The Optical Variability of Quasars as seen by the Sloan Digital Sky Survey

Chelsea L. MacLeod

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Washington

2012

Željko Ivezić, Chair

Eric Agol

Scott Anderson

Program Authorized to Offer Degree:
Astronomy

University of Washington

Abstract

The Optical Variability of Quasars
as seen by the
Sloan Digital Sky Survey

Chelsea L. MacLeod

Chair of the Supervisory Committee:
Professor Željko Ivezić
Astronomy

I provide a quantitative analysis of the database of 3.5 million photometric measurements for 80,000 spectroscopically confirmed quasars recently assembled by the Sloan Digital Sky Survey (SDSS). This database is an excellent data set to study quasar variability, and includes 9,000 well-sampled light curves from Stripe 82 and a 25,000-object two-epoch sample from the northern Galactic sky. I show that the damped random walk model provides a robust statistical description of these data. This model characterizes the variability using two parameters, the long-term amplitude and characteristic timescale, and these model parameters show trends with physical parameters such as the luminosity and black hole mass. This mathematical model supports accretion disk instabilities as the dominant variability mechanism. However, there is substantial scatter around the mean trends, as other sources of stochasticity are yet to be discovered. If magnetohydrodynamic models could be improved to reproduce the observed characteristics of variability, they should be able to shed light on the source of stochasticity inherent in quasar light curves.

In this dissertation, I discuss the application of a quasar variability model to study the underlying physics of accretion disks. I also discuss the prospect of using variability as a selection method for quasars. Lastly, I use results from SDSS to discuss the prospects of studying quasar variability in upcoming large time-domain sky surveys.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Many of the crucial steps in completing this work would not have been possible without the vision and guidance of other individuals, especially Zeljko Ivezic, the man behind the phrase, "a drunk married man's walk." The author wishes to express sincere gratitude to Christopher S. Kochanek and collaborators who made many valuable suggestions and their code available, and were instrumental in our efficiency and success. The author would like to thank Amy Kimball and Branimir Sesar who were valuable resources, along with Andrew C. Becker who maintained the Stripe 82 database. This work was mainly funded by an NSF grant. Local computing power was made available by the University of Washington Student Technology Fund. Finally, I owe so much more to my husband, Matthew Sean Singer, than words can express.

Chapter 1

INTRODUCTION

Quasars are among the most distant objects, detectable over almost the entire electromagnetic spectrum. While interesting objects in and of themselves, they are also cosmologically important through their effects on regulating star formation and ionization throughout the universe. Shortly after their discovery (Matthews & Sandage 1963), it was realized that the energy production in quasars originates from accretion onto a black hole residing in a distant and powerful active galactic nucleus (AGN). Central to this claim were the redshifted broad emission lines (BELs) revealed by spectroscopic studies. The discovery of quasars and their time variability were nearly simultaneous, opening a new avenue for quasar exploration. The observational studies of quasar physics since then can be classified into roughly three domains: spectroscopic studies, which include high-resolution spectroscopy and low-resolution spectral energy distributions (SEDs), the time domain, and spatial resolution, which is only possible for an AGN by indirect means such as gravitational microlensing. Over the last decade, time domain surveys have drastically increased the amount