $f_{UV}\dot{m}$, resulting from the dual dependency of L_{UV}/L_X on f_{UV} and $\dot{m}$. Furthermore, because the fraction of bolometric luminosity emitted by the disk should decrease with increasing f_X, then, if the fraction of bolometric luminosity emitted by the disk increases with increasing $\dot{m}$, we would expect a weaker dependence of L_{UV}/L_X on the quantity $f_X\dot{m}$. This is because an increase in $\dot{m}$ causes an increase in the disk emission relative to the corona emission, which will then work against the decrease in disk emission relative to the corona that results from an increase in f_X. The end result is a weaker dependence of L_{UV}/L_X on the product $f_X\dot{m}$. Indeed, this is what we observe, where $L_{UV}/L_X \propto (L_{UV}/L_{Edd})^{2.5}$ and $L_{UV}/L_X \propto (L_X/L_{Edd})^{-1.5}$. Therefore, we conclude that the disk emission relative to the corona emission increases with increasing $\dot{m}$. This is in agreement with some models of corona with a slab geometry (e.g., Czerny et al., 1997; Janiuk & Czerny, 2000; Merloni & Fabian, 2002; Liu et al., 2003), where the $\alpha_{\rm ox}$–$\dot{m}$ correlation arises due to a dependency of the size of the corona on $\dot{m}$.

Our result that $\alpha_{\rm ox}$ is correlated with L_{UV}/L_{Edd} and anti-correlated with L_X/L_{Edd} is inconsistent with a constant bolometric correction to both the optical/UV and X-ray luminosities. Instead, an increase in L_{UV}/L_X with increasing $\dot{m}$ implies that the bolometric correction depends on $\dot{m}$. Because we conclude that the fraction of bolometric luminosity emitted by the disk increases with increasing $\dot{m}$, this implies that the bolometric correction to the optical/UV luminosity decreases with increasing $\dot{m}$, while the bolometric correction to the X-ray luminosity increases with increasing $\dot{m}$. The direction of this trend is consistent with the results of Vasudevan & Fabian (2007), who find that the bolometric correction to the X-ray

luminosity increases with increasing L_{bol}/L_{Edd}. Similarly, we have found evidence that the fraction of bolometric luminosity emitted by the disk depends on M_{BH}, therefore implying that the bolometric correction also depends on M_{BH}. Even if the fraction of bolometric luminosity emitted by the disk is independent of M_{BH}, the bolometric correction will still depend on M_{BH} because the location of the peak in the disk emission will shift toward longer wavelengths as M_{BH} increases. As M_{BH} varies, the luminosity at 2500Å probes a different region of the quasi-blackbody disk emission, thereby producing a dependence of bolometric correction on M_{BH}.

7.8.3 Implications for Black Hole Feedback

A significant amount of recent work suggests that radiative and mechanical feedback energy from AGN plays an important part in galaxy and supermassive black hole coevolution (e.g., Fabian, 1999; Wyithe & Loeb, 2003; Di Matteo et al., 2005). Within the context of these models, a nuclear inflow of gas, possibly the result of a galaxy merger, feeds the SMBH, thus igniting a quasar. The SMBH grows until feedback energy from the quasar is able to drive out the accreting gas, thus halting the accretion process. Hydrodynamic calculations of accretion flows have shown that the efficiency of the quasar in driving an outflow depends on the fraction of energy emitted through he UV/disk component as compared to the X-ray/corona component (Proga, 2007). The disk component produces luminosity in the UV, which is responsible for driving an outflow via radiation pressure on lines, whereas the corona component produces luminosity in the X-rays, which is responsible for driving an outflow via thermal expansion. Calculations by Proga (2007) have shown that radiation driving produces an outflow that carries more mass and energy than thermal driving. If the efficiency of black hole feedback depends on the quasar SED, any dependence on M_{BH} and $\dot{m}$ of the fraction of AGN

energy emitted in the UV as compared to the X-ray has important consequences for models of black hole growth.

Because we have found evidence that the fraction of bolometric luminosity emitted by the disk increases with increasing $\dot{m}$ and M_{BH}, this implies that black holes becomes more efficient at driving an outflow with increasing $\dot{m}$ and M_{BH}. However, the α_{ox}–M_{BH} correlation may be due to the combination of both a correlation between M_{BH} and $\dot{m}$, and a dependence of the location peak in the disk SED on M_{BH}. If the fraction of energy emitted by the disk only depends weakly on M_{BH}, as some theoretical models have suggested (e.g., Czerny et al., 2003; Liu et al., 2003), the fraction of energy emitted in the UV will still decrease with increasing M_{BH} becuase the peak of the disk emission will shift away from the UV. In this case, at a given $\dot{m}$ we would expect that black holes will become less efficient at driving an outflow with increasing M_{BH}.

7.8.4 Dependence of Γ_X on M_{BH} and $\dot{m}$

In this work we have also found evidence that Γ_X and $M_{BH}, L_{UV}/L_{Edd}$, and L_X/L_{Edd} are not statistically independent. Moreover, the dependence of Γ_X on black hole mass or Eddington ratio appears to follow a non-monotonic form, although the Γ_X–M_{BH} trend is weak compared to the dependency of Γ_X on Eddington ratio. For the Γ_X–M_{BH} relationship, the X-ray continuum hardens with increasing black hole mass until $M_{BH} \sim 3 \times 10^8 M_\odot$, after which the X-ray continuum softens with increasing black hole mass. The opposite is true for the Γ_X–L_{UV}/L_{Edd} and Γ_X–L_X/L_{Edd} trends, and further work is needed to confirm this result. Previous studies have not seen this non-monotonic trend because they have only employed the Hβ emission line, and therefore their samples have been dominated by low-z, low-M_{BH} sources.

7.8.4.1 Selection Effects

It is unlikely that the dependence of Γ_X on $M_{BH}, L_{UV}/L_{Edd}$, or L_X/L_{Edd} is due to redshifting of the observable spectra range. If this were the case, then as M_{BH} increases, so does z due to selection effects, and thus we would observe a decrease in Γ_X as the 'soft excess' shifts out of the observed 0.3–7 keV spectral range, while the compton reflection component shifts into the observed spectral range. However, there are lines of evidence that suggest that the Γ_X correlation are not due to redshifting of the observable spectral region, and that at least some of the observed dependency of Γ_X on $M_{BH}, L_{UV}/L_{Edd}$, and L_X/L_{Edd} is real. First, in § 7.7.3 we tested whether a sample of nine $z > 1$ test sources with $M_{BH} \lesssim 3 \times 10^8 M_\odot$ were better described by a regression fit using the $z > 1.5, M_{BH} \gtrsim 3 \times 10^8 M_\odot$ sources, or by a regression fit using the $z < 1, M_{BH} \lesssim 3 \times 10^8 M_\odot$ sources. We found that the test sources were better fit using the regression of similar M_{BH}, and therefore that the difference in the Γ_X–M_{BH} correlations primarily depends on M_{BH}. Second, similar trends at low redshift between Γ_X and M_{BH} or L_{bol}/L_{Edd} have been seen in other studies that only analyze the hard X-ray spectral slope (typically 2–12 keV, e.g., Piconcelli et al., 2005; Shemmer et al., 2006), and thus these studies are not effected by the soft excess. Third, the compton reflection hump is unlikely to shift into the observabable spectral range until $z \sim 1$. However, the contribution to the inferred Γ_X from compton reflection at $z \gtrsim 1$ is likely weak, if not negligible, as our $z \gtrsim 1$ sources have $M_{BH} \gtrsim 10^8 M_\odot$ and are highly luminous, and therefore are expected to only have weak reflection components (Mineo et al., 2000; Ballantyne et al., 2001; Bianchi et al., 2007).

There are two scenarios in which the non-monotonic behavior of Γ_X with M_{BH} or Eddington ratio may be artificially caused by selection. We will focus on the Eddington ratio dependency, as it is the strongest; however, our argument also

applies to M_{BH}. First, the intrinsic dependency of Γ_X on Eddington ratio could be linear with increasing intrinsic scatter at high L_{bol}/L_{Edd}. Then, an inferred non-monotonic trend would occur if we were to systematically miss quasars with high L_{bol}/L_{Edd} and steep X-ray spectra. Alternatively, there could be no intrinsic dependency of Γ_X on Eddington ratio. In this case, we would infer a non-monotonic trend if we were to systematically miss quasars with steep X-ray spectra at low and high L_{bol}/L_{Edd}, and quasars with flat X-ray spectra at moderate L_{bol}/L_{Edd}.

We do not consider it likely that the observed non-monotonic dependence of Γ_X on Eddington ratio is due solely to selection effects. K07 describes the sample selection for sources with Γ_X. With the exception of some of the $z > 4$ quasars, all sources from K07 were selected by cross-correlating the SDSS DR3 quasars with public *Chandra* observations. Almost all SDSS sources in K07 had serendipitious *Chandra* observations, and therefore were selected without regard to their X-ray properties. K07 estimated Γ_X for all sources that were detected in X-rays at the level of 3σ or higher. Therefore, the only additional criterion beyond the SDSS selection imposed by K07 is the requirement that the source had to be detected in X-ray, which was fulfilled by 90% of the quasars; the undetected sources were slightly more likely to be found at lower redshift, probably due to the presence of unidentified BAL quasars. As a result, the K07 sample selection function is essentially equivalent to the SDSS quasar selection function. Because the SDSS selects quasars based on their optical colors, the most likely cause of selection effects is the optical color selection. There is evidence that Γ_X is correlated with the slope of the optical continuum, where the X-ray continuum flattens (hardens) as the optical continuum steepens (softens) (Gallagher et al., 2005, K07). The SDSS selection probability is lower for red sources (Richards et al., 2006), so we might expect to systematically miss sources with smaller Γ_X. However, for the

two scenarios described above, this is opposite the trend needed to explain the Γ_X–L_X/L_{Edd} relationship, where we need to at least systematically miss sources with larger Γ_X. Furthermore, the drop in SDSS selection efficiency with optical spectral slope only occurs at $2 < z < 4$ (Richards et al., 2006), thus we would expect a redshift dependence for this selection effect. As we have argued above, and in § 7.7.3, the non-monotonic trends for Γ_X cannot be completely explained as the result of different redshift ranges being probed.

7.8.4.2 Implications for Accretion Physics

The dependence of Γ_X on L_{UV}/L_{Edd} and L_X/L_{Edd} is likely due to a dependence of Γ_X on $\dot{m}$. If these Γ_X correlations were due to a dependence of Γ_X on f_{UV} or f_X, then we would expect opposite trends for L_{UV}/L_{Edd} and L_X/L_{Edd}, as f_{UV} and f_X should be anti-correlated. The fact that the regression results for the Γ_X–L_{UV}/L_{Edd} and Γ_X–L_X/L_{Edd} relationships are similar implies that Γ_X depends on $\dot{m}$, and at most only weakly on f_{UV} or f_X.

A non-monotonic dependence of Γ_X on $\dot{m}$ is predicted from the accreting corona model of Janiuk & Czerny (2000), as well as a non-monotonic dependence of Γ_X on the viscosity (Bechtold et al., 2003). In addition, Γ_X is expected to steepen with increasing optical depth (e.g., Haardt & Maraschi, 1991, 1993; Czerny et al., 2003). One could then speculate that the dependence of Γ_X on M_{BH} or $\dot{m}$ is due to a non-monotonic dependence of the corona optical depth on $\dot{m}$, which may indicate a change in the structure of the disk/corona system at $\sim 3 \times 10^8 M_\odot$ or some critical $\dot{m}$. Recent work also suggests a non-monotonic dependence of the optical/UV spectral slope, α_{UV}, on $\dot{m}$ (Bonning et al., 2007; Davis et al., 2007). From this work, it has been inferred that the optical/UV continuum becomes more red with increasing $\dot{m}$ until $L_{bol}/L_{Edd} \approx 0.3$, after which the optical/UV continuum becomes more blue with increasing $\dot{m}$. Assuming the bolometric cor-

rections described in § 7.5.2, the turnover in the $\Gamma_X - L_{bol}/L_{Edd}$ relationship also occurs at $L_{bol}/L_{Edd} \approx 0.3$. Bonning et al. (2007) suggested that the turnover in spectral slope at $L_{bol}/L_{Edd} \sim 0.3$ may be due to a change in accretion disk structure, where the inner part of the accretion disk becomes thicker due to increased radiation pressure (Abramowicz et al., 1988). Bonning et al. (2007) performed a simple approximation to this 'slim disk' solution and found that it is able to produce a non-monotonic trend between optical color and Eddington ration. If the inner disk structure changes at high $\dot{m}$, this change could alter the corona structure, producing the observed trend between Γ_X and Eddington ratio.

Unfortunately, current models for corona geometry make a number of simplifying assumptions, and do not yet predict a specific relationship between $\alpha_{ox}, \Gamma_X, M_{BH}$, and $\dot{m}$. Ideally, full magneto-hydrodynamic simulations (e.g., De Villiers et al., 2003; Turner, 2004; Krolik et al., 2005) that include accretion disk winds (e.g., Murray et al., 1995; Proga & Kallman, 2004) should be used to interpret the results found in this work. However, MHD simulations have not advanced to the point where they predict the dependence of α_{ox} and Γ_X on quasar fundamental parameters, but hopefully recent progress in analytical descriptions of the magneto-rotational instability (Pessah et al., 2006, 2007) will help to overcome some of the computational difficulties and facilitate further advancement.

7.9 SUMMARY

In this work we have investigated the dependence of α_{ox} and Γ_X on black hole mass and Eddington ratio using a sample of 318 radio-quiet quasars with X-ray data from *ROSAT* (Strateva et al., 2005) and *Chandra* (Kelly et al., 2007), and optical data mostly from the SDSS; 153 of these sources have estimates of Γ_X from *Chandra*. Our sample spans a broad range in black hole mass ($10^6 \lesssim M_{BH}/M_\odot \lesssim$

10^{10}), redshift ($0 < z < 4.8$), and luminosity ($10^{43} \lesssim \lambda L_\lambda(2500\text{Å})[\text{erg s}^{-1}] \lesssim 10^{48}$). To date, this is the largest study of the dependence of RQQ X-ray parameters on $M_{BH}, L_{UV}/L_{Edd}$, and L_X/L_{Edd}. Our main results are summarized as follows:

- We show that α_{ox} is correlated with L_{UV}/L_{Edd} and anti-correlated with L_X/L_{Edd}. This result is inconsistent with a constant bolometric correction being applicable to both the optical/UV luminosity and the X-ray luminosity. This result, when taken in combination with recent work by Vasudevan & Fabian (2007), implies that constant bolometric corrections can be considerably unreliable and lead to biased results. Instead, we argue that L_{UV}/L_X increases with increasing $\dot{m}$ and increasing M_{BH}, therefore implying that the bolometric correction depends on $\dot{m}$ and M_{BH}.

- We performed a linear regression of α_{ox} on luminosity, black hole mass, L_{UV}/L_{Edd}, and L_X/L_{Edd}, and found significant evidence that α_{ox} depends on all four quantities: $L_{UV}/L_X \propto L_{UV}^{0.31\pm0.03}$, $L_{UV}/L_X \propto M_{BH}^{0.43\pm0.06}$, $L_{UV}/L_X \propto (L_{UV}/L_{Edd})^{2.57\pm0.45}$, and $L_{UV}/L_X \propto L_X/L_{Edd}^{-1.48\pm0.14}$. The dependence of α_{ox} on L_{UV} may be due to the dual dependence of α_{ox} on M_{BH} and $\dot{m}$. Because we have attempted to correct for the statistical uncertainties in α_{ox} and the broad line estimates of M_{BH}, these results refer to the *intrinsic* relationships involving α_{ox} and M_{BH}, and are not merely the relationships between α_{ox} and the broad line mass estimates, $\hat{M}_{BL} \propto L^\gamma FWHM^2$.

- A correlation between α_{ox} and M_{BH} is expected from the fact that the peak in the disk emission will shift to longer wavelengths as M_{BH} increases, even if the fraction of the bolometric luminosity emitted by the disk does not change with M_{BH}. Using a simple model for RQQ spectra, we argue that the observed α_{ox}–M_{BH} correlation is steeper than that expected if both $\dot{m}$ and

the fraction of bolometric luminosity produced by the disk are independent of M_{BH}. The observed α_{ox}–M_{BH} relationship therefore implies that either the fraction of bolometric luminosity emitted by the disk increases with increasing M_{BH}, that M_{BH} is correlated with $\dot{m}$, or both.

- A correlation between α_{ox} and $\dot{m}$ is predicted from several models of 'slab'-type corona. We argue that the weaker dependence of α_{ox} on L_X/L_{Edd} implies that L_{UV}/L_X increases with increasing $\dot{m}$. Considering that the efficiency of quasar feedback energy in driving an outflow may depend on the ratio of UV to X-ray luminosity, a correlation between α_{ox} and both M_{BH} and $\dot{m}$ has important consequences for models of black hole growth. In particular, if supermassive black holes become more X-ray quiet at higher $\dot{m}$, they will become more efficient at driving away their accreting gas, thus halting their growth.

- Because of a possible nonlinear dependence of Γ_X on $M_{BH}, L_{UV}/L_{Edd}$, or L_X/L_{Edd}, we performed seperate regressions for the black hole mass estimates obtained from each emission line. We confirmed the significant dependence of Γ_X on L_{UV}/L_{Edd} and L_X/L_{Edd} seen in previous studies as inferred from the broad line mass estimates based on the Hβ line; however, we also find evidence that the Γ_X correlations change direction when including the C IV line. In particular, for the Hβ sample, the X-ray continuum hardens with increasing M_{BH}, while for the C IV sample, the X-ray continuum softens with increasing M_{BH}. Similar but opposite trends are seen with respect to L_{UV}/L_{Edd} and L_X/L_{Edd}, and we conclude that these relationships can be interpreted as resulting from a correlation between Γ_X and $\dot{m}$. Results obtained from the Mg II line were too uncertain to interpret.

We analyzed two test samples to argue that this non-monotonic behavior is not due to the different redshifts probed by the two samples, or to problems with the estimates of M_{BH} derived from the two lines; the different trends may be due to the difference in M_{BH} probed by the two samples. A non-monotonic dependence of Γ_X on M_{BH} and/or $\dot{m}$ may imply a change in the disk/corona structure, although a non-monotonic dependence of Γ_X on $\dot{m}$ and the viscosity is predicted by some models of 'slab'-type coronal geometries.

CHAPTER 8

DEPENDENCE OF QUASAR OPTICAL VARIABILITY AMPLITUDE AND CHARACTERISTIC TIME SCALE ON BLACK HOLE MASS

8.1 CHAPTER ABSTRACT

We analyze a sample of optical light curves for 100 quasars, 70 of which have black hole mass estimates. The sources in our sample have $z < 2.8$, $10^{42} \lesssim \lambda L_\lambda(5100\text{Å}) \lesssim 10^{46}$, and $10^6 \lesssim M_{BH}/M_\odot \lesssim 10^{10}$. We model the light curves as a continuous time stochastic process, providing a natural means of modeling the characteristic time scale and amplitude of quasar variations. We employ a Bayesian approach to estimate the characteristic time scale and amplitude of flux variations; our approach is not affected by biases introduced from finite sampling effects. We find that the characteristic time scales stongly correlate with black hole mass and luminosity, and are consistent with disk orbital or thermal time scales. In addition, the amplitude of short time scale variations is significantly anti-correlated with black hole mass and luminosity. We interpret the luminosity variations as resulting from thermal fluctuations that are driven by an underlying stochastic process, such as turbulence driven by a time-varying magnetic field. In addition, the intranight variations in optical flux implied by our stochastic model are ~ 0.01 mag, consistent with current studies. Our stochastic model is therefore able to unify both long and short time scale optical variations. Our sample is the largest and broadest used yet for modeling quasar variability.

8.2 CHAPTER INTRODUCTION

It is widely accepted that the extraordinary activity associated with quasars involves accretion onto a supermassive black hole, with the UV/optical emission

arising from a geometrically thin, optically thick cold accretion disk. Aperiodic variability across all wavebands is ubiquitous in AGN, with the most rapid variations occuring in the X-rays (for a review, see Ulrich et al., 1997). The source of quasar variability is unclear, and several models have been proposed for describing the optical variability of quasars, including accretion disk instabilities (e.g., Kawaguchi et al., 1998), supernovae (e.g., Aretxaga et al., 1997), microlensing (Hawkins, 2000), and more general Poisson process models (e.g., Cid Fernandes et al., 2000). However, recent results from reverberation mapping have shown that the broad emission lines respond to variations in the continuum emission after some time lag (e.g., Peterson et al., 2004), implying that the continuum variations are dominated by processes intrinsic to the accretion disk. Variability is therefore a potentially important and powerful probe of the quasar central engine and accretion disk physics.

A successful model for quasar X-ray variability describes the X-ray variations on long time scales as being the result of perturbations in the accretion rate that occur outside of the X-ray emitting region (e.g., Lyubarskii, 1997; Mayer & Pringle, 2006; Janiuk & Czerny, 2007). These accretion rate perturbations then travel inward, modulating the X-ray emitting region. It has been suggested that the origin of such perturbations is the result of a magnetic field randomly varying in time, and may be related to the appearance of an outflow (King et al., 2004). If this model for the X-ray variability is correct, we would expect to also see variations in the optical luminosity, whose origin lies in the disk at radii farther from the central source. Therefore, understanding the origin of quasar optical variations will not only lead to a better understanding of accretion disk physics, but may also lead to a better understanding of the origin of quasar X-ray variability, possibly unifying the source of variability in the two bands.

There have been numerous previous investigations of quasar optical variability. However, because of the difficulty in obtaining high quality, well sampled light curves that cover a long time span, most previous work has involved ensemble studies of quasars, or analysis of simple correlations involving variability amplitude. The most well known result from is a tendency for AGN to become less variable as their luminosity increases (e.g., Hook et al., 1994; Garcia et al., 1999; Giveon et al., 1999; Geha et al., 2003; Vanden Berk et al., 2004; de Vries et al., 2005, and references therein). There have also been claims of a variability–redshift correlation (e.g., Cristiani et al., 1990; Cid Fernandes et al., 1996; Cristiani et al., 1996; Trèvese & Vagnetti, 2002), although the sign of this correlation varies between studies (e.g., see the list in Giveon et al., 1999). If real, the variabilty–redshift correlation is most likely caused by the fact that quasars are more variable at shorter wavelengths (e.g., Cutri et al., 1985; di Clemente et al., 1996; Helfand et al., 2001; Vanden Berk et al., 2004), corresponding to regions in the disk closer to the central black hole. Recently, a correlation between optical variability and black hole mass has been claimed (Wold et al., 2007). While many of these correlations are formally statistically signficant, they often exhibit considerable scatter when one measures variability for individual objects.

A few previous studies have employed spectral techniques, such as power spectra and structure functions, in the analysis of quasar optical light curves. From these studies it has been inferred that quasar optical light curves generally have variations of $\sim 10\%$ on timescales of months, and that the power spectra of optical light curves is well described as $P(f) \propto 1/f^2$ (Giveon et al., 1999; Collier & Peterson, 2001); power spectra of this form are consistent with random walk, or more generally, autoregressive processes. In addition Collier & Peterson (2001) analyzed a sample of optical light curves from 8 low-z Seyfert 1 galaxies. They

found that the characteristic time scales of optical variations for the AGN in their sample are $\sim$ 10–100 days and correlate with black hole mass, consistent with disk orbital or thermal time scales.

Motivated by the potential of variability for increasing our understanding of the structure of quasar accretion disks, we have compiled a sample of well-sampled optical light curves from the literature. We directly model the quasar optical light curves as a stochastic process, in contrast to previous work based on more traditional Fourier (i.e., spectral) techniques. Our method allows us to describe quasar light curves with three free parameters: a characteristic time scale, amplitude of short time scale variability, and the mean value of the light curve. In addition, our method enables us to estimate the characteristic time scale of quasar variations without the windowing effects that can plague spectral approaches. Our sample consists of 100 AGN, 70 of which have black hole mass estimates. The sources in our sample have $z < 2.8$, $10^{42} \lesssim \lambda L_\lambda(5100\text{Å}) \lesssim 10^{46}$, and $10^6 \lesssim M_{BH}/M_\odot \lesssim 10^{10}$, making this by far the largest sample yet used for this kind of study.

In this work we adopt a cosmology based on the WMAP results ($h = 0.71, \Omega_m = 0.27, \Omega_\Lambda = 0.73$, Spergel et al., 2003).

8.3   DATA

In this work we analyze a sample of 100 quasar optical light curves, compiled from the literature. Our sample consists of 55 AGN from the *MACHO* survey (Geha et al., 2003), 37 Palomar Green (PG) quasars from the sample of Giveon et al. (1999), and 8 Seyfert galaxies from the AGN Watch[1] database. We were able to obtain black hole mass estimates for 71 of the AGN, where M_{BH} has been

---

[1]http://www.astronomy.ohio-state.edu/ agnwatch/

estimated for 20 of them from reverberation mapping (Peterson et al., 2004), and M_{BH} is estimated for the remaining 51 AGN from the broad emission lines using standard scaling relationships (e.g., Vestergaard & Peterson, 2006).

8.3.1   Macho Quasars from Geha et al.(2003)

We collected R band light curves from AGN selected by the *MACHO* survey (Alcock et al., 1997, 1999). The motivation for the *MACHO* survey was to study Galactic microlensing events behind the Magellanic Clouds. However, the survey was also able to select quasars via their variability, producing 59 quasars with well-sampled light curves over a broad range in redshift ($0.1 \lesssim z \lesssim 2.8$) (Geha et al., 2003). The quasar light curves span ~ 7.5 years, and have a typical sampling interval of ~ 2–10 days, although much longer gaps exist for some light curves. In general, the *MACHO* light curves are the highest quality in our sample, often being frequently and regularly sampled. See Geha et al. (2003) for more details of the *MACHO* quasar catalogue.

Spectra for the *MACHO* quasars are presented in Geha et al. (2003), and spectra for 27 of these quasars were kindly provided to us by Marla Geha. We calculated black hole mass estimates from the source luminosity and the $FWHM$ of the Hβ, Mg II, or C IV broad emission line using standard scaling relationships (e.g., Vestergaard & Peterson, 2006; Vestergaard et al., 2008). However, the spectra are not flux-calibrated, so the luminosities at $1350, 3000$, and 5100Å were estimated from the photometric data, assuming a power-law continuum with $\alpha = 0.5, f_\nu \propto \nu^{-\alpha}$ (Richards et al., 2001). Typical uncertainties on the broad line mass estimates are ~ 0.4 dex (e.g., Vestergaard & Peterson, 2006). When combined with the measurement error in the $FWHM$ measurements, and the uncertainty due to the intrinsic scatter in quasar spectral slopes (~ 0.3, Richards et al., 2001), our typical adopted uncertainty in M_{BH} was ~ 0.45 dex for the *MACHO*

quasars.

8.3.2 PG Quasars from Giveon et al.(1999)

Giveon et al. (1999) report B and R photometric light curves for 42 PG quasars spanning ~ 7 years with a typical sampling interval of ~ 40 days. These quasars are all bright ($B < 16$ mag) and nearby ($z < 0.4$). Peterson et al. (2004) report estimates of M_{BH} and $L_\lambda(5100\text{Å})$ calculated from reverberation mapping for 12 of these quasars. Estimates of M_{BH} and $L_\lambda(5100\text{Å})$ were taken from Vestergaard & Peterson (2006) for the remaining Giveon et al. (1999) quasars. See Giveon et al. (1999) for further details.

8.3.3 Seyfert Galaxies from AGN Watch Database

The light curves for the remaining 8 AGN in our sample are from the AGN watch project. The optical light curves for the Seyfert Galaxies AKN 564, Fairall 9, MRK 279, MRK 509, NGC 3783, NGC 4051, NGC 4151, NGC 5548, and NGC 7469 were taken from the AGN watch website. In general, we used the light curves at 5100Å, with the exception of AKN 564, for which we used the R-band light curve. We excluded 3C390.3 from our analysis because it is a classified as an optically-violent variable, and thus may experience a different variability mechanism than the other AGN in our sample. Black hole masses and 5100Å luminosities were taken from Peterson et al. (2004). Although the AGN Watch sample is small compared to the other two, these sources are particularly important in our analysis because they help to anchor the low-L and low-M_{BH} end of the correlations analyzed in § 8.5.

8.4 THE STATISTICAL MODEL: CONTINUOUS AUTOREGRESSIVE PROCESS

The previous analysis of Giveon et al. (1999) and Collier & Peterson (2001) suggests that the optical variability can be well-described by a power-law power

spectrum with slope ~ 2. The power spectra of the *MACHO* quasars have not been analyzed for individual objects, although Hawkins (2007) investigated the power spectra of the ensemble of objects. In Figure 8.1 we show the geometric mean R-band power spectra for the *MACHO* quasars, along with the 90% confidence region on the geometric mean. The power spectra for the *MACHO* sources are well described by a power law of slope ~ 2, consistent with previous work. The lack of any peaks in the power spectra, as well as the aperiodic and noisy appearance of quasar light curves, suggests that quasar light curves are best modelled as a stochastic process.

8.4.1   Description of Autoregressive Processes

Power spectra of the form $P(f) \propto 1/f^2$ are consistent with a first order autoregressive (AR(1)) process. The AR(1) process is a well-studied stochastic process (e.g., see Scargle, 1981, and references therein), generated according to

$$x_i = \alpha x_{i-1} + \epsilon_i, \tag{8.1}$$

where ϵ_i is a normally-distributed random variable with zero mean and variance σ_{AR}^2, and the data x_i are observed at regular time intervals. The parameters of the AR(1) model are α and σ_{AR}^2, and α is usually constrained as $|\alpha| < 1$ in order to ensure stationarity. The case $\alpha = 1$ corresponds to a random walk. For the case of quasar light curves, the n observed data points $x_1, \ldots, x_n$ correspond to the observed fluxes at times $t_1, \ldots, t_n$, for $t_i = t_1 + (i - 1)\Delta t$.

The discrete AR(1) process is only defined for regularly sampled time series. However, astronomical time series are rarely regularly sampled, and often large gaps in time can exist. Furthermore, the AR(1) process is a discrete process, but the underlying physical process that gives rise to the observed flux is continuous. Because of these two considerations, we instead model the quasar light curves as
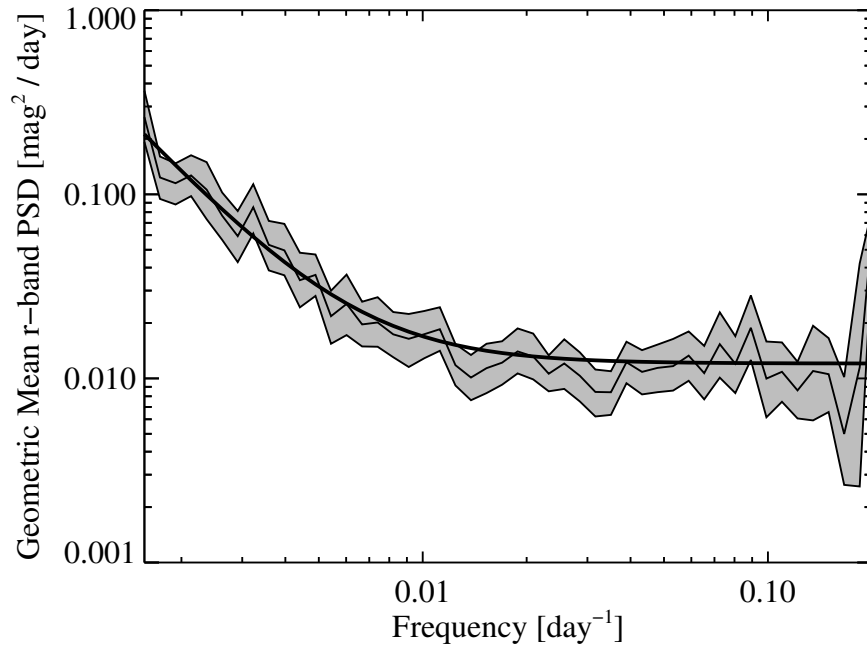
Figure 8.1 The geometric mean power spectrum (solid noisy line) for the R-band light curves of the *MACHO* quasars, along with 90% confidence region (shaded region). The thick solid line is a power spectrum of the form $P(f) \propto 1/f^2$ with an additive measurement error contribution. The optical light curves for the *MACHO* quasars are well described by a $1/f^2$ power spectrum, consistent with other samples of quasars. Power spectra of the form $1/f^2$ are suggestive of random walk and related stochastic processes.

a first order continuous autoregressive (CAR(1)) process. The CAR(1) process is described by the following stochastic differential equation[2] (e.g., Browkwell & Davis, 2002):

$$dX(t) = -\frac{1}{\tau}X(t)dt + \sigma\sqrt{dt}\epsilon(t) + b\,dt, \quad \tau, \sigma, t > 0. \tag{8.2}$$

Here, τ is called the 'relaxation time' of the process $X(t)$, and $\epsilon(t)$ is a white noise process with zero mean and variance equal to one. Throughout this work we will assume that the white noise process is also Gaussian. As we will show below, τ can be interpreted as the time required for the time series to become roughly uncorrelated, and σ can be interpreted as describing the variability of the time series on very short time scales. In the physics literature, Equation (8.2) is often referred to as an Ornstein-Uhlenbeck (O-U) process, and plays a central role in the mathememmatics of Brownian motion; see Gillespie (1995) for a review of the O-U process. Within the context of this work, $X(t)$ is the quasar light curve.

The solution to Equation (8.2) is

$$X(t) = e^{-t/\tau}X(0) + b\tau(1 - e^{-t/\tau}) + \sigma\int_0^t e^{-(t-s)/\tau}dB(s). \tag{8.3}$$

Here, $X(0)$ is a random variable describing the initial value of the time series, $dB(s)$ is a temporally uncorrelated normally distributed random variable with zero mean and variance dt. Strictly speaking, $dB(t)$ is an interval of Brownian motion, and the integral on the right side represents the stochastic component of the time series. The mean value of $X(t)$ is $b\tau$ and the variance is $\tau\sigma^2/2$. From Equation (8.3) it can be seen that if $\sigma = 0$, i.e., if there is no stochastic component, and if $X(0)$ represents a random perturbation, $X(t)$ relaxes to it's mean value with an e-folding time scale τ; hence the identification of τ as the relaxation time.

---

[2]Strictly speaking, the stochastic differential equation is complicated by the fact that white noise does not exist as a derivative in the usual sense. However, we ignore the mathematical technicalities for ease of interpretation of Equation (8.2)

If $\sigma > 0$, then the path that $X(t)$ takes will vary randomly about the expected exponential relaxation. Within the context of quasar light curves, it is tempting to associate τ with a characteristic time scale, such as the time required for diffusion to smooth out local accretion rate perturbations, and σ to represent the variability resulting from local random deviations in the accretion disk structure, such as caused by turbulence and other random magneto-hydrodynamic (MHD) effects.

The expected value of $X(t)$ given $X(s)$ for $s < t$ is

$$E(X(t)|X(s)) = e^{-\Delta t/\tau}X(s) + b\tau(1 - e^{-\Delta t/\tau}) \tag{8.4}$$

and the variance in $X(t)$ given $X(s)$ is

$$Var(X(t)|X(s)) = \frac{\tau\sigma^2}{2}\left[1 - e^{-2\Delta t/\tau}\right] \tag{8.5}$$

where $\Delta t = t - s$. If $2\Delta t/\tau \ll 1$, then Equation (8.5) implies that the variance on short time scales is $\approx \sigma^2 \Delta t$. Therefore, σ^2 can be interpreted as representing the variance in the light curve on short time scale, as stated above. In addition, one can show that when the time sampling is regular with $\Delta t = 1$, then the CAR(1) process reduces to an AR(1) process with $\alpha = e^{(-1/\tau)}$ and $\sigma_{AR}^2 = \tau\sigma^2(1-e^{-2/\tau})/2$.

In astronomical time series analysis it is common to interpret a light curve in terms of its autocorrelation function and power spectrum. The autocovariance function at time t' is defined to be the expected value of the product of $X(t)$ and $X(t + t')$, and the autocorrelation function is calculated by dividing the autocovariance function by the variance of the time series. The autocorrelation function of the CAR(1) process is

$$ACF(t') = e^{-t'/\tau}. \tag{8.6}$$

Equation (8.6) states that the correlations in CAR(1) light curve fall off exponentially with lag t', with an e-folding time equal to the relaxation time, τ. Following

Gillespie (1995), the power spectrum of a process is computed from the autocovariance function $r_X(t')$ as

$$P_X(f) = 4 \int_0^\infty r_X(t') \cos(2\pi f t') \, dt', \quad f \geq 0. \tag{8.7}$$

For a CAR(1) process, $r_X(t') = \tau\sigma^2 e^{-t'/\tau}/2$, from which if follows that the power spectrum of a CAR(1) process is

$$P_X(f) = \frac{2\sigma^2\tau^2}{1 + (2\pi\tau f)^2}. \tag{8.8}$$

From Equation (8.8) we infer that there are two important regimes for $P_X(f)$: $P_X(f) \propto 1/f^2$ for $f \gtrsim (2\pi\tau)^{-1}$ and $P_X(f) \propto 1$ for $f \lesssim (2\pi\tau)^{-1}$. Therefore, the CAR(1) process has a power spectrum that falls off as $1/f^2$ at time scales short compared to the relaxation time, and flattens to white noise at time scales long compared to the relaxation time. Because 'characteristic' time scales of quasar light curves are often defined by a break in the power spectrum, this is an additional justification of associating τ with a characteristic time scale. In addition, because the power spectra of quasar optical light curves are well described by $P_X(f) \propto 1/f^2$, it suggests that a CAR(1) process should provide a good description of the light curves, with τ being on the order of the length of the light curves or longer.

To illustrate the CAR(1) process, we simulate four CAR(1) light curves. The light curves were simulated by first simulating a random variable from a normal distribution with mean τb and variance $\tau\sigma^2/2$; note that this is the mean and variance of the CAR(1) process. Then, from this random initial value, we simulated the rest of the light curve using Equations (8.4) and (8.5). These simulated light curves span a length of 7 yrs and are sampled every 5 days. The simulated light curves span a period in time similar to the quasar light curves analyzed in this work, but are better sampled than most of the quasar light curves. Collier

& Peterson (2001) list four characteristic time scales of interest for quasars: the light crossing time, the ADAF accretion timescale (Manmoto et al., 1996), the gas orbital time scale, and the accretion disk thermal time scale. These time scales are

$$t_{lc} = 1.1 \times \left(\frac{M_{BH}}{10^8 M_\odot} \right) \left(\frac{R}{100 R_S} \right) \text{ days} \tag{8.9}$$

$$t_{acc} = 16 \times \left(\frac{M_{BH}}{10^8 M_\odot} \right) \left(\frac{R}{100 R_S} \right)^{3/2} \text{ days} \tag{8.10}$$

$$t_{orb} = 104 \times \left(\frac{M_{BH}}{10^8 M_\odot} \right) \left(\frac{R}{100 R_S} \right)^{3/2} \text{ days} \tag{8.11}$$

$$t_{th} = 4.6 \times \left(\frac{\alpha}{0.01} \right)^{-1} \left(\frac{M_{BH}}{10^8 M_\odot} \right) \left(\frac{R}{100 R_S} \right)^{3/2} \text{ yrs,} \tag{8.12}$$

where M_{BH} is the mass of the black hole, R is the emission distance from the central black hole, $R_S = 2GM_{BH}/c^2$ is the Schwarzschild radius, and α is the standard disk viscosity parameter. For the simulated quasar light curves, we use $M_{BH} = 10^8 M_\odot, \alpha = 0.01$, and $R = 100 R_S$, and set τ equal to each of these four time scales. In addition, we use $b = 0$ and $\sigma = 1$. The simulated light curves are shown in Figure 8.2, and their corresponding power spectra are shown in Figure 8.3. The increased amount of variation on long time scales with increasing τ is apparent. In addition, because t_{lc} is smaller than the time sampling, the first simulated light curve is only sampling frequencies on the flat part of the power spectrum, giving it the appearance of white noise. In contrast, the two simulated light curves with the longest time scales are sampled on the $1/f^2$ part of the power spectrum, giving them more of a 'red noise' appearance. In addition, 'red noise' leak affects the estimated power spectrum of the light curve with $\tau = 4.6$ yrs, evidenced by the constant offset between the true power spectrum and the estimated one. Red noise leak occurs when power from time scales longer than the span of the time series 'leaks' into the shorter time scales, biasing the power spectrum when estimated as the modulus of the discrete Fourier transform

(e.g., van der Klis, 1997).

8.4.2 Estimating the Parameters of a CAR(1) Process

The parameters for a CAR(1) process are commonly estimated by maximum-likelihood directly from the observed time series. This is an advantage over non-parameteric approaches, such as the discrete power spectrum or the structure function. The observed power spectrum and structure function can both suffer from windowing effects caused by the finite duration and sampling of the light curve, whereby power from high frequencies can leak to low frequencies (*aliasing*), and power at low frequencies can leak to high frequencies (e.g., red noise leak). For ground based optical observations, an additional complication is the regularity enforced in the sampling caused by the Earth's rotation around the sun, as objects are only observable during certain times of the year. For example, this periodic sampling can be seen in the light curve for the *MACHO* source shown in Figure 8.4. All of these effects can bias the power spectrum or structure function when estimated directly from the light curve in a non-parameteric fashion. In contrast, estimating a 'characteristic' time scale and variance directly from the observed time series, instead of from the observed power spectrum or structure function, has the advantage of being free of windowing effects, giving unbiased estimates of τ and σ^2. Of course, this requires one to assume a parameteric model for the time series, but as we will show below the CAR(1) process provides a good description of most of the AGN light curves analyzed in this work. Furthermore, higher-order terms can be added to Equation (8.2) to allow additional flexibility (e.g., Browkwell & Davis, 2002), but this is beyond the scope of the current work.

When the data are measured without error, the likelihood function for the CAR(1) process may be derived from Equations (8.4) and (8.5), and noting that
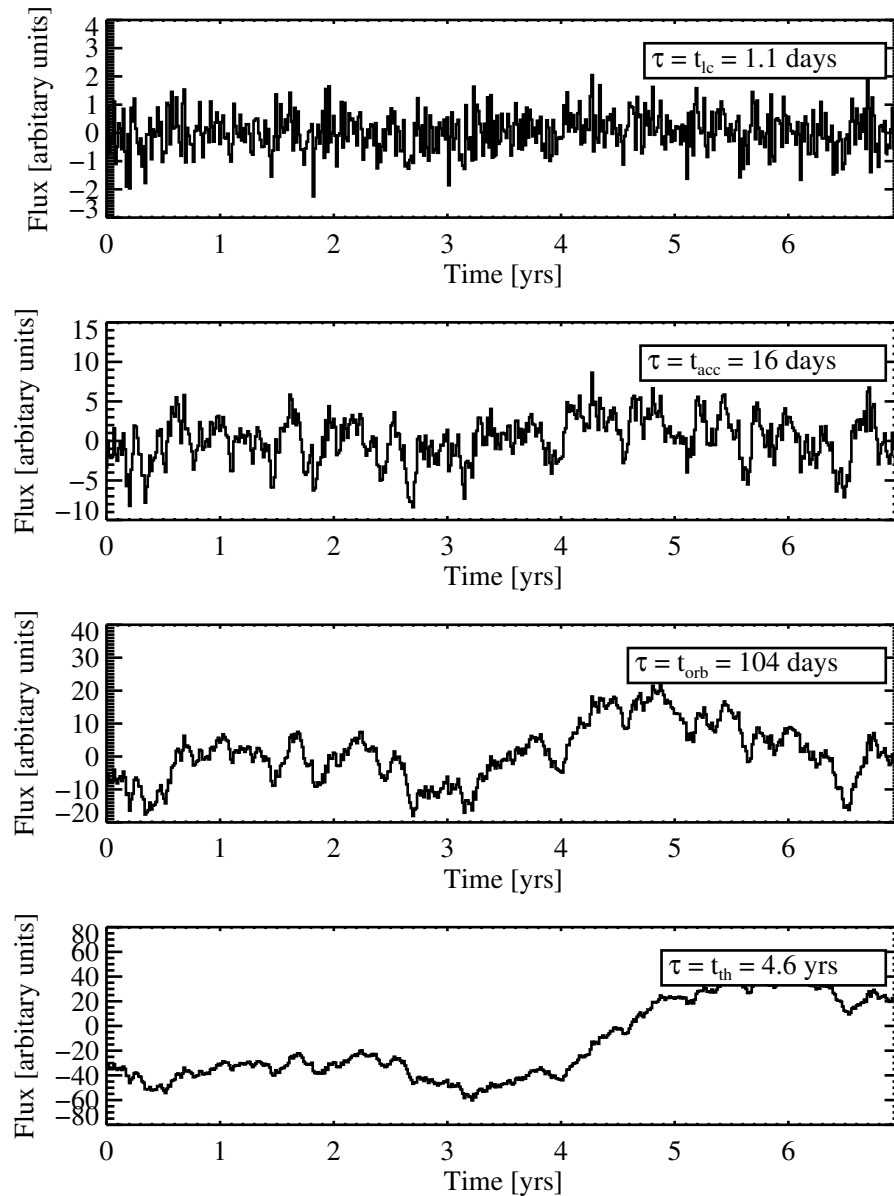
Figure 8.2 Light curves simulated from a CAR(1) process for four different characteristic time scales, assuming typical parameters for quasars (see Eq.[8.9]–[8.12]). From top to bottom, these are the light crossing time, $\tau = 1.1$ days, the ADAF accretion time scale, $\tau = 16$ days, the disk orbital time scale, $\tau = 104$ days, and the disk thermal time scale, $\tau = 4.6$ yrs. The stochastic nature of the CAR(1) process is apparent, and the light curve exhibits more variability on longer time scales as the characteristic time scale increases.
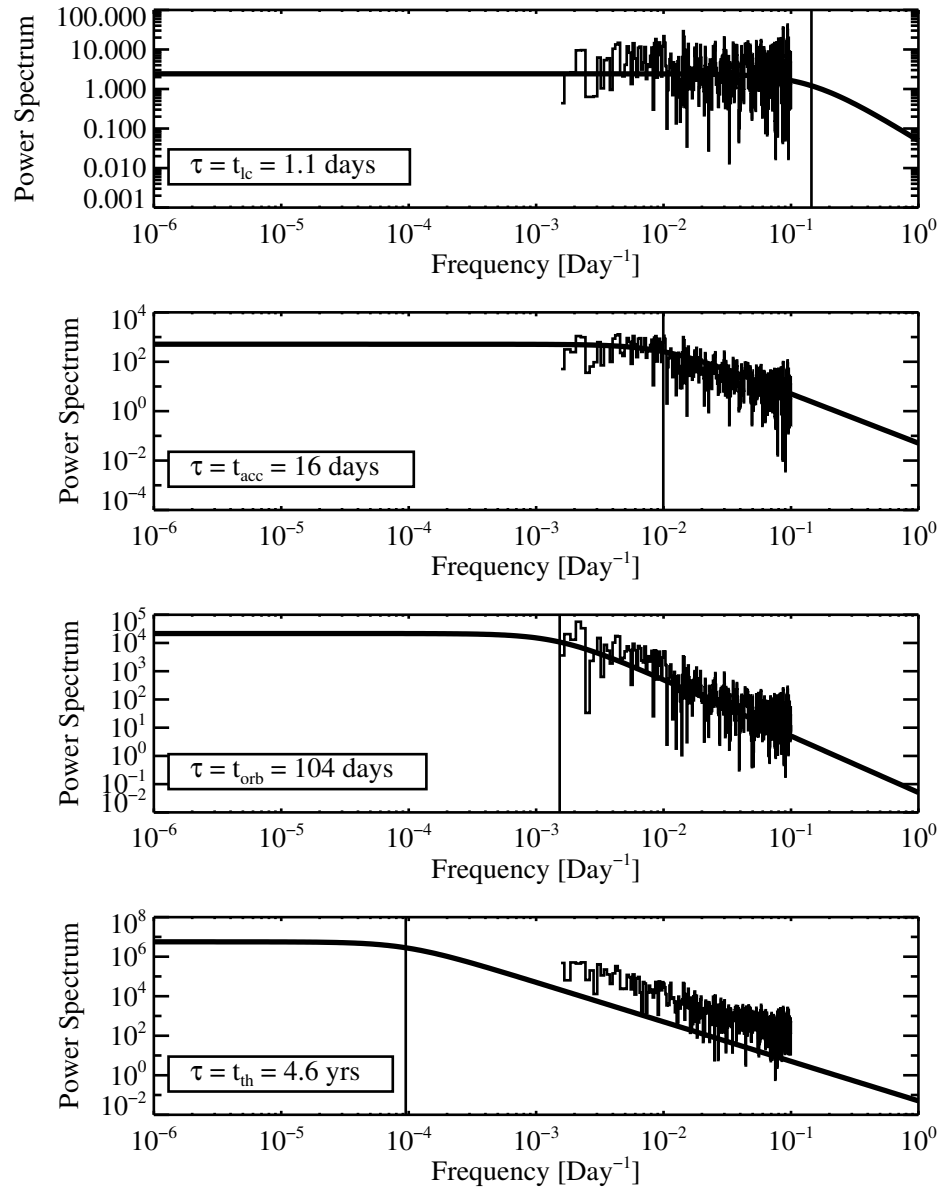
Figure 8.3 Power spectra for the simulated CAR(1) light curves shown in Figure 8.2. The actual power spectra are shown with a solid line, and the empirical power spectra estimated directly from the light curves are the noisy curves. The power spectra are flat on the 'white noise' part of the curve, corresponding to frequencies $f \lesssim (2\pi\tau)^{-1}$, and fall off as $1/f^2$ on the 'red noise' part of the curve, $f \gtrsim (2\pi\tau)^{-1}$. As τ increases, the break in the power spectra, marked with a vertical line, shifts toward shorter frequencies. For the CAR(1) process with $\tau = t_{th}$, red noise leak biases the power spectrum estimated directly from the simulated light curve.
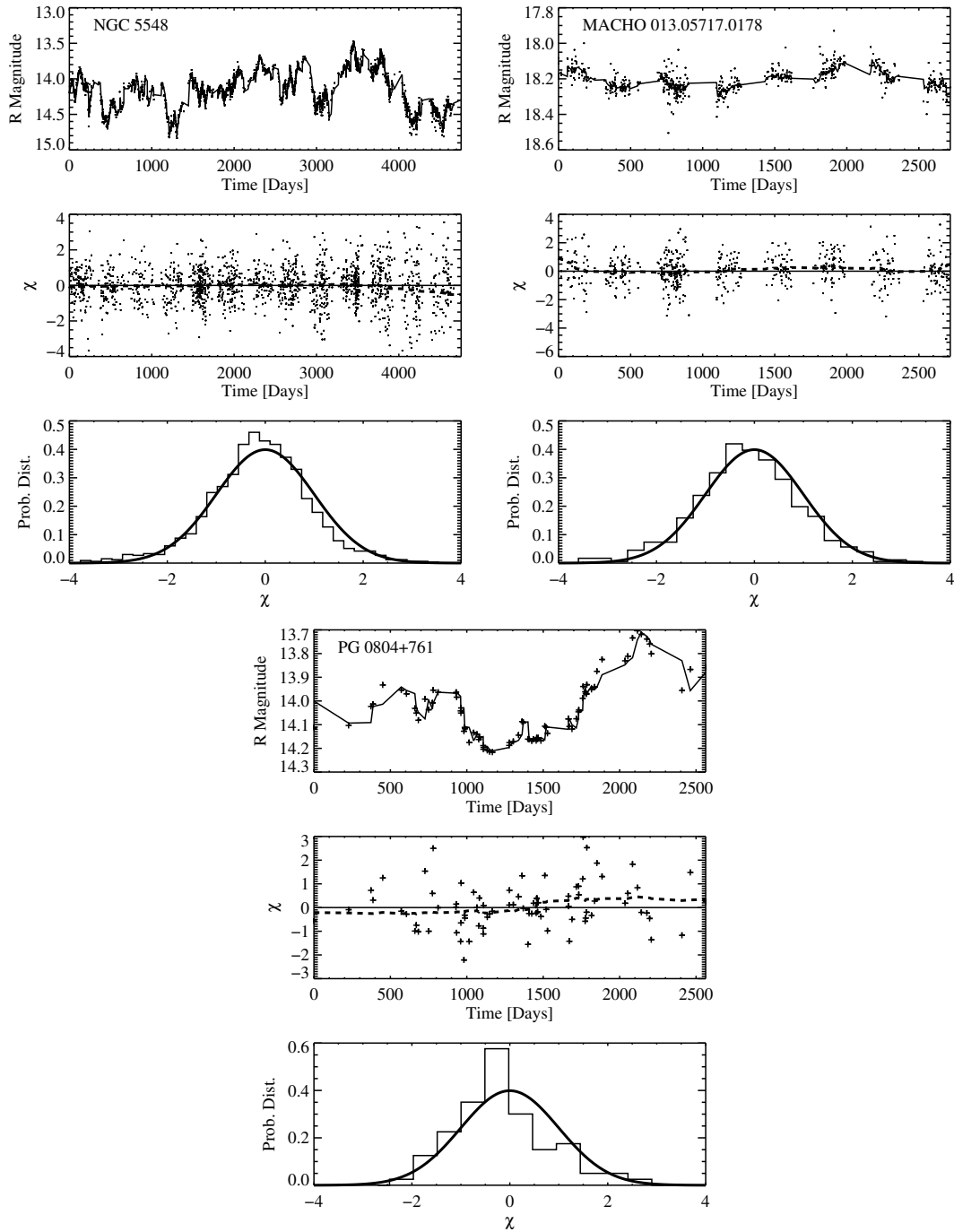
Figure 8.4 Light curves and best fit CAR(1) processes for NGC 5548 (top left), MACHO source 013.05717.0178 (top right), and PG 0804+761 (bottom). The top panels show the light curves (data points) along with the best fit CAR(1) process (solid line), the middle panels show the standardized residuals (Eq.[8.21]) as a function of time (data points) and a moving average estimate (dashed line), and the bottom panels compare a histogram of the standardized residuals with the

the initial value of the light curve follows a normal distribution with mean $b\tau$ and variance $\sigma^2\tau/2$. However, it is almost always the case that the observed fluxes are measured with error. The likelihood function can then be calculated using Equations (8.4) and (8.5) in combination with a 'state-space' representation of the time series (e.g., Browkwell & Davis, 2002). Denoting the measured fluxes as $x_1, \ldots, x_n$, observed at times $t_1, \ldots, t_n$ with measurement error variances $\sigma_1^2, \ldots, \sigma_n^2$, the likelihood function is

$$
\begin{aligned}
p(x_1, \ldots, x_n | b, \sigma, \tau) &= \prod_{i=1}^{n} \left[2\pi(\Omega_i + \sigma_i^2) \right]^{-1/2} \exp \left\{ -\frac{1}{2} \frac{(\hat{x}_i - x_i^*)^2}{\Omega_i + \sigma_i^2} \right\} & (8.13) \\
x_i^* &= x_i - b\tau & (8.14) \\
\hat{x}_0 &= 0 & (8.15) \\
\Omega_0 &= \frac{\tau\sigma^2}{2} & (8.16) \\
\hat{x}_i &= a_i \hat{x}_{i-1} + \frac{a_i \Omega_{i-1}}{\Omega_{i-1} + \sigma_{i-1}^2} \left(x_{i-1}^* - \hat{x}_{i-1} \right) & (8.17) \\
\Omega_i &= \Omega_0 \left(1 - a_i^2 \right) + a_i^2 \Omega_{i-1} \left(1 - \frac{\Omega_{i-1}}{\Omega_{i-1} + \sigma_{i-1}^2} \right) & (8.18) \\
a_i &= e^{-(t_i - t_{i-1})/\tau}. & (8.19)
\end{aligned}
$$

The maximum-likelihood estimate is then found by maximizing Equation (8.13) with respect to b, τ, and σ.

In this work we employ a Bayesian approach in order to directly compute the probability distribution of b, τ, and σ, given our observed light curves. The probability distribution of the parameters, given the observed data (i.e., the *posterior* distribution), is calculated as the product of the likelihood function with a prior probability distribution. In this work we assume a uniform prior on b and σ. For deriving a prior on τ, we note that when the data are regularly sampled, then the CAR(1) process reduces to the AR(1) process described by Equation (8.1) with $\alpha = e^{-1/\tau}$. In this work we assume a uniform prior on α. For an AR(1) process, α gives the correlation of x_i and x_{i-1}. Therefore, we consider it reasonable to as-

sume that any value of α is *a priori* likely, i.e., we do not assume anything *a priori* about the correlations between subsequent data points, and therefore we assume a uniform prior on α from 0 to 1. This prior is non-informative in the sense that all of the information on α comes from the data. Because $p(\tau)d\tau = p(\alpha)d\alpha$, this corresponds to taking the prior on τ to be $p(\tau) \propto \tau^{-2}e^{-1/\tau}$. Therefore, the probability distribution of the parameters, given the observed data, is

$$p(b, \sigma, \tau | x_1, \ldots, x_n) = \frac{1}{\tau^2} e^{-1/\tau} p(x_1, \ldots, x_n | b, \sigma, \tau), \tag{8.20}$$

where the likelihood function is given by Equation (8.13).

The accuracy of the fit can be assessed by comparing the residuals of the light curve with the values expected under the assumption of a CAR(1) process. From Equation (8.13) it is apparent that if the CAR(1) process provides a good model of the observed data, then the residuals should be uncorrelated and follow a normal distribution:

$$\chi \equiv \frac{x_i^* - \hat{x}_i}{\sqrt{\Omega_i + \sigma_i^2}} \sim N(0, 1). \tag{8.21}$$

Here, the notation $\chi \sim N(0, 1)$ means that χ is distributed according to a normal distribution with mean equal to zero and variance equal to one. The goodness of fit can then be assessed by inspecting a plot of the residuals with time to ensure that they are uncorrelated, and by comparing a histogram of the residuals with the expected standard normal distribution.

8.4.3   Fitting the Quasar Light Curves

In this work we model the logarithm of the flux as following a CAR(1) process, or equivalently the apparent magnitudes. We do this because the assumption of a Gaussian white noise process in Equation (8.2) produces both positive and negative values of $X(t)$, while flux is a strictly positive quantity. The logarithm maps a strictly positive quantity to the interval $(-\infty, \infty)$, and therefore Equation

(8.2) is likely to be a better description of the light curve for the source apparent magnitude, as opposed to the source flux.

An additional correction is needed to correct for cosmological time dilation. Because the quasars are fit in the observed frame, and time scales decrease as $(1 + z)$, spurious correlations may arise if one does not correct to the quasar rest frame. This is particularly problematic when dealing with flux limited samples, which create an artificial correlation between z and both luminosity and M_{BH}. Noting that $dt_{obs} = (1 + z)dt_{rest}$, Equation (8.2) can be expressed in the forms

$$
\begin{aligned}
dX(t) &= -\frac{1}{\tau_{obs}}X(t)dt_{obs} + \sigma_{obs}\sqrt{dt_{obs}}\epsilon(t) + b_{obs}\,dt_{obs} && (8.22)\\
&= -\frac{1+z}{\tau_{obs}}X(t)dt_{rest} + \sigma_{obs}\sqrt{(1+z)dt_{rest}}\epsilon(t) + (1+z)b_{obs}\,dt_{rest} && (8.23)\\
&= -\frac{1}{\tau_{rest}}X(t)dt_{rest} + \sigma_{rest}\sqrt{dt_{rest}}\epsilon(t) + b_{rest}\,dt_{rest}. && (8.24)
\end{aligned}
$$

From Equations (8.22)–(8.24) it is apparent that the observed and rest frame parameters are related as

$$
\begin{aligned}
\tau_{rest} &= (1+z)^{-1}\tau_{obs} && (8.25)\\
\sigma_{rest} &= (1+z)^{1/2}\sigma_{obs} && (8.26)\\
b_{rest} &= (1+z)b_{obs}. && (8.27)
\end{aligned}
$$

Noting that the mean of the CAR(1) process is $b\tau$, and that the variance is $\tau\sigma^2/2$, Equations (8.25)–(8.27) imply that the mean and variance of a CAR(1) process are unaffected by cosmological time dilation. However, the variance observed from a light curve with a finite duration and sampling is still affected by time dilation, as the observed variance over a time interval Δt is the integral of Equation (8.8) over that time interval. In what follows, the quantities τ, σ, and b will always refer to the quasar rest frame quantities, unless specified otherwise.

Random draws of $b, \tau,$ and σ from the posterior probability distribution are obtained using a Metropolis-Hastings algorithm. We calculated an estimate of each parameter as the median of the posterior, and the posterior medians, standard deviations, and 95% (2σ) confidence intervals are computed using these random draws. In general the posterior median values were not significantly different from a maximum likelihood fit. The goodness of fit for the light curves was determined by examining a histogram of the residuals and a plot of the residuals against time, as described in § 8.4.2. Occasionally outlying values of the flux are present for the light curves with more data points, possibly due to unidentified systematic error. These outlying data points were removed and the light curves were refit. In Figure 8.4 we show the light curve and best fit CAR(1) model for the most densely sampled object in our sample, NGC 5548, for a representative light curve from the *MACHO* sample, and for a representative light curve from the PG sample of Giveon et al. (1999). In general, the CAR(1) model provided a good fit to the quasar light curves analyzed in this work, and we only flagged 9 out of 109 light curves as having a bad fit. The 9 objects for which the fit was deemed unacceptable are not used in the regression analysis.

8.5   RESULTS

The distribution of $\sigma, \tau,$ and light curve standard deviations for our sample are shown in Figure 8.5. The light curve standard deviation is calculated as the square root of the light curve variance, $s = \sigma\sqrt{\tau/2}$. The best fit quasar relaxation times have a median value of 540 days and a dispersion of 0.64 dex, and show typical long time scale optical variations of 3–30 per cent. However, the uncertainties on τ are large and make a considerable contribution to the observed scatter. Correcting for the contribution from the uncertainties implies an intrinsic
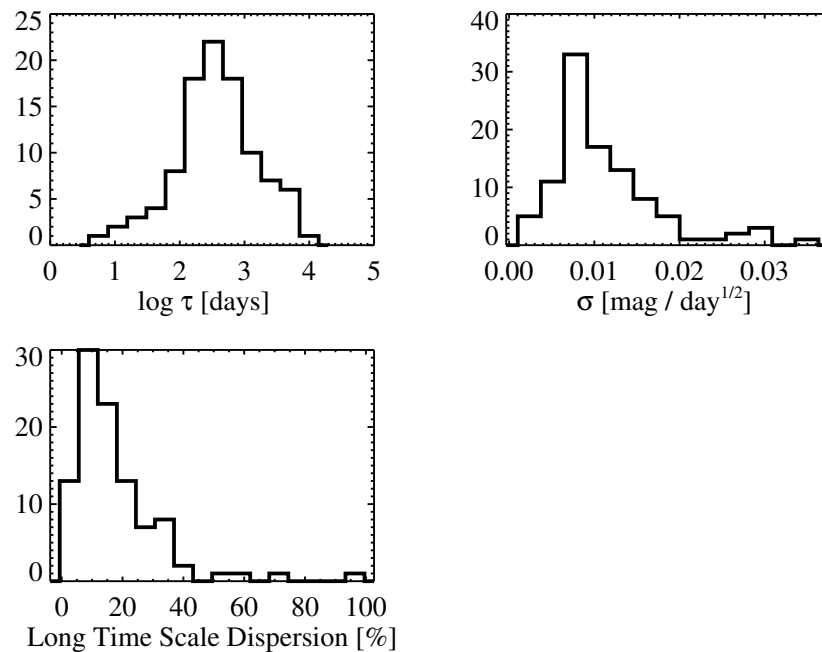
Figure 8.5 Distribution of the best fit CAR(1) process parameters for the quasars in our sample. The characteristic time scales of AGN optical light curves are $10 \lesssim \tau \lesssim 10^4$ days, the amplitudes of short time scale scale variations are $\sigma \lesssim 0.02\,\mathrm{mag\,day}^{-1/2}$, and the amplitudes of long time scale variations are $\lesssim 40\%$. The uncertainties on the characteristic time scales are large, and the true dispersion in τ is likely ~ 0.3 dex.

dispersion in relaxation time of ~ 0.3 dex.

In order to look for correlations of quasar variability properties on luminosity, redshift, black hole mass, and Eddington ratio, we used the linear regression method of Kelly (2007). The method of Kelly (2007) takes a Bayesian approach to linear regression, and accurately accounts for intrinsic scatter in the regression relationships, as well as measurement errors in both the dependent and independent variables. Measurement errors can be large for both the estimates of τ and

M_{BH}, and thus can have a significant effect on the observed correlations (e.g., Kelly et al., 2007). Therefore, it is necessary to correct for the measurement errors when attempting to recover any underlying trends.

8.5.1 Dependence of Quasar Variability on Luminosity and Redshift

In order to investigate whether quasar variability properties depend on luminosity, redshift, or both, we performed a regression analysis. Throughout this section, the luminosity will always be taken to be λL_λ at 5100Å. There has been considerable debate over whether quasar variability is correlated with luminosity or redshift, and the artificial correlation between the two has made it difficult for previous work to uncover the true intrinsic correlation. However, because we perform a linear regresion of variability properties on both L and z simultaneously, we are able to break the degeneracy between L and z. This is because the multiple linear regression describes how variability depends on L at a given z, and likewise for z at a given L.

In Figure 8.6 we show the relaxation time τ as a function of luminosity and redshift, and in Figure 8.7 we show σ as a function of luminosity and redshift. The results of the regressions for τ are

$$\log \tau = (-10.29 \pm 3.76) + (0.29 \pm 0.08) \log \lambda L_\lambda \ \text{[days]} \tag{8.28}$$

$$\log \tau = (2.32 \pm 0.10) + (1.12 \pm 0.41) \log(1+z) \ \text{[days]} \tag{8.29}$$

$$\log \tau = (-8.13 \pm 0.12) + (0.24 \pm 0.12) \log \lambda L_\lambda +$$
$$(0.34 \pm 0.58) \log(1+z) \ \text{[days]}, \tag{8.30}$$

and the results of the regression for σ are

$$\log \sigma = (4.73 \pm 2.34) - (0.19 \pm 0.05) \log \lambda L_\lambda \ \text{[R mag}^2/\text{day]} \tag{8.31}$$

$$\log \sigma = (-3.84 \pm 0.06) - (0.32 \pm 0.25) \log(1+z) \ \text{[R mag}^2/\text{day]} \tag{8.32}$$

$$\begin{aligned} \log \sigma \ &= \ (8.00 \pm 3.29) - (0.27 \pm 0.07) \log \lambda L_\lambda + \\ &\quad (0.47 \pm 0.33) \log(1 + z) \ \ [\text{R mag}^2/\text{day}]. \end{aligned} \tag{8.33}$$

There is a statistically significant correlation between τ and both L and z, where the light curve relaxation time increases with increasing L and z. In addition, there is a statistically significant anti-correlation between σ and both L and z, implying that the short time scale variance decreases with increasing L and z. However, the multiple regression results show that the luminosity trends are the dominant ones, and that there is no significant trend between either τ or σ and z, at a given L. This therefore implies that the observed trends with redshift are caused by the artificial correlation between L and z resulting from selection effects.

We also looked for trends of the light curve variance, $\tau\sigma^2/2$, with L and z and found statistically significant evidence for a correlation between the variance of the light curve and z. Both the Spearman and Kendall rank correlation statistic was significant at 3σ, although there is considerable scatter in the correlation. This correlation is most likely a reflection of the well-known fact that quasar emission at shorter wavelengths is more variable (Vanden Berk et al., 2004, e.g.,[)

8.5.2   Dependence of Quasar Characteristic Time Scale and Variability on M_{BH}

We also looked for trends in quasar variability properties with M_{BH} and the Eddington ratio, L/L_{Edd}. In this work we assume a constant bolometric correction of $C_{bol} = 10$ to the luminosity at 5100Å(Kaspi et al., 2000). However, we stress that a constant bolometric correction can introduce significant error in the Eddington ratio (e.g., Vasudevan & Fabian, 2007; Kelly et al., 2008), and that our use of $C_{bol} = 10$ is only suggestive. Strictly speaking, what is being used in the following regressions is the ratio $\lambda L_\lambda(5100\text{Å})/M_{BH}$, and in general this will not
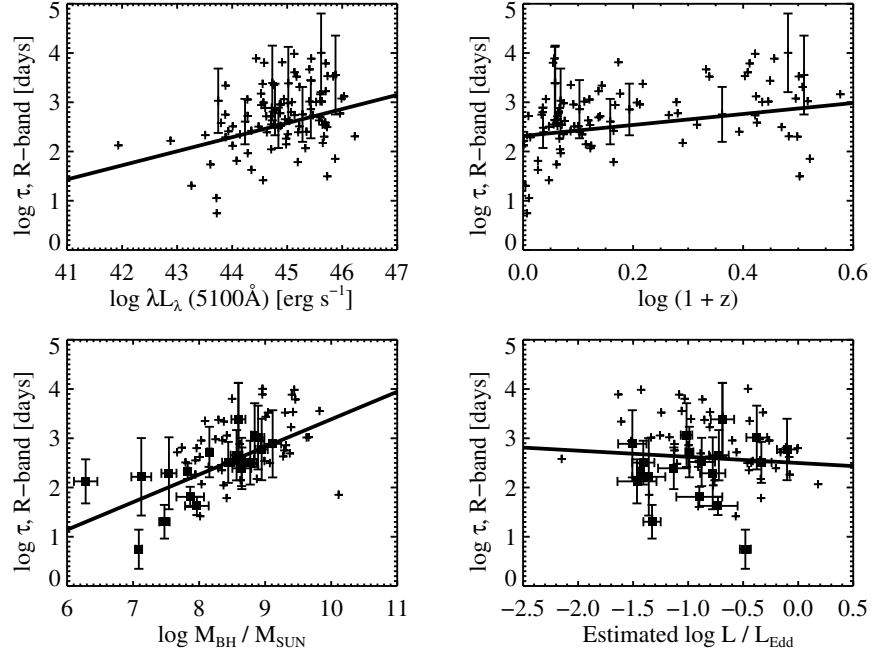
Figure 8.6 The characteristic time scale of the optical light curves for the AGN in our sample as a function of optical luminosity, redshift, black hole mass, and estimated Eddington ratio. For clarity, we only show error bars for a random fraction of the data points in the top two panels, and we only show the error bars for the sources with M_{BH} estimated from reverberation mapping in the bottom two panels. The straight lines denote the best fit linear regression. There is a significant trend for τ to increase with increasing M_{BH}, and less significant trends between τ and λL_λ or z. There is no significant trend between τ and L/L_{Edd}.

Figure 8.7 Same as Figure 8.6, but for the variance in the short time scale variations, σ^2. There is a significant trend for σ to decrease with increasing M_{BH}, and a similar but less significant trend between σ and λL_λ.
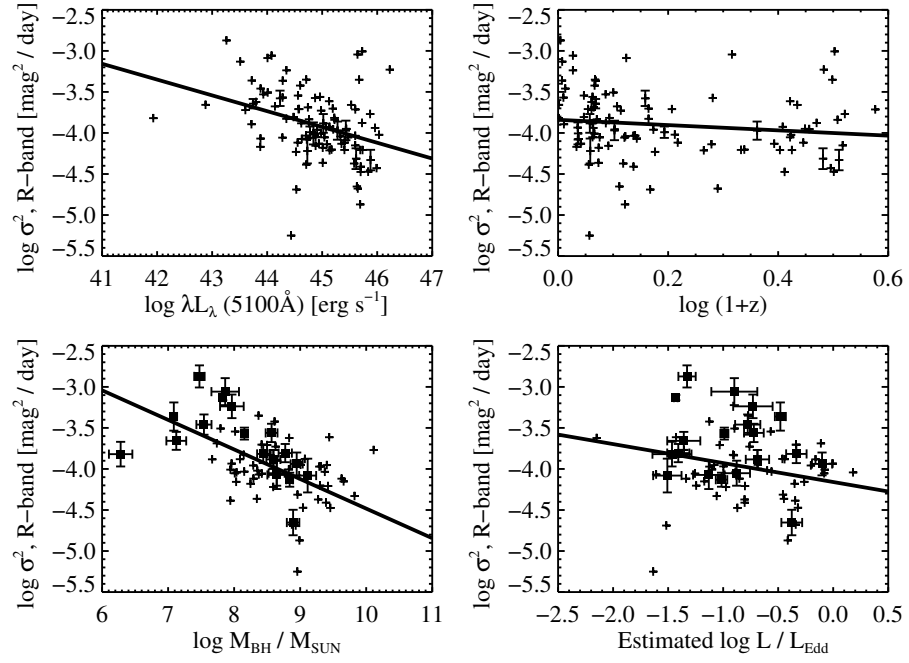
equal the true Eddington ratio. In Figure 8.6 we also show τ as a function of M_{BH} and L/L_{Edd}, and in Figure 8.7 we also show σ as a function of M_{BH} and L/L_{Edd}.

The results of the regressions for τ are

$$\log \tau \;\; = \;\; (-2.29 \pm 1.17) + (0.56 \pm 0.14) \log M_{BH} \;\; \text{[days]} \qquad (8.34)$$

$$\log \tau \;\; = \;\; (2.50 \pm 0.24) - (0.06 \pm 0.27) \log L/L_{bol} \;\; \text{[days]}. \qquad (8.35)$$

and the results of the regression for σ are

$$\log \sigma \;\; = \;\; (0.33 \pm 0.73) - (0.52 \pm 0.08) \log M_{BH} \;\; [\text{R mag}^2/\text{day}] \qquad (8.36)$$

$$\log \sigma \;\; = \;\; (-4.33 \pm 0.19) - (0.25 \pm 0.22) \log L/L_{Edd} \;\; [\text{R mag}^2/\text{day}]. \quad (8.37)$$

Based on these regression results, there is a statistically significant correlation between τ and M_{BH}, where the relaxation time increases with increasing M_{BH}. In addition, there is significant evidence that σ decreases with increasing M_{BH}. However, there is no evidence for a dependence of τ or σ on the Eddington ratio.

Due to the correlation between L and M_{BH}, it is unclear whether the observed dependency of τ and σ on these quantities is real for both L and M_{BH}, or whether one correlation is simply a reflection of the other. Similar to breaking the L–z degeneracy, we can investigate which correlation is the fundamental one, or if both are, by performing a multiple regression of τ and σ on L and M_{BH}. The results are

$$\log \sigma \;\; = \;\; (-3.83 \pm 0.17) - (0.09 \pm 0.19) \log \left(\frac{\lambda L_\lambda}{10^{45} \text{ erg s}^{-1}} \right)$$

$$- \;\; (0.25 \pm 0.24) \log \left(\frac{M_{BH}}{10^8 M_\odot} \right) \;\; [\text{R mag}^2/\text{day}] \qquad (8.38)$$

$$\tau \;\; = \;\; (80.4^{+66.9}_{-35.8}) \left(\frac{\lambda L_\lambda}{10^{45} \text{ erg s}^{-1}} \right)^{-0.42 \pm 0.28} \left(\frac{M_{BH}}{10^8 M_\odot} \right)^{1.03 \pm 0.38} \;\; \text{[days]} \;\; (8.39)$$

Here, we have expressed τ as a function of L and M_{BH} instead of $\log \tau$ for more direct comparison with the characteristic time scales described by Equations (8.9)–(8.12). The joint probability distributions of the slopes are shown in Figure 8.8

for both regressions. It is unclear whether σ depends on solely M_{BH}, solely L, or both M_{BH} and L, although the data favor a dependence on M_{BH} over one on L. However, there is significant evidence that the relaxation time scale depends on at least M_{BH}, and possibly on L as well. In fact, the dependence of τ on M_{BH} has steepened, and the relationship described by Equation (8.41) is similar to that for the orbital or thermal time scale, assuming a viscosity parameter of $\alpha \sim 10^{-3}$.

8.6   DISCUSSION

8.6.1   Comparison with Previous Work

Most previous work on quasar optical variability has been based on analysis of structure functions or power spectra, either of individual quasars or an ensemble of quasars. From these studies, an anti-correlation between variability and luminosity has often emerged (e.g., Hook et al., 1994; Garcia et al., 1999; Vanden Berk et al., 2004; Wilhite et al., 2008), while results on a variability–redshift correlation have remained mixed. Our result that long-term quasar variability is uncorrelated with luminosity may appear to be in conflict with previous work. However, the evidence for a variability–luminosity correlation is considerably weaker in the studies that have computed variability measures for individual objects. Indeed, studies that have compared variability with luminosity for individual objects have noted the signficant scatter in the relationship, producing a very weak correlation, often leading to a detection of 'moderate' statistical significance at best. Ensemble studies, on the other hand, cannot investigate the scatter in the relationship and therefore cannot assess the strength of the correlation. Instead, ensemble studies simply look for an average trend in variability properties, and, given enough quasars, are able to detect even a weak trend of variability with luminosity. Furthermore, quasars exhibit a range in characteristic time scale and
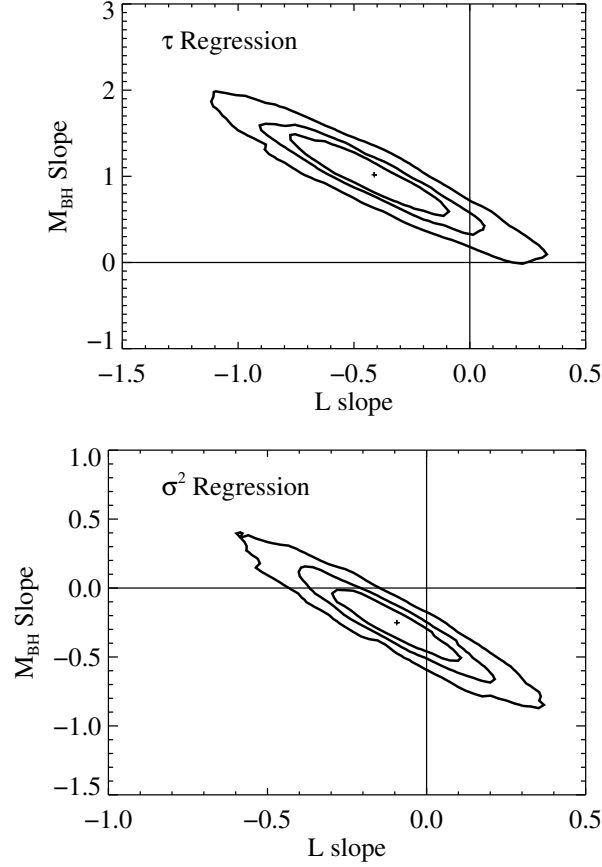
Figure 8.8 Probability distribution for the values of the coefficients in a linear regression of the characteristic time scale of quasar optical variations, τ, (top) and the magnitude of short time scale variations, σ, (bottom) as a function of M_{BH} and λL_λ (see Eq.[8.40] and [8.41]). The contours correspond to approximate $50\%, 75\%$, and 95% joint confidence regions. While there is significant evidence that τ depends on at least M_{BH}, it is unclear is there is an additional dependence on luminosity. In addition, while it is clear that σ depends on either M_{BH} or λL_λ, it is unclear whether the dependency is on M_{BH}, λL_λ, or both.

variability amplitude at a given luminosity or black hole mass, and it is unclear how this affects an ensemble structure function or power spectrum.

Recently Wold et al. (2007) have reported a correlation between optical variability and M_{BH} on time scales $t > 100$ days, but no correlation is seen on shorter time scales. In contrast, we observed an anti-correlation between M_{BH} and short time scale variability, but no correlation between M_{BH} and long time scale variability. The Wold et al. (2007) result is unexpected, since M_{BH} and L are correlated (e.g., Peterson et al., 2004), and previous studies have found that variability is anti-correlated with L, even on long time scales. The Wold et al. (2007) result is based on an ensemble structure function, and the uncertainties on the structure function for the high M_{BH} bins are large. Furthermore, the time sampling of the Wold et al. (2007) sample for the high M_{BH} bins is worse than for the low M_{BH} bins, and windowing effects due to the finite length of the time series may be at work here. When considering variability measurement of individual sources, Wold et al. (2007) still find a positive correlation between variability and M_{BH}. However, while this correlation is statistically significant, it is very weak and exhibits considerable scatter. We performed a Kendall and Spearman rank correlation test between long time scale variability and the estimated black hole mass, and also find a marginally significant correlation, but the significance disappeared when we accounted for the uncertainty in the mass estimates.

Collier & Peterson (2001) calculated structure functions of optical light curves for 12 low-z AGN with reverberation mapping data. Consistent with our work, they find a correlation between M_{BH} and characteristic time scale, where a characteristic time scale was defined as the location of a break in the structure function. The time scales found by Collier & Peterson (2001) were consistent with dynamical or disk thermal time scales, although they tended to be somewhat

shorter than those derived in this work. This difference may be explained by somewhat different definitions of 'characteristic' time scale between their work and our, and systematic errors in the estimated structure functions caused by finite time sampling effects.

8.6.2 Connection with Accretion Physics

In this work we have found that both the characteristic time scale of quasar light curves, and the magnitude of the short time scale variations, depend on black hole mass. This, in combination with the evidence from reverberation mapping, strongly argues that the source of quasar optical variability is intrinsic to the accretion disk. In addition, we have found that the characteristic time scales of quasar light curves are similar to what would be expected for disk dynamical or thermal time scales, assuming a viscosity parameter of $\alpha \sim 10^{-3}$. Recent MHD simulations have radiation-dominated AGN accretion disks have found that the thermal time scale is shorter than that implied by the standard α-prescription (Turner, 2004), making the association of τ with thermal time scales more consistent. This implies that on time scales shorter than t_{orb} or t_{th}, the accretion disk has difficulty generating variations in optical flux in response to random variations of some input process, such as, for example, a time varying magnetic field. Instead, these short time scale variations get 'smoothed out', creating $1/f^2$ power spectrum for frequencies higher than $\sim 1/t_{orb}$ or $\sim 1/t_{th}$. While the quasar characteristic time scales are consistent with both orbital and thermal time scales, we find it more appropriate to associate these time scales with t_{th}, as we would expect some sort of periodic activity in the light curves if the flux variations were driven by orbital motion. In addition, quasar optical emission is thought to be thermal emission from an optically thick accretion disk (e.g., Krolik, 1999a; Frank, King, & Raine, 2002), and quasars tend to be bluer as they brighten (e.g., Giveon et al.,

1999; Trèvese et al., 2001), suggesting that the flux variations are due to thermal variations.

In order to interpret the CAR(1) process in terms of accretion disk physics, we rewrite Equation (8.2) as

$$d \log L(t) = -\frac{1}{\tau}(\log L(t) - \mu)dt + \sigma[B(t + dt) - B(t)]. \qquad (8.40)$$

Here, $L(t)$ denotes the luminosity of the quasar at time t, $\mu = \tau b$ is the mean value of the quasar light curve, and $B(t)$ denotes Brownian motion. In Equation (8.42) we have used the fact that the derivative of Brownian motion is white noise, i.e., $\epsilon(t) = dB(t) = B(t+dt) - B(t)$. Brownian motion is a non-stationary random walk process that has a power spectrum $P(f) \propto 1/f^2$, and is described by Equation (8.2) in the limit $\tau \to \infty$. In addition, Equation (8.42) implicitly assumes that the variance in the random variable $dB(t) = B(t + dt) - B(t)$ is $Var(dB(t)) = dt$.

Writing the Equation for a CAR(1) process as Equation (8.42) reveals a number of interesting properties of this process. First, we note that the first term on the right side is what keeps the time series stationary. Considering only this term (i.e., $\sigma = 0$), $d \log L(t)/dt$ is negative when the value of $L(t)$ is brighter than the mean, and $d \log L(t)/dt$ is positive when $L(t)$ is fainter than the mean. Therefore, the first term on the right side stabilizes the process by always driving $L(t)$ toward its mean value, while the second term generates random perturbations to $d \log L(t)/dt$ that cause $L(t)$ to deviate from its expected path. For highly accreting objects like quasars, the accretion disks are expected to be radiation pressure dominated. Under the standard α-prescription for the viscosity, where the viscous torque is assumed to be proportional to the total pressure, a radiation pressure dominated disk is unstable to perturbations in the heating rate (e.g., Shakura & Sunyaev, 1976; Krolik, 1999a). The fact that the quasar light curves in our sample are described well by a CAR(1) process with relaxation times similar to disk

thermal time scales rules out instabilities in the disk that grow as $\sim t_{th}$, consistent with results obtained from full MHD simulations (e.g., Turner, 2004).

From Equation (8.42) it is apparent that the stochastic input into the differential equation, which drives the random variations in L, is itself a stochastic process that should resemble brownian motion. In particular, a random deviation in the input process, $B(t)$, over a time interval dt causes a random perturbation to the change in $\log L(t)$ expected over the interval dt, with σ controlling how sensitive $d \log L(t)/dt$ is to $dB(t)/dt$. It is not essential that the mechanism driving the stochastic variations be Brownian motion, but rather that the process 'looks' like Brownian motion on time scales much longer than the relaxation time scale of Equation (8.42). A CAR(1) process satisfies this requirement so long as the relaxation time for the input process is long compared to the relaxation time of the quasar light curve. For example, the input process could be due to variations in accretion rate, or perturbations caused by a time-varying magnetic field. Indeed, recently some authors have modelled quasar variability as being driven by variations in a magnetic field, with the magnetic field density being modelled as an AR(1) process (e.g., King et al., 2004; Mayer & Pringle, 2006; Janiuk & Czerny, 2007).

If the variations in $L(t)$ are driven by variations in the accretion rate, $\dot{M}(t)$, then the accretion rate variations will propagate on a viscous time scale, t_v, with the shorter time scale variations being smoothed out. Because the viscous time scale is much longer than the relaxation time scales of our sample, the variations in accretion rate will look like brownian motion on time scales $t \ll t_v$ if the power spectrum of $\dot{M}(t)$ is $\sim 1/f^2$ on time scales smaller than the viscous time scale. A CAR(1) process for $\dot{M}(t)$ with $\tau_{\dot{M}} = t_v$ satisfies this requirement, but other processes may as well. In this case, while the random short time scale variations

in $\dot{M}$ are damped, they still exist. If the relaxation time of quasar light curves is associated with the thermal time scale, this therefore implies that the damped short time scale variations in $\dot{M}$ randomly alter the heat content of the disk. The random variations in heat content then create random variations in $L(t)$, but because the disk cannot react to changes in heat content on time scales less than the thermal time scale, the shorter time scale variation in flux are smoothed out. In this sense, short time scale variations are correlated because the disk has not had time to completely react to the change in heat content. However, on time scales $t \gtrsim t_{th}$, the disk has had time to adjust to the heat content variations, thereby 'forgetting' about the previous perturbations in heat content. The result is a red noise power spectrum on time scales $t \lesssim t_{th}$, and a white noise power spectrum on time scales $t \gtrsim t_{th}$. A similar argument applies if a time-varying magnetic field drives the random variations in flux, so long as the characteristic time scale of the magnetic field is long compared to the relaxation time scale of the quasar light curve. Whatever the physical process is that can be associated with the stochastic input into Equation (8.42), the anti-correlation between σ and M_{BH} we have found implies that the magnitude of the random variations in $d \log L(t)/dt$ produced by the stochastic input process must decrease as M_{BH} increases.

Modelling quasar variability as a stochastic process provides an opportunity to unify both short and long time scale variability as the result of a single process. The source of variations in optical luminosity over time scales of hours (so-called 'microvariability' or 'intranight variability') has remained a puzzle, although reprocessing of X-rays or a weak blazar component have been suggested (Czerny et al., 2008). In general, microvariability in radio-quiet quasars is not detected above the photometric uncertainty (e.g., Gupta & Joshi, 2005; Carini et al., 2007); however, for those sources for which it is detected the standard deviation in the

variability over the course of a night is ~ 0.01 mag (Gopal-Krishna et al., 2003; Stalin et al., 2004, 2005; Gupta & Joshi, 2005). As noted above, for a CAR(1) process the standard deviation of short time scale variations is $\approx \sigma\sqrt{\Delta t}$. As can be seen from figure 8.5, assuming a CAR(1) process for quasar light curves predicts variations of ~ 0.01 mag over the course of a night, consistent with has has been observed. Therefore, it is not necessary to invoke an additional physical mechanism to explain the short time scale variations, as the CAR(1) process is able to explain both short and long time scale variations as being driven by the same process. Furthermore, microvariability is known to be stronger in radio-loud quasars (e.g., Gupta & Joshi, 2005). If a single underlying stochastic process drives both short and long time scale variability, then the increased microvariability of radio-loud objects may be a clue in understanding jet formation. In particular, if a time-varying magnetic field is what drives the optical variations, then the increased microvariability in radio-loud objects may be due to a more turbulent magnetic field, which may make it easier for these objects to form powerful outflows or jets.

8.7   SUMMARY

In this work we have modelled quasar light curves as a type of stochastic process called a first-order continuous autoregressive process. This statistical model has three free parameters: the characteristic time scale for the process to 'forget' about itself, τ, the magnitude of the small time scale variations, σ, and the mean of the time series, μ. We used this model to fit 100 quasar light curves at $z < 2.8$, including 70 quasars with black hole mass estimates. Our conclusions are summarized as follows:

- Quasar optical light curves are often well described by a continuous au-

toregressive process (CAR(1)). For the quasars in our sample, the long time scale variations are typically $\sim 3\text{--}30\%$, consistent with previous work. The characteristic time scales of the quasar light curves vary between ~ 10 days and ~ 10 yrs and are consistent with accretion disk orbital or thermal time scales, assuming a viscosity parameter of $\alpha \sim 10^{-3}$. In addition, the short time scale variations are $\lesssim 0.02$ mag day$^{-1/2}$.

- The characteristic time scales of quasar optical light curves are correlated with M_{BH} and luminosity, while the magnitude of the short time scale variations are anti-correlated with M_{BH} and luminosity. We did not find any evidence for an additional redshift correlation. A multiple regression analysis suggested that the primary correlation is with M_{BH}. At a given luminosity, the characteristic time scales depend on M_{BH} as

$$\tau = (80.4^{+66.9}_{-35.8}) \left(\frac{\lambda L_\lambda}{10^{45} \text{ erg s}^{-1}} \right)^{-0.42\pm0.28} \left(\frac{M_{BH}}{10^8 M_\odot} \right)^{1.03\pm0.38} \quad \text{[days]} \quad (8.41)$$

where the errors are quoted at 68% confidence (1σ).

- For the CAR(1) process, the random perturbations to $d \log L(t)/dt$ are caused by a time varying stochastic process that is well described by a power spectrum $P(f) \propto 1/f^2$ on time scales long compared to the characteristic time scale of the quasar light curves. This stochastic input process may be a time-varying magnetic field or accretion rate variations that propagate on a viscous time scale. Variations in the input process over an interval dt create variations in $L(t)$ which are smoothed out on an orbital or thermal time scale.

- The fact that quasar optical light curves can be well fit by a CAR(1) model suggest that it is not necessary to invoke an an additional physical mechanism to describe short time scale variations. Instead, within the CAR(1)

model, both short and long time scale variations are driven by an under-
lying stochastic process that causes random white noise perturbations to
$d \log L(t)/dt$. This is supported by the fact that the intranight variations pre-
dicted by our best fit CAR(1) processes are ~ 0.01 mag in R, consistent with
studies of intranight variability.

Before concluding, we stress that the CAR(1) model is a statistical model and
not a physical model. Quasar light curves are stochastic in nature, and the de-
pendence of luminosity on the time-varying properties of the disk is complex.
In this sense, the randomness in the stochastic model is not due to that fact that
the physical processes themselves are not deterministic, but rather is a reflection
of our lack of knowledge of the complex physical processes that generate varia-
tions in flux. While a physical model is needed in order to interpret the stochastic
model in terms of accretion disk physics, and thus lead to a proper understand-
ing of quasar light curves, the stochastic model is sufficient for modeling the data,
given our current knowledge. Furthermore, much of the mathematical formalism
of accretion physics is in the language of differential equations, suggesting that
stochastic differential equations are a natural choice for modeling quasar light
curves.

The field of stochastic processes is a rich field with well-developed method-
ology, predominantly because of its important in financial and economic model-
ing. We have utilized the CAR(1) model because of its simplicity, and because
it allows us to perform statistical inference without having our results biased by
the irregular sampling, measurement errors, and finite span of the time series.
However, the CAR(1) model is the simplest of stationary continuous autoregres-
sive processes, and additional flexibility may be achieved through the addition
of higher order derivatives to Equation (8.2). This provides a rich and flexible

method of modeling the power spectra of quasar light curves without suffering from the windowing effects that can bias traditional Fourier and structure function techniques. For example, quasi-periodic oscillations can be modeled through the addition of second order derivatives to Equation (8.2), as has been done in the analysis of the frequency of sun spot numbers (Phadke & Wu, 1974). In addition, Equation (8.2) can be generalized to a vector form, allowing the simultaneous modelling of quasar light curves accross multiple observing wavelengths, and thus introducing additional constraints on physical models of quasar variability. Both the use of higher order terms and multiwavelength modelling of quasar light curves will be the subject of future research.

CHAPTER 9

CONCLUDING REMARKS AND FUTURE WORK

In this thesis I have attempted to further our understanding of the structure of the quasar central engine through investigating trends in quasar SEDs and lightcurves with luminosity, redshift, black hole mass, and Eddington ratio. In addition, I have developed new statistical methods for dealing with the significant statistical uncertainty in estimates of M_{BH} based on the broad emission lines, as well as correcting for the presence of a selection function and non-detections. I used these new statistical methods to analyze correlations involving M_{BH} and the ratio of optical/UV flux or spectral slope of the X-ray continuum (see Chapter 7), and the characteristic time scale and variability of quasar optical lightcurves (see Chapter 8). In addition, I also use my new methods to estimate the $z < 0.5$ SMBH mass function for broad line AGN, based on quasars from the Bright Quasar Survey. Among my results, the most noteworthy are:

- The statistical uncertainty in the broad line mass estimates can lead to significant artificial broadening of the inferred distribution of M_{BH}. This therefore causes one to overestimate the number of AGN with large values of M_{BH} and small values of M_{BH}, and underestimate the number of AGN with average values of M_{BH}. In addition, the large statistical uncertainties in M_{BH} 'blur out' correlations involving M_{BH}, reducing their observed magnitude and statistical significance. It is important to correct for the statistical uncertainty in M_{BH} estimates in order to infer the intrsinic distributions and correlations.

- The $z = 0.2$ broad line quasar black hole mass function (BHMF) falls off

approximately as a power law with slope ~ 2 for $M_{BH} \gtrsim 10^8 M_\odot$. There is some evidence for evolution in the local quasar BHMF, with the BHMF shifting toward higher values of M_{BH} as z increases, and possibly flattening at the high M_{BH} end at $z \gtrsim 0.3$. In addition, my statistical model implies that at a given M_{BH}, $z < 0.5$ broad line quasars have a typical Eddington ratio of $L/L_{Edd} \sim 0.4$, with a dispersion of $\lesssim 0.5$ dex. However, these estimates are based on the assumption of a Gaussian distribution in L/L_{Edd}, and the estimated dispersion in L/L_{Edd} likely biased toward smaller values if the true distribution of L/L_{Edd} exhibits significant skew toward smaller values.

- There is evidence for a correlation between $\alpha_{\rm ox}$ and optical/UV luminosity, redshift, M_{BH}, and Eddington ratio. These correlations imply that radio-quiet quasars become more X-ray quiet as their L_{UV}, M_{BH}, and L/L_{Edd} increase, and become more X-ray loud as z increases at a given L_{UV}. The correlations involving L_{UV} and z are most likely driven by correlations involving M_{BH} and L/L_{Edd}. These correlations imply that radio-quiet quasars emit a larger fraction of their bolometric luminosity through the accretion disk component, as compared to the corona component, as M_{BH} and $\dot{m}$ increase. In addition, because radio-quiet quasars become more X-ray quiet as M_{BH} and $\dot{m}$ increase, these correlations imply that quasars become more efficient at injecting feedback energy into the ambient gas as M_{BH} and $\dot{m}$ increase, and are thus able to drive more powerful outflows.

- There is no evidence for a correlation between the spectral slope of the X-ray continuum, Γ_X, and luminosity or redshift. However, there is evidence for a significant trend between Γ_X and L/L_{Edd}. In addition, the trend between

Γ_X and L/L_{Edd} is non-monotonic, where the X-ray continuum softens with increasing Eddington ratio until $L/L_{Edd} \sim 0.3$, and then hardens with increasing Eddington ratio. A similar trend has been seen between the spectral slope of the optical/UV continuum and Eddington ratio (Bonning et al., 2007; Davis et al., 2007), and this non-monotonic behavior may be caused by a change in the structure of the accretion disk at $L/L_{Edd} \sim 0.3$, where the inner part of the accretion disk becomes thicker due to increased radiation pressure (Abramowicz et al., 1988).

- Many quasar optical light curves are well describes for a first order continuous autoregressive process (CAR(1)). The CAR(1) process is a type of stochastic process with a $1/f^2$ power spectra on time scales short compared to some characteristic time scale, τ, and a flat power spectra on time scales long compared to τ. I found significant evidence that the characteristic time scales of quasar optical flux variations increases with increasing M_{BH}, and that the amplitude of short time scale variability decreases with increasing M_{BH}. Both the magnitude of our estimated characteristic time scales and the trend with M_{BH} are consistent with disk orbital or thermal time scales. I interpret quasar optical light curves as being driven by thermal fluctuations, which in turn are driven by some other underlying stochastic process with characteristic time scale long compared to the disk thermal time scale. In addition, the CAR(1) stochastic model is able to explain both long and short time scale variations, thus eliminating the need to invoke seperate physical mechanisms to explain both.

There are a few natural directions for future research that build on the conclusions and methods of this thesis. First, understanding how and when SMBHs grow is currently of central importance in extragalactic astronomy. Currently,

it is unclear to what degree mergers of gas rich galaxies are important, as opposed to quiescent accretion of ambient gas. At this time, the rough picture that is emerging is that gas rich mergers are important for growing the most massive black holes at high redshift, where bright quasars dominate the AGN population, while lower mass black holes (e.g., Seyfert galaxies) grow more slowly through accretion of ambient gas, such as cold molecular clouds or the hotter IGM.

The broad line quasar BHMF offers a step forward in understanding black hole growth. By comparing the local quiescent BHMF with the integrated broad line quasar BHMF, we can place constraints on how much black hole growth occurs in broad line AGN. In addition, the local quiescent BHMF can be seperated into the quiescent BHMF for early and late type galaxies, where early type galaxies dominate the local BHMF at $M_{BH} \gtrsim 10^7 M_\odot$, and late type galaxies dominate at $M_{BH} \lesssim 10^7 M_\odot$ (e.g., Shankar et al., 2004). If the largest SMBHs grow during bright quasar phases, then we would expect that the integrated broad line AGN BHMF should roughly match the local BHMF for early type galaxies, modulo the additional SMBH growth that occurs in these AGN before entering quiescence. Any systematic differences between the two mass functions may also shed light on the importance of obscured (Type II) AGN for black hole growth, and whether the difference between obscured and unobscured AGN is an evolutionary or orientation effect. However, in order to accomplish these tasks, it is necessary to properly correct for a selection function and the large statistical uncertainty in broad line mass estimates, as was done in Chapter 5 of this thesis.

Using the broad line AGN BHMF to help understand how the local quiescent BHMF was built up also has important implications for methods that attempt to map black hole growth at $z > 0$ based on the argument of Soltan (1982). These methods assume that the local BHMF was built up by accretion during quasar

phases. Under this assumption, black hole growth is mapped by stepping backwards from the local BHMF, using the quasar luminosity function as a constraint (e.g., Yu & Tremaine, 2002; Marconi et al., 2004); however if this assumption is shown to be only partially true, than this will imply that mapping black hole growth is more complicated, and some modifications must be made to this technique. Furthermore, is calls into question the accuracy of estimated quasar accretion efficiencies derived from these methods.

An additional avenue for further research is to apply the methods of Chapter 8 to modeling the X-ray lightcurves of quasars. X-ray variability studies have been dominated by spectral analysis, where parameteric forms are often fit to the empirical power spectra obtained from a variety of missions that sample X-ray light curves at various time scales (e.g., Markowitz et al., 2003; Markowitz & Edelson, 2004; McHardy et al., 2004; Uttley et al., 2002). These studies have found that the X-ray power spectra can be modeled as a broken power-law, where the power spectrum is $P(f) \propto 1/f^2$ on time scales shorter than some break frequency, and $P(f) \propto 1/f$ on time scales longer than the break frequency. Most interestingly, the break frequency is inversely proportional to M_{BH}.

The existence of long time scale variations has been a problem for interpreting AGN X-ray variations. In particular, it is unclear how long time scale variations, presumably caused by processes in the accretin disk outside of the X-ray emitting region, can survive in the X-ray lightcurve. As a result, X-ray variability has most often been interpreted in terms of models of inwardly-propagating accretion rate fluctuations (e.g., Lyubarskii, 1997; King et al., 2004; Mayer & Pringle, 2006). Within the context of these models, perturbations in the accretion rate occur in the outer edges of the accretion disk and then propagate inward. Upon reaching the X-ray emitting region, they modulate the X-ray flux, creating a $1/f$

power spectrum. The break in the X-ray power spectrum is interpreted as corresponding to a size of the X-ray emitting regin, where variations get 'smoothed out' on time scales shorter than the characteristic time scale of the flux from the X-ray emitting region. The anti-correlation of the break frequency with M_{BH} has been interpreted as evidence that the size of the X-ray emitting region is correlated with M_{BH}, with the X-ray emitting region becoming larger with increasing M_{BH}.

Similar to the case of optical variability studies, spectral analysis has only been applied to X-ray lightcurves of a handful of low redshift Seyfert 1 galaxies. This is largely because of the difficulty in obtaining enough X-ray photons to create a binned lightcurve with both adequate time sampling and Gaussian error bars. However, the methods of Chapter 8, where the quasar lightcurve is directly modeled as a stochastic process, can be applied to Poisson distributed data (i.e., photon counts) as well (e.g., Browkwell & Davis, 2002). This therefore opens up the possibility of analyzing X-ray lightcurves for a large sample of quasars over a broad range in luminosity, redshift, and black hole mass. Furthermore, modeling the lightcurves directly as a stochastic process has the advantage that the analysis is not biased by finite time sampling effects, and that quantities of interest, such as characteristic time scales, are estimated directly from the observed lightcurve, as opposed to indirectly from the empirical power spectrum. Analysis of X-ray lightcurves for a large sample of quasars will shed light on the physical source of the X-ray fluctuations, and will enable further constraints on the disk/corona structure.

Appendix A

MAXIMUM-LIKELIHOOD VS BAYESIAN INFERENCE

In this section we compare the maximum likelihood approach with the Bayesian approach. We do this for readers who are unfamiliar with some of the more technical aspects of the two approaches, with the hope that the discussion in this section will facilitate interpretation of our results in the main body of the paper.

In maximum-likelihood analysis, one is interested in finding the estimate that maximizes the likelihood function of the data. For a given statistical model, parameterized by θ, the likelihood function, $p(x|\theta)$, is the probability of observing the data, denoted by x, as a function of the parameters θ. The maximum-likelihood estimate, denoted by $\hat{\theta}$, is the value of θ that maximizes $p(x|\theta)$. Under certain regularity conditions, $\hat{\theta}$ enjoys a number of useful properties. In particular, as the sample size becomes infinite, $\hat{\theta}$ becomes an unbiased estimate of θ. An unbiased estimator is an estimator with expectation equal to the true value, i.e., $E(\hat{\theta}) = \theta_0$, where θ_0 is the true value of θ. Therefore, on average, an unbiased estimator will give the true value of the parameter to be estimated.

Because the maximum likelihood estimate is a function of the data, $\hat{\theta}$ has a sampling distribution. The sampling distribution of $\hat{\theta}$ is the distribution of $\hat{\theta}$ under repeated sampling from the probability distribution of the data. Under certain regularity conditions, the sampling distribution of $\hat{\theta}$ is asymptotically normal with covariance matrix equal to the average value of the inverse of the Fisher information matrix, $I(\theta)$, evaluated at θ_0. The Fisher information matrix is the expected value of the matrix of second derivatives of the log-likelihood, multiplied by -1. Formally, this result states that as $n \to \infty$, then

$$\hat{\theta} \quad \sim \quad N_p(\theta_0, I^{-1}(\theta_0)), \tag{A.1}$$

$$I(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2}\ln p(x|\theta)\right), \tag{A.2}$$

where p is the number of parameters in the model, and the expectation in Equation (A.2) is taken with respect to the sampling distribution of x, $p(x|\theta_0)$. Because we do not know θ_0, it is common to estimate $I^{-1}(\theta_0)$ by $I^{-1}(\hat{\theta})$. In addition, it is common to estimate $I(\theta)$ as the matrix of second derivatives of the log-likelihood of one's data, since the sample average is a consistent estimate for the expectation value. Qualitatively, Equation(A.1) states that as the sample size becomes large, $\hat{\theta}$ approximately follows a normal distribution with mean θ_0 and covariance matrix $I^{-1}(\hat{\theta})$. This fact may be used to construct confidence intervals for θ.

While the asymptotic results are useful, it is not always clear how large of a sample is needed until Equation (A.1) is approximately true. The maximum likelihood estimate can be slow to converge for models with many parameters, or if most of the data is missing. Within the context of luminosity function estimation, the maximum-likelihood estimate will be slower to converge for surveys with shallower flux limits. In addition, Equation (A.1) does not hold if the regularity conditions are not met. In general, this is not a concern, but it is worth noting that the asymptotics do not hold if the true value of θ lies on the boundary of the parameter space. For example, in the case of a Schechter luminosity function, if the true value of the shape parameter, α (see [4.14]), is $\alpha_0 = -1$, then Equation (A.1) does not hold, since $\alpha > -1$. If $\alpha_0 \approx -1$, then Equation (A.1) is still valid, but it will take a large sample before the asymptotics are valid, as α_0 lies near the boundary of the parameter space.

In Bayesian analysis, one attempts to estimate the probability distribution of the model parameters, θ, given the observed data x. The probability distribution

of θ given x is related to the likelihood function as

$$p(\theta|x) \propto p(x|\theta)p(\theta). \tag{A.3}$$

The term $p(x|\theta)$ is the likelihood function of the data, and the term $p(\theta)$ is the prior probability distribution of θ; the result, $p(\theta|x)$ is called the posterior distribution. The prior distribution, $p(\theta)$, should convey information known prior to the analysis. In general, the prior distribution should be constructed to ensure that the posterior distribution integrates to one, but to not have a significant effect on the posterior. In particular, the posterior distribution should not be sensitive to the choice of prior distribution, unless the prior distribution is constructed with the purpose of placing constraints on the posterior distribution that are not conveyed by the data. The contribution of the prior to $p(\theta|x)$ becomes negligible as the sample size becomes large.

From a practical standpoint, the primary difference between the maximum likelihood approach and the Bayesian approach is that the maximum likelihood approach is concerned with calculating a point estimate of θ, while the Bayesian approach is concerned with mapping out the distribution of θ. The maximum likelihood approach uses an estimate of the sampling distribution of $\hat{\theta}$ to place constraints on the true value of θ. In contrast, the Bayesian approach directly calculates the probability distribution of θ, given the observed data, to place constraints on the true value of θ. It is illustrative to consider the case when the prior is taken to be uniform over θ; assuming the posterior integrates to one, the posterior is then proportional to the likelihood function, $p(\theta|x) \propto p(x|\theta)$. In this case, the goal of maximum likelihood is to calculate an estimate of θ, where the estimate is the most probable value of θ, given the observed data. Then, confidence intervals on θ are derived from the maximum likelihood estimate, $\hat{\theta}$, usually by assuming Equation (A.1). In contrast, the Bayesian approach is not concerned

with optimizing the likelihood function, but rather is concerned with mapping out the likelihood function. Under the Bayesian approach with a uniform prior, confidence intervals on θ are derived directly from likelihood function, and an estimate of θ can be defined as, for example, the value of θ averaged over the likelihood function. So, the maximum likelihood attempts to obtain the 'most likely' value of θ, while the Bayesian approach attempts to directly obtain the probability distribution of θ, given the observed data. Because the Bayesian approach directly estimates the probability distribution of θ, and because it does not rely on any asymptotic results, we consider the Bayesian approach to be preferable for most astronomical applications.

APPENDIX B

DERIVATION OF THE MARGINAL POSTERIOR DISTRIBUTION FOR

TRUNCATED DATA

Here, we give a derivation of the posterior probability distribution of θ, given by Equation (4.12). If we assume a uniform prior on $\log N$, then this is equivalent to assuming the prior $p(\theta, N) \propto N^{-1}p(\theta)$. In this case, the posterior distribution is given by

$$p(\theta, N|L_{obs}, z_{obs}) \propto N^{-1}p(\theta)C_n^N \left[p(I = 0|\theta)\right]^{N-n} \prod_{i\in\mathcal{A}_{obs}} p(L_i, z_i|\theta). \qquad \text{(B.1)}$$

The marginal posterior distribution of θ is obtained by summing the joint posterior over all possible values of N. For the choice of prior $p(\theta, \log N) \propto p(\theta)$, the marginal posterior of θ is

$$p(\theta|L_{obs}, z_{obs}) \quad \propto \quad p(\theta) \left[\prod_{i\in\mathcal{A}_{obs}} p(L_i, z_i|\theta)\right] \sum_{N=n}^{\infty} N^{-1}C_n^N \left[p(I = 0|\theta)\right]^{N-n} \qquad \text{(B.2)}$$

$$\propto \quad p(\theta) \left[p(I = 1|\theta)\right]^{-n} \left[\prod_{i\in\mathcal{A}_{obs}} p(L_i, z_i|\theta)\right]$$

$$\times \quad \sum_{N=n}^{\infty} C_{n-1}^{N-1} \left[p(I = 0|\theta)\right]^{N-n} \left[p(I = 1|\theta)\right]^{n}, \qquad \text{(B.3)}$$

where we arrived at the second Equation by multiplying and dividing the first Equation by $p(I = 1|\theta)^n$ and noting that $C_n^N = C_{n-1}^{N-1}(N/n)$. The term within the sum is the mathematical expression for a negative binomial distribution as a function of N (see Eq.[C.1]). Because probability distributions must be equal to unity when summed over all possible values, the sum is just equal to one. We therefore arrive at Equation (4.12) by replacing the summation in Equation (B.3) with the value of one.

APPENDIX C

SOME PROBABILITY DISTRIBUTIONS USED IN THIS WORK

In this chapter of the appendix we briefly describe some probability distribution that we employ, but may be unfamiliar to some astronomers.

C.1   Negative Binomial

The negative binomial distribution is closely related to the binomial distribution. The binomial distribution gives the probability of observing n 'successes', given that there have been N trials and that the probability of success is p. In contrast, the negative binomial distribution gives the probability of needing N trials before observing n successes, given that the probability of success is p. Within the context of this work, the binomial distribution gives the probability of detecting n sources, given that there are N total sources and that the detection probability is p. The negative binomial distribution gives the probability that the total number of sources is N, given that we have detected n sources and that the detection probability is p. The negative binomial distribution is given by

$$p(N|n,p) = C_{n-1}^{N-1} p^n (1-p)^{N-n}, \quad N \geq n. \tag{C.1}$$

A random draw from the negative binomial distribution with parameters n and p may be simulated by first drawing n random values uniformly distributed on $[0,1]$, $u_1, \ldots, u_n \sim \mathrm{Uniform}(0,1)$. Then, calculate the quantity

$$m = \sum_{i=1}^{n} \left\lfloor \frac{\log u_i}{\log(1-p)} \right\rfloor, \tag{C.2}$$

where $\lfloor \cdot \rfloor$ is the floor function, i.e., $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x. The quantity $N = n + m$ will then follow a negative binomial distribution with parameters n and p.

C.2 Dirichlet

The Dirichlet distribution is a multivariate generalization of the Beta distribution, and it is commonly used when modelling group proportions. Dirichlet random variables are constrained to be positive and sum to one. The Dirichlet distribution with argument $\theta_1, \ldots, \theta_k$ and parameters $\alpha_1, \ldots, \alpha_k$ is given by

$$p(\theta_1, \ldots, \theta_k | \alpha_1, \ldots, \alpha_k) = \frac{\Gamma(\alpha_1 + \ldots + \alpha_k)}{\Gamma\alpha_1 \cdots \Gamma\alpha_k} \prod_{i=1}^{k} \theta_i^{\alpha_i - 1}, \ \ \theta_1, \ldots, \theta_k \geq 0, \ \alpha_1, \ldots, \alpha_k > 0, \ \sum_{i=1}^{k} \theta_i = 1. \tag{C.3}$$

To draw a random value $\theta_1, \ldots, \theta_k$ from a Dirichlet distribution with parameters $\alpha_1, \ldots, \alpha_k$, first draw $x_1, \ldots, x_k$ independently from Gamma distributions with shape parameters $\alpha_1, \ldots, \alpha_k$ and common scale parameter equal to one. Then, set $\theta_j = x_j / \sum_{i=1}^{k} x_i$. The set of θ will then follow a Dirichlet distribution.

C.3 Multivariate Student-t and Cauchy Distribution

The Student-t distribution is often used as a robust alternative to the normal distribution because it is more heavily tailed than the normal distribution, and therefore reduces the effect of outliers on statistical analysis. A t distribution with $\nu = 1$ degree of freedom is referred to as a Cauchy distribution, and it is functionally equivalent to a Lorentzian function. A p-dimensional multivariate t distribution with p-dimensional argument $\mathbf{x}$, p-dimensional mean vector μ, $p \times p$ scale matrix Σ, and degrees of freedom ν is given by

$$p(\mathbf{x}|\mu, \Sigma, \nu) = \frac{\Gamma((\nu + p)/2)}{\Gamma(\nu/2)\nu^{p/2}\pi^{p/2}} |\Sigma|^{-1/2} \left[1 + \frac{1}{\nu} (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \right]^{-(\nu+p)/2}. \tag{C.4}$$

The 1-dimensional t distribution is obtained by replacing matrix and vector operations in Equation (C.4) with scalar operations.

Although we do not simulate from a t distribution in this work, for completeness we include how to do so. To simulate a random vector $\mathbf{t}$ from a multivariate

t distribution with mean vector μ, scale matrix Σ, and degrees of freedom ν, first draw $\mathbf{z}$ from a zero mean multivariate normal distribution with covariance matrix Σ. Then, draw x from a chi-square distribution with ν degrees of freedom, and compute the quantity $\mathbf{t} = \mu + \mathbf{z}\sqrt{\nu/x}$. The quantity $\mathbf{t}$ is then distributed according to the multivariate t distribution.

C.4 Wishart and Inverse Wishart

The Wishart distribution describes the distribution of the $p \times p$ sample covariance matrix, given the $p \times p$ population covariance matrix, for data drawn from a multivariate normal distribution. Conversely, the inverse Wishart distribution describes the distribution of the population covariance matrix, given the sample covariance matrix, when the data are drawn from a multivariate normal distribution. The Wishart distribution can be thought of as a multivariate extension of the χ^2 distribution. A Wishart distribution with $p \times p$ argument S, $p \times p$ scale matrix Σ, and degrees of freedom ν is given by

$$p(S|\Sigma, \nu) = \left[2^{\nu p/2} \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma\left(\frac{\nu + 1 - i}{2} \right) \right]^{-1} |\Sigma|^{-\nu/2} |S|^{(\nu-p-1)/2} \exp\left\{ -\frac{1}{2} tr(\Sigma^{-1}S) \right\},$$

(C.5)

where the matrices S and Σ are constrained to be positive definite. An inverse Wishart distribution with $p \times p$ argument Σ, $p \times p$ scale matrix S, and degrees of freedom ν is

$$p(\Sigma|S, \nu) = \left[2^{\nu p/2} \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma\left(\frac{\nu + 1 - i}{2} \right) \right]^{-1} |S|^{\nu/2} |\Sigma|^{-(\nu+p+1)/2} \exp\left\{ -\frac{1}{2} tr(\Sigma^{-1}S) \right\},$$

(C.6)

where the matrices S and Σ are constrained to be positive definite.

To draw a $p \times p$ random matrix from a Wishart distribution with scale matrix Σ and ν degrees of freedom, first draw $\mathbf{x}_1, \ldots, \mathbf{x}_\nu$ from a zero mean multivariate normal distribution with $p \times p$ covariance matrix Σ. Then, calculate the sum $S =$

$\sum_{i=1}^{\nu} \mathbf{x}_i \mathbf{x}_i^T$. The quantity S is then a random draw from a Wishart distribution. Note that this technique only works when $\nu \geq p$. A random draw from the inverse Wishart distribution with scale matrix S and degrees of freedom ν may be obtained by first obtaining a random draw W from a Wishart distribution with scale matrix S^{-1} and degrees of freedom ν. The quantity $\Sigma = W^{-1}$ will then follow an inverse Wishart distribution.

APPENDIX D

THE KULLBACK-LEIBLER INFORMATION

The Kullback-Leibler Information (KLI, Kullback & Leibler, 1951) is a common method for comparing models and their representation of data. The KLI may be thought of as representing the information lost when a parametric model is used to approximate the true distribution that gave rise to the (Anderson et al., 2000). This approach is appealing because it attempts to find the model that is 'closest' to the true distribution. This approach differs from the classical method of statistical hypothesis testing, in that the KLI looks for the model that best describes the data, without assuming that the true model is among the set of models considered. Furtermore, the KLI allows for the comparison of both nested and nonnested models. This is important, because the parametric models considered in Equations (6.3)–(6.7) are idealizations that are unlikely to be completely true, and do not form a set of nested parametric forms. However, among these idealizations, we can attempt to find the model that best describes the observed data, while fully admitting that such a model is unlikely to be true exactly. In contrast, the classical approach assumes that some null hypothesis is correct, and then tests whether the model parameters are compatible with this null hypothesis at some set significance level. The significance level is usually set such that, if the null hypothesis is true, then one would incorrectly reject it with low probability. In other words, the classical approach assumes that a 'null' model is correct, and then looks for overwhelming evidence to the contrary. However, a comparison of models based on the KLI does not make any *a priori* assumptions about which model is correct, but rather assesses which (flawed) model best describes the observed data; i.e., is 'closest' to the true probability density that generated

the data. In this sense, the KLI evaluates the evidence *for* the models considered, whereas the classical approach only evaluates the evidence *against* some assumed null model. In addition, comparing models based on the KLI has the advantage that all models may be compared simultaneously, whereas the classical approach can only compare two models at a time.

Problems with the classical approach to statistical hypothesis testing and model selection have been known for some time, and many authors have proposed using the KLI as an alternative (e.g., Akaike, 1974). There is a large literature on these issues; see Anderson et al. (2000) and references therein for a more thorough discussion of the problems with classical hypothesis testing and the advantages of the information-theoretic approach.

The KLI measures the discrepancy between the model distribution for the data, $p(y|\theta)$, parameterized by θ, and the true distribution of the data $f(y)$; note that $p(y|\theta)$ is the likelihood function for y. The KLI may be thought of as the relative entropy of a statistical model, and is given by

$$H(\theta) = \int \log \left(\frac{f(y)}{p(y|\theta)} \right) f(y) dy. \tag{D.1}$$

The difference in KLI between two statistical models, $p_j(y|\theta_j)$ and $p_k(y|\theta_k)$, is then

$$H(\theta_j) - H(\theta_k) = \int [\log p_k(y|\theta_k) - \log p_j(y|\theta_j)] f(y) dy. \tag{D.2}$$

As is apparent from Equation (D.2), the difference in KLI between two statistical models is the expectation of the difference in their log-likelihoods, multiplied by -1. The Kullback-Leibler information describes the information lost when a statistical model, $p(y|\theta)$, is used to approximate the true distribution, $f(y)$.

For this work $y = \alpha_{\mathrm{ox}}$, and the true sampling distribution of α_{ox} is denoted as $f(\alpha_{\mathrm{ox}}|l_{UV}, z)$. The distribution is made conditional on (l_{UV}, z) because we are interested in how α_{ox} is distributed at a given l_{UV} and z.

In order to choose the model that minimizes the KLI, it is necessary to find the values of the model parameters, θ, that minimize Equation (D.1). Unfortunately, this requires knowledge of the unknown sampling density, $f(\alpha_{\mathrm{ox}}|l_{UV}, z)$. However, the maximum likelihood estimate of θ, $\hat{\theta}$, provides a good estimate of the θ that minimizes the KLI, and in fact converges to it as the sample size becomes large (e.g., Shibata, 1997). Furthermore, for non-censored Gaussian data, $\hat{\theta}$ corresponds to the θ that minimizes the squared error between the data and the model predictions, and thus finding the model that minimizes the KLI is asymptotically equivalent to finding the model that minimizes the expected squared error. In this work the amount of censoring is small, and the censored data only contribute to the log-likelihood at the $\approx 5\%$ level. Because the residuals from the α_{ox} regressions are approximately Gaussian, and because the amount of censoring is small, we expect our statistical models to behave similarly to the usual uncensored case for Gaussian data. Therefore, finding the parameterization that minimizes the KLI has the straight-forward interpretation of finding the parameterization that approximately minimizes the expected squared error.

Because the difference in KLI between two models is the expected difference in their log-likelihoods, one can estimate $H(\hat{\theta}_j) - H(\hat{\theta}_k)$ for a single data point using the sample mean of the difference in log-likelihoods, evaluated at the maximum likelihood estimates of θ. However, this produces a biased estimate of $H(\theta_j) - H(\theta_k)$, as we use the same data to fit the model as to estimate its KLI. Akaike (1974) showed that this bias is on the order of the number of free parameters in the statistical model. This led him to define the Akaike Information Criterion (AIC):

$$AIC = -2\log p(y|\hat{\theta}) + 2d. \tag{D.3}$$

Here, $\log p(y|\hat{\theta})$ is the log-likelihood of the data evaluated at the maximum-likelihood

estimate of θ, and d is the number of free parameters in the model. The differences in AIC between models may then be used as estimates of the differences in KLI between the models. The model with the best estimated KLI is the model that minimizes the AIC.

APPENDIX E

KENDALL'S PARTIAL τ

To further assess the behavior of Kendall's partial τ we performed additional simulations, varying the degree of correlation between L_{UV} and z. We drew 165 values of l_{UV} and $\log z$ from a multivariate normal density for 1000 simulations, with correlations between l_{UV} and $\log z$ of $\rho = 0.0, 0.3, 0.6$, and 0.9. The remainder of the simulations were performed in an identical manner to those described in § 6.4.5 and § 6.4.6, with the exception that we did not include the effects of variability and measurement error. For each of these cases, we calculated the average values of $\tau_{xz,l}$ and $\tau_{\alpha z,l}$ under both the null and alternative hypotheses. From the notation of § 6.4.6, $\tau_{xz,l}$ denotes the value of Kendall's partial τ between L_X and z, controlling for L_{UV}, and $\tau_{\alpha z,l}$ denotes the value of Kendall's partial τ between α_{ox} and z, controlling for L_{UV}.

In addition, we also calculate the power of the test when using either $\tau_{xz,l}$ and $\tau_{\alpha z,l}$ under both hypotheses; the power of a statistical test is the probability of choosing for the alternative hypothesis. We have chosen a significance level of 0.05, meaning that we reject the null hypothesis, H_0, when $|\tau - E(\tau|H_0)|/\sigma_\tau > 1.96$, where σ_τ is the standard deviation in τ. Here, the notation $E(x|H)$ denotes the conditional expectation value of x, given hypothesis H. Using this significance level, we would expect to incorrectly reject the null hypothesis $\approx 5\%$ of the time. To illustrate the effect of incorrectly assuming that $E(\tau|H_0) = 0$, we calculate the power assuming $E(\tau|H_0) = 0$. Therefore, the null hypothesis is rejected when the observed value of τ falls within the region, $|\tau| > 1.96\sigma_\tau$.

The results of our simulations are shown in Table E.1. As can be seen, incorrectly assuming $E(\tau|H_0) = 0$ has a significant effect on the conclusions drawn

Table E.1.   Behavior of Kendall's Generalized Partial τ From Simulation

		$H_0 : L_X \propto L_{UV}^{0.65}$				$H_1 : L_X \propto L_{UV}^{0.4} e^{-t(z)/5.5}$			
ρ^{a}	τ_{lz}^{b}	$\tau_{xz,l}^{\text{c}}$	$\text{pow}(\tau_{xz,l}\|H_0)^{\text{d}}$	$\tau_{\alpha z,l}^{\text{e}}$	$\text{pow}(\tau_{\alpha z,l}\|H_0)^{\text{f}}$	$\tau_{xz,l}^{\text{c}}$	$\text{pow}(\tau_{xz,l}\|H_1)^{\text{d}}$	$\tau_{\alpha z,l}^{\text{e}}$	$\text{pow}(\tau_{\alpha z,l}\|H_1)^{\text{f}}$
0.9	0.715	0.153	0.988	0.105	0.845	0.290	1.000	0.011	0.061
0.6	0.412	0.101	0.696	0.074	0.390	0.338	1.000	-0.126	0.935
0.3	0.197	0.048	0.196	0.038	0.126	0.329	1.000	-0.204	1.000
0.0	0.000	0.000	0.050	0.000	0.050	0.299	1.000	-0.258	1.000

[a] The correlation between l_{UV} and $\log z$ used for the simulation.

[b] The average value of Kendall's τ between L_{UV} and z.

[c] The average value of Kendall's generalized partial τ between L_X and z, controlling for L_{UV}. The left value of $\tau_{xz,l}$ corresponds to when the null hypothesis, H_0, is true, and the right value corresponds to when the alternative hypothesis, H_1, is true.

[d] The power of the test when $\tau_{xz,l}$ is used and the null hypothesis H_0 is true (left) or the alternative hypothesis H_1 is true (right).

[e] Same as $\tau_{xz,l}$, but when using α_{ox} instead of L_X.

[f] Same as $\text{pow}(\tau_{xz,l}|\cdot)$, but when using $\tau_{az,l}$ instead of $\tau_{xz,l}$.

from Kendall's generalized partial τ. This is particularly notable when the correlation between the two 'independent' variables is high, in this case between L_{UV} and z. For high correlations between L_{UV} and z, one incorrectly rejects the null hypothesis with probability $p \approx 0.988$ when using $\tau_{xz,l}$, and $p \approx 0.845$ when using $\tau_{\alpha z,l}$. Furthermore, when L_{UV} and z are highly correlated and one is using $\tau_{\alpha z,l}$, one rarely ($p \approx 0.061$) rejects the null hypothesis if the alternative hypothesis is true. However, one does reject the null hypothesis when the alternative is true with near certainty when using $\tau_{xz,l}$. These problems are ameliorated when the degree of correlation between the independent variables is reduced; however, even when $\rho = 0.3$ one still incorrectly rejects the null hypothesis with probability a factor of $\sim$ 2–4 times higher than the expected 5%.

We also performed our simulations experimenting with different sample sizes. We did not notice any dependence of $E(\tau|H_0)$ on the sample size. This results in the seemingly paradoxical behavior that, as the sample size increases, the null hypothesis is rejected with increasing probability even if it is true, if one assumes $E(\tau|H_0) = 0$. While the expected value of τ under the null hypothesis does not appear to depend on sample size, the variance in τ, σ_τ^2, does, decreasing as the sample size increases. Thus, as the sample size increases, σ_τ decreases, and the observed values of τ become more concentrated around $E(\tau|H_0)$. As a result, if one incorrectly assumes $E(\tau|H_0) = 0$, it becomes more likely that τ falls inside of the region where $|\tau| > 1.96\sigma_\tau$, and thus one is more likely to incorrectly reject the null hypothesis as the sample size increases.

Similar results using Monte Carlo Simulations were found by Steffen et al. (2006). S06 concluded that their results implied that spurious but false correlations between L_X and z result when there is a high degree of correlation between L_{UV} and z. However, the results do not imply that spurious but false correlation arise when there is a high degree of correlation between L_{UV} and z, but rather the simulations show that the expected value of τ under the null hypothesis varies as the degree of correlation between L_{UV} and z varies. This may then result in apparently significant correlations because one is incorrectly assuming $E(\tau|H_0) = 0$. This incorrect assumption can also result in apparently insignificant correlations, even if such correlations are significant. Our simulations clearly show this, since they were constructed to ensure that under the null hypothesis α_{ox} is independent of z, given L_{UV},, *independent* of the distribution of L_{UV} and z.

Akritas & Siebert (1996) also performed simulations to assess the behavior of Kendall's partial τ. They investigated the power of the test and found that the test becomes more powerful as the departure from the null hypothesis increases.

They also found that when the null hypothesis is true, and when the two independent variable are statistically independent, the probability of rejecting H_0 is equal to the chosen significance level. Similar to us, they also use a significance level of 0.05 (1.96σ). However, when investigating the probability of incorrectly rejecting the null hypothesis, they only used simulations where the two independent variables (L_{UV} and z in this work) are statistically independent. We have confirmed this result here, but have also expanded upon it, showing that the null hypothesis can be incorrectly rejected with high probability when the two independent variables have a moderate to high correlation. Furthermore, Akritas & Siebert (1996) concluded that the Kendall's partial τ test can become more powerful when the alternative hypothesis is far from the null hypothesis, rejecting the null hypothesis with high probability. However, we have shown here that this is not always true, and that the power of the test depends on the degree of correlation between the two independent variables. Indeed, when L_{UV} and z are highly correlated, and when one is investigating the partial correlation between α_{ox} and z, one almost always incorrectly claims that the data are consistent with a null hypothesis of no evolution in α_{ox}.

It should be noted that these results only apply to Kendall's *partial* τ, and not to the usual Kendall's τ. Both Kendall's τ and the partial linear correlation have values of zero under the null hypothesis of statistical independence between the variables of interest (Nelson & Yang, 1988).

Our simulations show that the expected value of Kendall's partial τ under the null hypothesis depends on the distributional properties of the sample, and that this can significantly effect the power of the test. Unfortunately, we know of no way in which to analytically calculate the expected value of τ under the null hypothesis, and it must likely be calculated using simulation. However, one must

likely employ parametric methods in order to simulate data, and this undermines the nonparametric nature of Kendall's partial τ. We also note that these results are not meant to be a complete dismissal of the use of Kendall's generalized partial τ, but rather to point out the problems that can arise when using Kendall's partial τ. If one does not know $E(\tau|H_0)$, then one is not able to calibrate the partial τ statistic against a physically meaningful null hypothesis, therefore making statistical hypothesis testing based on it suspect.

REFERENCES

Abazajian, K., et al. 2004, AJ, 128, 502

Abramowicz, M. A., Czerny, B., Lasota, J. P., & Szuszkiewicz, E. 1988, ApJ, 332, 646

Adams, F. C., Graff, D. S., & Richstone, D. O. 2001, ApJL, 551, L31

Aitken, M. 1981, Technometrics, 23, 161

Aitken, M., & Rocci, R., 2002, Statistics and Computing, 12, 163

Akaike, H. 1974, IEEE Transactions on Automatic Control, 19, 716

Akritas, M. G., & Bershady, M. A. 1996, ApJ, 470, 706

Akritas, M. G., & Siebert, J. 1996, MNRAS, 278, 919

Alcock, C., et al. 1997, ApJ, 486, 697

Alcock, C., et al. 1999, PASP, 111, 1539

Anderson, D. R., Burnham, K. P., Thompson, W. L. 2000, J. Wildl. Manage., 64, 912

Anderson, S. F., et al. 2001, AJ, 122, 503

Aretxaga, I., Cid Fernandes, R., & Terlevich, R. J. 1997, MNRAS, 286, 271

Avni, Y., & Bahcall, J. N. 1980, ApJ, 235, 694

Avni, Y., & Tananbaum, H. 1982, ApJL, 262, L17

Avni, Y., & Tananbaum, H. 1986, ApJ, 305, 83 (AT86)

Babbedge, T. S. R., et al. 2006, MNRAS, 370, 1159

Balbus, S. A., & Hawley, J. F. 1991, ApJ, 376, 214

Balbus, S. A., & Hawley, J. F. 1998, Reviews of Modern Physics, 70, 1

Ballantyne, D. R., Ross, R. R., & Fabian, A. C. 2001, MNRAS, 327, 10

Barger, A. J., Cowie, L. L., Mushotzky, R. F., Yang, Y., Wang, W.-H., Steffen, A. T., & Capak, P. 2005, AJ, 129, 578

Barker, D. R., & Diana, L. M. 1974, Am. J. Phys., 42, 224

Baskin, A., & Laor, A. 2004, MNRAS, 350, L31

Baskin, A., & Laor, A. 2005, MNRAS, 356, 1029

Bechtold, J., Dobrzycki, A., Wilden, B., Morita, M., Scott, J., Dobrzycka, D., Tran, K., & Aldcroft, T. L. 2002, ApJS, 140, 143

Bechtold, J., et al. 2003, ApJ, 588, 119

Begelman, M. C., & Nath, B. B. 2005, MNRAS, 361, 1387

Bentz, M. C., Peterson, B. M., Pogge, R. W., Vestergaard, M., & Onken, C. A. 2006, ApJ, 644, 133

Bian, W.-H. 2005, Chinese Journal of Astronomy and Astrophysics Supplement, 5, 289

Bianchi, S., Guainazzi, M., Matt, G., & Fonseca Bonilla, N. 2007, A&A, 467, L19

Bisnovatyi-Kogan, G. S., & Blinnikov, S. I. 1977, A&A, 59, 111

Blandford, R. D., & McKee, C. F. 1982, ApJ, 255, 419

Blanton, M. R., et al. 2003, ApJ, 592, 819

Boller, T., Brandt, W. N., & Fink, H. 1996, A&A, 305, 53

Bonning, E. W., Cheng, L., Shields, G. A., Salviander, S., & Gebhardt, K. 2007, ApJ, 659, 211

Boroson, T. A., & Green, R. F. 1992, ApJS, 80, 109

Bower, R. G., Benson, A. J., Malbon, R., Helly, J. C., Frenk, C. S., Baugh, C. M., Cole, S., & Lacey, C. G. 2006, MNRAS, 370, 645

Brandt, W. N., Mathur, S., & Elvis, M. 1997, MNRAS, 285, L25

Brandt, N., & Boller, T. 1998, Astronomische Nachrichten, 319, 163

Brocksopp, C., Starling, R. L. C., Schady, P., Mason, K. O., Romero-Colmenero, E., & Puchnarewicz, E. M. 2006, MNRAS, 366, 953

Brockwell, P. .J., & Davis, R. A. 2002 Introduction to Time Series and Forecasting (2nd Ed.; New York, NY: Springer)

Bromley, J. M., Somerville, R. S., & Fabian, A. C. 2004, MNRAS, 350, 456

Brown, M. J. I., et al. 2006, ApJ, 638, 88

Brown, M. J. I., Dey, A., Jannuzi, B. T., Brand, K., Benson, A. J., Brodwin, M., Croton, D. J., & Eisenhardt, P. R. 2007, ApJ, 654, 858

Brunner, H., Mueller, C., Friedrich, P., Doerrer, T., Staubert, R., & Riffert, H. 1997, A&A, 326, 885

Budavári, T., et al. 2005, ApJL, 619, L31

Bundy, K., et al. 2006, ApJ, 651, 120

Cao, X., & Xu, Y.-D. 2007, MNRAS, 377, 425

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245

Carini, M. T., Noble, J. C., Taylor, R., & Culler, R. 2007, AJ, 133, 303

Carroll, R. J., Roeder, K., & Wasserman, L., 1999, Biometrics, 55, 44

Carroll, R. J., Ruppert, D., & Stefanski, L. A., 1995, Measurement Error in Nonlinear Models (London:Chapman & Hall)

Cash, W. 1979, ApJ, 228, 939

Chib, S., & Greenberg, E. 1995, Amer. Stat., 49, 327

Cid Fernandes, R. J., Aretxaga, I., & Terlevich, R. 1996, MNRAS, 282, 1191

Cid Fernandes, R., Sodré, L., Jr., & Vieira da Silva, L., Jr. 2000, ApJ, 544, 123

Ciotti, L., & Ostriker, J. P. 2001, ApJ, 551, 131

Cirasuolo, M., et al. 2007, MNRAS, 380, 585

Clutton-Brock, M. 1967, Technometrics, 9, 261

Collier, S., & Peterson, B. M. 2001, ApJ, 555, 775

Collin, S., Boisson, C., Mouchet, M., Dumont, A.-M., Coupé, S., Porquet, D., & Rokaki, E. 2002, A&A, 388, 771

Collin, S., Kawaguchi, T., Peterson, B. M., & Vestergaard, M. 2006, A&A, 456, 75

Congdon, P. 2006, Bayesian Statistical Modelling (2nd. Ed.; West Sussex:John Wiley & Sons Ltd.)

Constantin, A., Shields, J. C., Hamann, F., Foltz, C. B., & Chaffee, F. H. 2002, ApJ, 565, 50

Corbett, E. A., Robinson, A., Axon, D. J., Young, S., & Hough, J. H. 1998, MNRAS, 296, 721

Corbett, E. A., et al. 2003, MNRAS, 343, 705

Cristiani, S., Vio, R., & Andreani, P. 1990, AJ, 100, 56

Cristiani, S., Trentini, S., La Franca, F., Aretxaga, I., Andreani, P., Vio, R., & Gemmo, A. 1996, A&A, 306, 395

Croom, S. M., Smith, R. J., Boyle, B. J., Shanks, T., Miller, L., Outram, P. J., & Loaring, N. S. 2004, MNRAS, 349, 1397

Croton, D. J., et al. 2005, MNRAS, 356, 1155

Croton, D. J., et al. 2006, MNRAS, 365, 11

Cutri, R. M., Wisniewski, W. Z., Rieke, G. H., & Lebofsky, M. J. 1985, ApJ, 296, 423

Czerny, B., Nikołajuk, M., Różańska, A., Dumont, A.-M., Loska, Z., & Zycki, P. T. 2003, A&A, 412, 317

Czerny, B., Rózańska, A., & Kuraszkiewicz, J. 2004, A&A, 428, 39

Czerny, B., Siemiginowska, A., Janiuk, A., & Gupta, A. C. 2008, in press at MN-RAS(arXiv:0802.4396)

Czerny, B., Witt, H. J., & Zycki, P. 1997, The Transparent Universe, 382, 397

Dahlen, T., Mobasher, B., Somerville, R. S., Moustakas, L. A., Dickinson, M., Ferguson, H. C., & Giavalisco, M. 2005, ApJ, 631, 126

Dai, X., Chartas, G., Eracleous, M., & Garmire, G. P. 2004, ApJ, 605, 45

Davis, S. W., Woo, J.-H., & Blaes, O. M. 2007, ApJ, 668, 682

Davison, A. C., & Hinkley, D. V. 1997, Bootstrap Methods and their Application (Cambridge:Cambridge University Press)

Dellaportas, P., & Papageorgiou, I. 2006, Stat. Comput., 16, 57

Dellaportas, P., Stephens, D. A., 1995, Biometrics, 51, 1085

Dempster, A., Laird, N., & Rubin, D. 1977, J. R. Statist. Soc. B., 39, 1

De Villiers, J.-P., Hawley, J. F., & Krolik, J. H. 2003, ApJ, 599, 1238

de Vries, W. H., Becker, R. H., White, R. L., & Loomis, C. 2005, AJ, 129, 615

Dickey, J. M., & Lockman, F. J. 1990, ARA&A, 28, 215

di Clemente, A., Giallongo, E., Natali, G., Trevese, D., & Vagnetti, F. 1996, ApJ, 463, 466

Dietrich, M., Appenzeller, I., Vestergaard, M., & Wagner, S. J. 2002, ApJ, 564, 581

Dietrich, M., & Hamann, F. 2004, ApJ, 611, 761

Di Matteo, T., Springel, V., & Hernquist, L. 2005, Nature, 433, 604

Done, C., & Krolik, J. H. 1996, ApJ, 463, 144

Dunlop, J. S., McLure, R. J., Kukula, M. J., Baum, S. A., O'Dea, C. P., & Hughes, D. H. 2003, MNRAS, 340, 1095

Efron, B. 1979, Ann. Statist., 7, 1

Efron, B. 1984, J. Amer. Stat. Assoc., 79, 791

Efron, B. 1987, J. Americ. Stat. Assoc., 82, 171

Efron, B., & Petrosian, V. 1992, ApJ, 399, 345

Elvis, M., et al. 1994, ApJS, 95, 1

Elvis, M. 2006, Memorie della Societa Astronomica Italiana, 77, 573

Elvis, M., Marengo, M., & Karovska, M. 2002, ApJL, 567, L107

Elvis, M., Risaliti, G., & Zamorani, G. 2002, ApJL, 565, L75

Edmunds, M. G., & Eales, S. A. 1998, MNRAS, 299, L29

Esin, A. A., McClintock, J. E., & Narayan, R. 1997, ApJ, 489, 865

Evans, I. N., & Koratkar, A. P. 2004, ApJS, 150, 73

Faber, S. M., et al. 2007, ApJ, 665, 265

Fabian, A. C. 1999, MNRAS, 308, L39

Fabian, A. C., et al. 2002, MNRAS, 335, L1

Fan, X., et al. 2001, AJ, 121, 54

Fan, X., et al. 2006, AJ, 131, 1203

Feigelson, E. D., 1992, in Statistical Challenges in Modern Astronomy, ed. E. Feigelson & G. Babu, (New York:Springer-Verlag)

Feigelson, E. D., & Nelson, P. I. 1985, ApJ, 293, 192

Ferrarese, L., Pogge, R. W., Peterson, B. M., Merritt, D., Wandel, A., & Joseph, C. L. 2001, ApJL, 555, L79

Finlator, K., Davé, R., Papovich, C., & Hernquist, L. 2006, ApJ, 639, 672

Fox, J. 1997, Applied Regression Analysis, Linear Models, and Related Methods (Thousand Oaks:Sage Publications, Inc.)

Frank, J., King, A., & Raine, D. 2002, Accretion Power in Astrophysics (3rd Edition;Cambridge, UK:Cambridge Univ. Press)

Freedman, L. S., Fainberg, V., Kipnis, V., Midthune, D., & Carrol, R. J., 2004, Biometrics, 60, 172

Freeman, P. E., Graziani, C., Lamb, D. Q., Loredo, T. J., Fenimore, E. E., Murakami, T., & Yoshida, A. 1999, ApJ, 524, 753

Freeman, P. E., Doe, S., & Siemiginowska, A. 2001, SPIE Proceeedings, 4477, 76

Fromerth, M. J., & Melia, F. 2000, ApJ, 533, 172

Fuller, W. A. 1987, Measurement Error Models (New York:John Wiley & Sons)

Galeev, A. A., Rosner, R., & Vaiana, G. S. 1979, ApJ, 229, 318

Gallagher, S. C., Brandt, W. N., Chartas, G., & Garmire, G. P. 2002, ApJ, 567, 37

Gallagher, S. C., Richards, G. T., Hall, P. B., Brandt, W. N., Schneider, D. P., & Vanden Berk, D. E. 2005, AJ, 129, 567

Gallagher, S. C., Brandt, W. N., Chartas, G., Priddey, R., Garmire, G. P., & Sambruna, R. M. 2006, ApJ, 644, 709

Garcia, A., Sodré, L., Jablonski, F. J., & Terlevich, R. J. 1999, MNRAS, 309, 803

Gebhardt, K., et al. 2000, ApJL, 539, L13

Gebhardt, K., et al. 2000, ApJL, 543, L5

Geha, M., et al. 2003, AJ, 125, 1

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. 2004, Bayesian Data Analysis (2nd ed.; Boca Raton:Chapman & Hall/CRC)

Gelman, A., Meng, X. L., & Stern, H. S. 1998, Statistica Sinica, 6, 733

Gelman, A., Roberts, G., & Gilks, W. 1995, in Bayesian Statistics 5, ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Oxford:Oxford University Press), 599

Gierliński, M., & Done, C. 2004, MNRAS, 349, L7

Gillespie, D. T. 1996, Am. J. Phys., 64, 225

Gilli, R., Comastri, A., & Hasinger, G. 2007, A&A, 463, 79

Giveon, U., Maoz, D., Kaspi, S., Netzer, H., & Smith, P. S. 1999, MNRAS, 306, 637

Gopal-Krishna, Stalin, C. S., Sagar, R., & Wiita, P. J. 2003, ApJL, 586, L25

Green, P. J., Aldcroft, T. L., Mathur, S., Wilkes, B. J., & Elvis, M. 2001, ApJ, 558, 109

Green, R. F., Schmidt, M., & Liebert, J. 1986, ApJS, 61, 305

Greene, J. E., & Ho, L. C. 2007, ApJ, 667, 131

Grupe, D., Thomas, H.-C., & Beuermann, K. 2001, A&A, 367, 470

Grupe, D., Mathur, S., Wilkes, B., & Osmer, P. 2006, AJ, 131, 55

Gull, S. F. 1989, in Maximum Entropy and Bayesian Methods, ed. J. Skilling, (Dordrecht:Kluwer Academic Publishers) 511

Gupta, A. C., & Joshi, U. C. 2005, A&A, 440, 855

Haardt, F., & Maraschi, L. 1991, ApJL, 380, L51

Haardt, F., & Maraschi, L. 1993, ApJ, 413, 507

Haehnelt, M. G., & Kauffmann, G. 2000, MNRAS, 318, L35

Haiman, Z., & Menou, K. 2000, ApJ, 531, 42

Hao, L., et al. 2005, AJ, 129, 1795

Harsono, D., & de Propris, R. 2007, MNRAS, 380, 1036

Hastie, T., Tibshirani, R., & Friedman, J. 2001, The Elements of Statistical Learning
    (New York:Springer-Verlag)

Hastings, W. K. 1970, Biometrika, 57, 97

Hatziminaoglou, E., Mathez, G., Solanes, J.-M., Manrique, A., & Salvador-Solé, E.
    2003, MNRAS, 343, 692

Hawkins, M. R. S. 2000, A&AS, 143, 465

Hawkins, M. R. S. 2007, A&A, 462, 581

Helfand, D. J., Stone, R. P. S., Willman, B., White, R. L., Becker, R. H., Price, T.,
    Gregg, M. D., & McMahon, R. G. 2001, AJ, 121, 1872

Hirashita, H., & Ferrara, A. 2002, MNRAS, 337, 921

Ho, L. C. 2002, ApJ, 564, 120

Hook, I. M., McMahon, R. G., Boyle, B. J., & Irwin, M. J. 1994, MNRAS, 268, 305

Hopkins, P. F., Hernquist, L., Cox, T. J., Di Matteo, T., Martini, P., Robertson, B., &
    Springel, V. 2005, ApJ, 630, 705

Hopkins, P. F., & Hernquist, L. 2006, ApJS, 166, 1

Hopkins, P. F., Hernquist, L., Cox, T. J., Di Matteo, T., Robertson, B., & Springel, V. 2006, ApJS, 163, 1

Hopkins, P. F., Narayan, R., & Hernquist, L. 2006, ApJ, 643, 641

Hopkins, P. F., Richards, G. T., & Hernquist, L. 2007, ApJ, 654, 731

Hoyle, F., Rojas, R. R., Vogeley, M. S., & Brinkmann, J. 2005, ApJ, 620, 618

Huang, X., Stefanski, L. A., & Davidian, M. 2006, Biometrika, 93, 53

Huynh, M. T., Frayer, D. T., Mobasher, B., Dickinson, M., Chary, R.-R., & Morrison, G. 2007, ApJL, 667, L9

Isobe, T., Feigelson, E. D., & Nelson, P. I. 1986, ApJ, 306, 490

Isobe, T., Feigelson, E. D., Akritas, M. G., & Babu, G. J. 1990, ApJ, 364, 104

Ivezić, Z., et al. 2004, ASP Conf. Ser. 311: AGN Physics with the Sloan Digital Sky Survey, 311, 347

Janiuk, A., & Czerny, B. 2000, New Astronomy, 5, 7

Janiuk, A., Czerny, B., & Madejski, G. M. 2001, ApJ, 557, 408

Janiuk, A., & Czerny, B. 2007, A&A, 466, 793

Jasra, A., Holmes, C.C., & Stephens, D.A. 2005, Statistical Science, 20, 50

Jester, S. 2005, ApJ, 625, 667

Jiang, L., et al. 2006, AJ, 131, 2788

Jiang, L., et al. 2006, AJ, 132, 2127

Just, D. W., Brandt, W. N., Shemmer, O., Steffen, A. T., Schneider, D. P., Chartas, G., & Garmire, G. P. 2007, ApJ, 665, 1004

Kaspi, S., Smith, P. S., Netzer, H., Maoz, D., Jannuzi, B. T., & Giveon, U. 2000, ApJ, 533, 631

Kaspi, S., Maoz, D., Netzer, H., Peterson, B. M., Vestergaard, M., & Jannuzi, B. T. 2005, ApJ, 629, 61

Kauffmann, G., & Haehnelt, M. 2000, MNRAS, 311, 576

Kawaguchi, T., Mineshige, S., Umemura, M., & Turner, E. L. 1998, ApJ, 504, 671

Kelly, B. C. 2007, ApJ, 665, 1489

Kelly, B. C., & Bechtold, J. 2007, ApJS, 168, 1

Kelly, B. C., Bechtold, J., Siemiginowska, A., Aldcroft, T., & Sobolewska, M. 2007, ApJ, 657, 116

Kelly, B. C., Bechtold, J., Trump, J. R., Vestergaard, M., & Siemiginowska, A. 2008, in press at ApJS(arXiv:0801.2383)

Kelly, B. C, Fan, X., & Vestergaard, M. 2008, submitted to ApJ

Kennefick, J. D., de Carvalho, R. R., Djorgovski, S. G., Wilber, M. M., Dickson, E. S., Weir, N., Fayyad, U., & Roden, J. 1995, AJ, 110, 78

Kim, D.-W., et al. 2006, ApJ, 652, 1090

King, A. R., Pringle, J. E., West, R. G., & Livio, M. 2004, MNRAS, 348, 111

Kollmeier, J. A., et al. 2006, ApJ, 648, 128

Kormendy, J., & Richstone, D. 1995, ARA&A, 33, 581

Krolik, J. H. 1999, Active Galactic Nuclei: From the Central Engine to the Galactic Environment (Princeton, NJ:Princeton Univ. Press)

Krolik, J. H. 1999, High Energy Processes in Accreting Black Holes, 161, 315

Krolik, J. H. 2001, ApJ, 551, 72

Krolik, J. H., Hawley, J. F., & Hirose, S. 2005, ApJ, 622, 1008

Kukula, M. J., Dunlop, J. S., McLure, R. J., Miller, L., Percival, W. J., Baum, S. A., & O'Dea, C. P. 2001, MNRAS, 326, 1533

Kuraszkiewicz, J. K., Green, P. J., Crenshaw, D. M., Dunn, J., Forster, K., Vestergaard, M., & Aldcroft, T. L. 2004, ApJS, 150, 165

Kullback, S., & Leibler, R. A. 1951, Ann. Math. Statist., 22, 76

La Franca, F., et al. 2005, ApJ, 635, 864

Landy, S. D., & Szalay, A. S. 1992, ApJ, 391, 494

Laor, A., Fiore, F., Elvis, M., Wilkes, B. J., & McDowell, J. C. 1997, ApJ, 477, 93

Lauer, T. R., et al. 2007, ApJ, 662, 808

Lee, J. C., Iwasawa, K., Houck, J. C., Fabian, A. C., Marshall, H. L., & Canizares, C. R. 2002, ApJL, 570, L47

Leipski, C., Falcke, H., Bennert, N., Hüttemeister, S. 2006, A&A, 455, 161

Liang, E. P. T., & Price, R. H. 1977, ApJ, 218, 247

Lightman, A. P., & Eardley, D. M. 1974, ApJL, 187, L1

Lin, Y.-T., & Mohr, J. J. 2007, ApJS, 170, 71

Little, R. J. A., & Rubin, D. B. 2002, Statistical Analysis with Missing Data (2nd ed.; Hoboken:John Wiley & Sons)

Liu, B. F., Mineshige, S., & Ohsuga, K. 2003, ApJ, 587, 571

Loredo, T. J., 1992, in Statistical Challenges in Modern Astronomy, ed. E. Feigelson & G. Babu, (New York:Springer-Verlag)

Lu, Y., & Yu, Q. 1999, ApJL, 526, L5

Lynden-Bell, D. 1969, Nature, 223, 690

Lynden-Bell, D. 1971, MNRAS, 155, 95

Lyubarskii, Y. E. 1997, MNRAS, 292, 679

Magorrian, J., et al. 1998, AJ, 115, 2285

Mainieri, V., et al. 2007, ApJS, 172, 368

Malkan, M. 1991, IAU Colloq. 129: The 6th Institute d'Astrophysique de Paris (IAP) Meeting: Structure and Emission Properties of Accretion Disks, 165

Maloney, A., & Petrosian, V. 1999, ApJ, 518, 32

Malzac, J., Beloborodov, A. M., & Poutanen, J. 2001, MNRAS, 326, 417

Manmoto, T., Takeuchi, M., Mineshige, S., Matsumoto, R., & Negoro, H. 1996, ApJL, 464, L135

Marchesini, D., Celotti, A., & Ferrarese, L. 2004, MNRAS, 351, 733

Marchesini, D., et al. 2007, ApJ, 656, 42

Marchesini, D., & van Dokkum, P. G. 2007, ApJL, 663, L89

Marconi, A., & Hunt, L. K. 2003, ApJL, 589, L21

Marconi, A., Risaliti, G., Gilli, R., Hunt, L. K., Maiolino, R., & Salvati, M. 2004, MNRAS, 351, 169

Markowitz, A., Edelson, R., & Vaughan, S. 2003, ApJ, 598, 935

Markowitz, A., & Edelson, R. 2004, ApJ, 617, 939

Marshall, H. L., Tananbaum, H., Avni, Y., & Zamorani, G. 1983, ApJ, 269, 35

Marshall, H. L., 1992, in Statistical Challenges in Modern Astronomy, ed. E. Feigelson & G. Babu, (New York:Springer-Verlag)

Marulli, F., Branchini, E., Moscardini, L., & Volonteri, M. 2007, MNRAS, 375, 649

Marziani, P., Sulentic, J. W., Zamanov, R., Calvani, M., Dultzin-Hacyan, D., Bachev, R., & Zwitter, T. 2003, ApJS, 145, 199

Mathur, S. 2000, MNRAS, 314, L17

Matute, I., La Franca, F., Pozzi, F., Gruppioni, C., Lari, C., & Zamorani, G. 2006, A&A, 451, 443

Mauch, T., & Sadler, E. M. 2007, MNRAS, 375, 931

Mayer, M., & Pringle, J. E. 2006, MNRAS, 368, 379

McHardy, I. M., Papadakis, I. E., Uttley, P., Page, M. J., & Mason, K. O. 2004, MNRAS, 348, 783

McIntosh, D. H., Rieke, M. J., Rix, H.-W., Foltz, C. B., & Weymann, R. J. 1999, ApJ, 514, 40

McLeod, K. K., & McLeod, B. A. 2001, ApJ, 546, 782

McLure, R. J., Kukula, M. J., Dunlop, J. S., Baum, S. A., O'Dea, C. P., & Hughes, D. H. 1999, MNRAS, 308, 377

McLure, R. J., & Dunlop, J. S. 2001, MNRAS, 327, 199

McLure, R. J., & Dunlop, J. S. 2002, MNRAS, 331, 795

McLure, R. J., & Dunlop, J. S. 2004, MNRAS, 352, 1390

McLure, R. J., & Jarvis, M. J. 2002, MNRAS, 337, 109

McLure, R. J., & Jarvis, M. J. 2004, MNRAS, 353, L45

Merloni, A. 2004, MNRAS, 353, 1035

Merloni, A., & Fabian, A. C. 2002, MNRAS, 332, 165

Merloni, A., Heinz, S., & di Matteo, T. 2003, MNRAS, 345, 1057

Merritt, D., & Ferrarese, L. 2001, ApJ, 547, 140

Merritt, D., & Poon, M. Y. 2004, ApJ, 606, 788

Metropolis, N., & Ulam, S. 1949, J. Amer. Stat. Assoc., 44, 335

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, J. Chem. Phys., 21, 1087

Miller, K. A., & Stone, J. M. 2000, ApJ, 534, 398

Mineo, T., et al. 2000, A&A, 359, 471

Müller, P., & Roeder, K., 1997, Biometrika, 84, 523

Murray, N., Chiang, J., Grossman, S. A., & Voit, G. M. 1995, ApJ, 451, 498

Nandra, K., George, I. M., Mushotzky, R. F., Turner, T. J., & Yaqoob, T. 1997, ApJ, 477, 602

Nayakshin, S. 2000, ApJ, 534, 718

Neal, R. M. 1996, Statistics and Computing, 6, 353

Nelson, C. H., Green, R. F., Bower, G., Gebhardt, K., & Weistrop, D. 2004, ApJ, 615, 652

Nelson, P. I., & Yang, S. S., 1988, Statis.Prob.Lett., 6, 147

Netzer, H. 1990, in Active Galactic Nuclei, ed. R. D. Blandford, H. Netzer, & L. Woltjer (Berlin:Springer), 137

Netzer, H. 1990, 20. Saas-Fee Advanced Course of the Swiss Society for Astrophysics and Astronomy: Active galactic nuclei, p. 57 - 160, 57

Netzer, H., & Laor, A. 1993, ApJL, 404, L51

Netzer, H. 2003, ApJL, 583, L5

Nicastro, F., Martocchia, A., & Matt, G. 2003, ApJL, 589, L13

Nicastro, F., et al. 2005, Nature, 433, 495

Nishihara, E., Yamashita, T., Yoshida, M., Watanabe, E., Okumura, S.-I., Mori, A., & Iye, M. 1997, ApJL, 488, L27

Nolan, L. A., Dunlop, J. S., Kukula, M. J., Hughes, D. H., Boroson, T., & Jimenez, R. 2001, MNRAS, 323, 308

Novak, G. S., Faber, S. M., & Dekel, A. 2006, ApJ, 637, 96

Nowak, M. A., Wilms, J., & Dove, J. B. 2002, MNRAS, 332, 856

Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713

Omont, A., Cox, P., Bertoldi, F., McMahon, R. G., Carilli, C., & Isaak, K. G. 2001, A&A, 374, 371

O'Neill, P. M., Nandra, K., Papadakis, I. E., & Turner, T. J. 2005, MNRAS, 358, 1405

Onken, C. A., Ferrarese, L., Merritt, D., Peterson, B. M., Pogge, R. W., Vestergaard, M., & Wandel, A. 2004, ApJ, 615, 645

Osterbrock, D. E., Koski, A. T., & Phillips, M. M. 1976, ApJ, 206, 898

Page, K. L., Reeves, J. N., O'Brien, P. T., Turner, M. J. L., & Worrall, D. M. 2004, MNRAS, 353, 133

Page, M. J., & Carrera, F. J. 2000, MNRAS, 311, 433

Page, K. L., Turner, M. J. L., Reeves, J. N., O'Brien, P. T., & Sembay, S. 2003, MNRAS, 338, 1004

Paltani, S., et al. 2007, A&A, 463, 873

Percival, W. J., Miller, L., McLure, R. J., & Dunlop, J. S. 2001, MNRAS, 322, 843

Péroux, C., Storrie-Lombardi, L. J., McMahon, R. G., Irwin, M., & Hook, I. M. 2001, AJ, 121, 1799

Pessah, M. E. 2007, ApJ, 655, 66

Pessah, M. E., Chan, C.-K., & Psaltis, D. 2006, MNRAS, 372, 183

Pessah, M. E., Chan, C.-k., & Psaltis, D. 2007, ApJL, 668, L51

Pessah, M. E., Chan, C.-K., & Psaltis, D. 2008, MNRAS, 383, 683

Peterson, B. M. 1993, PASP, 105, 247

Peterson, B. M., et al. 2004, ApJ, 613, 682

Peterson, B. M., et al. 2005, ApJ, 632, 799

Pettini, M. 2004, Cosmochemistry. The melting pot of the elements, 257

Phadke, M. S., & Wu, S. M. 1974, J. Amer. Statist. Assoc., 69, 325

Piconcelli, E., Cappi, M., Bassani, L., Di Cocco, G., & Dadina, M. 2003, A&A, 412, 689

Piconcelli, E., Jimenez-Bailón, E., Guainazzi, M., Schartel, N., Rodríguez-Pascual, P. M., & Santos-Lleó, M. 2005, A&A, 432, 15

Popesso, P., Biviano, A., Böhringer, H., & Romaniello, M. 2006, A&A, 445, 29

Porquet, D., Reeves, J. N., O'Brien, P., & Brinkmann, W. 2004, A&A, 422, 85

Poutanen, J., Krolik, J. H., & Ryde, F. 1997, MNRAS, 292, L21

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numberical Recipes (Second ed.; Cambridge:Cambridge Unv. Press), 660

Proga, D. 2005, ApJL, 630, L9

Proga, D. 2007, ApJ, 661, 693

Proga, D., & Kallman, T. R. 2004, ApJ, 616, 688

Protassov, R., van Dyk, D. A., Connors, A., Kashyap, V. L., & Siemiginowska, A. 2002, ApJ, 571, 545

Ptak, A., Mobasher, B., Hornschemeier, A., Bauer, F., & Norman, C. 2007, ApJ, 667, 826

Rees, M. J. 1984, ARA&A, 22, 471

Reeves, J. N., Turner, M. J. L., Ohashi, T., & Kii, T. 1997, MNRAS, 292, 468

Reeves, J. N., & Turner, M. J. L. 2000, MNRAS, 316, 234

Reichard, T. A., et al. 2003, AJ, 126, 2594

Richards, G. T., et al. 2001, AJ, 121, 2308

Richards, G. T., et al. 2002, AJ, 123, 2945

Richards, G. T., Vanden Berk, D. E., Reichard, T. A., Hall, P. B., Schneider, D. P., SubbaRao, M., Thakar, A. R., & York, D. G. 2002, AJ, 124, 1

Richards, G. T., et al. 2006, AJ, 131, 2766

Richardson, S., & Leblond, L., 1997, Statistics in Medicine, 16, 203

Richstone, D., et al. 1998, Nature, 395, A14

Ripley, B. D. 1987, Stochastic Simulation (New York:John Wiley & Sons)

Risaliti, G., & Elvis, M. 2005, ApJL, 629, L17

Roeder, K., & Wasserman, L. 1997, J. Amer. Stat. Assoc., 92, 894

Roy, S., Banerjee, T., 2006, Ann. Instit. Statist. Math., 58, 153

Rubin, D. B. 1981, J. Educational Statistics, 6, 377

Rubin, D. B. 1984, Annals of Statistics, 12, 1151

Salpeter, E. E. 1964, ApJ, 140, 796

Sandage, A. 1965, ApJ, 141, 1560

Scargle, J. D. 1981, ApJS, 45, 1

Schafer, C. M. 2007, ApJ, 661, 703

Schafer, D. W., 1987, 74, 385, Biometrika

Schafer, D. W., 2001, Biometrics, 57, 53

Scheines, R., Hoijtink, H., Boomsma, A., 1999, Psychometrika, 64, 37

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525

Schmidt, M. 1963, Nature, 197, 1040

Schmidt, M. 1968, ApJ, 151, 393

Schmidt, M., & Green, R. F. 1983, ApJ, 269, 352

Schmitt, J. H. M. M. 1985, ApJ, 293, 178

Schwartz, G. 1979, Ann. Statist., 6, 461

Schneider, D. P., Schmidt, M., & Gunn, J. E. 1991, AJ, 101, 2004

Schneider, D. P., Schmidt, M., & Gunn, J. E. 1997, AJ, 114, 36

Schneider, D. P., et al. 2005, AJ, 130, 367

Scott, J., Bechtold, J., Dobrzycki, A., & Kulkarni, V. P. 2000, ApJS, 130, 67

Scoville, N., et al. 2007, ApJS, 172, 1

Shakura, N. I., & Syunyaev, R. A. 1973, A&A, 24, 337

Shakura, N. I., & Sunyaev, R. A. 1976, MNRAS, 175, 613

Shankar, F., Salucci, P., Granato, G. L., De Zotti, G., & Danese, L. 2004, MNRAS, 354, 1020

Shapiro, S. L., Lightman, A. P., & Eardley, D. M. 1976, ApJ, 204, 187

Shemmer, O., Netzer, H., Maiolino, R., Oliva, E., Croom, S., Corbett, E., & di Fabrizio, L. 2004, ApJ, 614, 547

Shemmer, O., et al. 2006, ApJ, 644, 86

Shemmer, O., Brandt, W. N., Netzer, H., Maiolino, R., & Kaspi, S. 2006, ApJL, 646, L29

Shen, Y., Greene, J. E., Strauss, M., Richards, G. T., & Schneider, D. P. 2007, in press at ApJ, (arXiv:0709.3098)

Shibata, R. 1997, Statistica Sinica, 7, 375

Shields, G. A., Gebhardt, K., Salviander, S., Wills, B. J., Xie, B., Brotherton, M. S., Yuan, J., & Dietrich, M. 2003, ApJ, 583, 124

Silk, J., & Rees, M. J. 1998, A&A, 331, L1

Silverman, B. 1986, Density Estimation for Statistics and Data Analysis (London:Chapman and Hall)

Silverman, J. D., et al. 2005, ApJ, 618, 123

Sobolewska, M. A., Siemiginowska, A., & Życki, P. T. 2004a, ApJ, 608, 80

Sobolewska, M. A., Siemiginowska, A., & Życki, P. T. 2004b, ApJ, 617, 102

Soltan, A. 1982, MNRAS, 200, 115

Spergel, D. N., et al. 2003, ApJS, 148, 175

Springel, V., Di Matteo, T., & Hernquist, L. 2005, ApJL, 620, L79

Stalin, C. S., Gopal-Krishna, Sagar, R., & Wiita, P. J. 2004, MNRAS, 350, 175

Stalin, C. S., Gupta, A. C., Gopal-Krishna, Wiita, P. J., & Sagar, R. 2005, MNRAS, 356, 607

Stapleton, D. C. & Young, D. J. 1984, Econometrica, 52, 737

Steed, A., & Weinberg, D. H. 2006, submitted to ApJ(astro-ph/0311312)

Steffen, A. T., Barger, A. J., Cowie, L. L., Mushotzky, R. F., & Yang, Y. 2003, ApJL, 596, L23

Steffen, A. T., Strateva, I., Brandt, W. N., Alexander, D. M., Koekemoer, A. M., Lehmer, B. D., Schneider, D. P., & Vignali, C. 2006, AJ, 131, 2826

Storrie-Lombardi, L. J., McMahon, R. G., Irwin, M. J., & Hazard, C. 1996, ApJ, 468, 121

Strateva, I. V., Brandt, W. N., Schneider, D. P., Vanden Berk, D. G., & Vignali, C. 2005, AJ, 130, 387 (S05)

Tananbaum, H., et al. 1979, ApJL, 234, L9

Tremaine, S., et al. 2002, ApJ, 574, 740

Tremonti, C. A., Moustakas, J., & Diamond-Stanic, A. M. 2007, ApJL, 663, L77

Trèvese, D., Kron, R. G., & Bunone, A. 2001, ApJ, 551, 103

Trèvese, D., & Vagnetti, F. 2002, ApJ, 564, 624

Trump, J. R., et al. 2006, ApJS, 165, 1

Trump, J. R., et al. 2007, ApJS, 172, 383

Trump, J., Impey, C., et al. 2008, in preparation

Turner, N. J. 2004, ApJL, 605, L45

Ueda, Y., Akiyama, M., Ohta, K., & Miyaji, T. 2003, ApJ, 598, 886

Ulrich, M.-H., Maraschi, L., & Urry, C. M. 1997, ARA&A, 35, 445

Ulvestad, J. S., Antonucci, R. R. J., & Barvainis, R. 2005, ApJ, 621, 123

Uttley, P., McHardy, I. M., & Papadakis, I. E. 2002, MNRAS, 332, 231

Vanden Berk, D. E., et al. 2004, ApJ, 601, 692

van der Klis, M. 1997, Statistical Challenges in Modern Astronomy II, 321

Vasudevan, R. V., & Fabian, A. C. 2007, MNRAS, 381, 1235

Véron-Cetty, M.-P., Joly, M., & Véron, P. 2004, A&A, 417, 515

Vestergaard, M. 2002, ApJ, 571, 733

Vestergaard, M. 2004, ApJ, 601, 676

Vestergaard, M. 2006, New Astronomy Review, 50, 817

Vestergaard, M., 2008, in preparation

Vestergaard, M., et al., 2008, submitted to ApJL

Vestergaard, M., & Peterson, B. M. 2006, ApJ, 641, 689

Vestergaard, M., & Wilkes, B. J. 2001, ApJS, 134, 1

Vignali, C., Brandt, W. N., Fan, X., Gunn, J. E., Kaspi, S., Schneider, D. P., & Strauss, M. A. 2001, AJ, 122, 2143

Vignali, C., Brandt, W. N., Schneider, D. P., Garmire, G. P., & Kaspi, S. 2003, AJ, 125, 418

Vignali, C., Brandt, W. N., & Schneider, D. P. 2003, AJ, 125, 433

Vignali, C., et al. 2003, AJ, 125, 2876

Vignali, C., Brandt, W. N., Schneider, D. P., & Kaspi, S. 2005, AJ, 129, 2519

Vignali, C., Comastri, A., Cappi, M., Palumbo, G. G. C., Matsuoka, M., & Kubo, H. 1999, ApJ, 516, 582

Waddington, I., Dunlop, J. S., Peacock, J. A., & Windhorst, R. A. 2001, MNRAS, 328, 882

Wandel, A. 2000, New Astronomy Review, 44, 427

Wandel, A., Peterson, B. M., & Malkan, M. A. 1999, ApJ, 526, 579

Wang, J.-M., Chen, Y.-M., & Zhang, F. 2006, ApJL, 647, L17

Wang, J.-M., Watarai, K.-Y., & Mineshige, S. 2004, ApJL, 607, L107

Wang, R., et al. 2008, AJ, 135, 1201

Wang, T.-G., Zhou, Y.-Y., & Gao, A.-S. 1996, ApJ, 457, 111

Warner, C., Hamann, F., & Dietrich, M. 2003, ApJ, 596, 72

Warren, S. J., Hewett, P. C., & Osmer, P. S. 1994, ApJ, 421, 412

Weiner, B. J., et al. 2006, ApJ, 653, 1049

Weiss, A. A., 1993, J. Econometrics., 56, 169.

Wilhite, B. C., Brunner, R. J., Grier, C. J., Schneider, D. P., & vanden Berk, D. E. 2008, MNRAS, 383, 1232

Wilkes, B. J., & Elvis, M. 1987, ApJ, 323, 243

Wilkes, B. J., Tananbaum, H., Worrall, D. M., Avni, Y., Oey, M. S., & Flanagan, J. 1994, ApJS, 92, 53 (W94)

Willott, C. J., Rawlings, S., Blundell, K. M., Lacy, M., & Eales, S. A. 2001, MNRAS, 322, 536

Wold, M., Brotherton, M. S., & Shang, Z. 2007, MNRAS, 375, 989

Wolf, C., Wisotzki, L., Borch, A., Dye, S., Kleinheinrich, M., & Meisenheimer, K. 2003, A&A, 408, 499

Wolf, C., et al. 2004, A&A, 421, 913

Woo, J., & Urry, C. M. 2002, ApJ, 579, 530

Worrall, D. M., Tananbaum, H., Giommi, P., & Zamorani, G. 1987, ApJ, 313, 596

Wu, X.-B., Wang, R., Kong, M. Z., Liu, F. K., & Han, J. L. 2004, A&A, 424, 793

Wyithe, J. S. B., & Loeb, A. 2003, ApJ, 595, 614

Xie, G.-Z., Chen, L.-E., Li, H.-Z., Mao, L.-S., Dai, H., Xie, Z.-H., Ma, L., & Zhou, S.-B. 2005, Chinese Journal of Astronony and Astrophysics, 5, 463

York, D. G., et al. 2000, AJ, 120, 1579

Yu, Q., & Tremaine, S. 2002, MNRAS, 335, 965

Yuan, W., Brinkmann, W., Siebert, J., & Voges, W. 1998a, A&A, 330, 108

Yuan, W., Siebert, J., & Brinkmann, W. 1998b, A&A, 334, 498

Zamorani, G., et al. 1981, ApJ, 245, 357

Zdziarski, A. A., Lubinski, P., & Smith, D. A. 1999, MNRAS, 303, L11

Zellner, A. 1971, An Introduction to Bayesian Inference in Econometrics (New York:John Wiley & Sons)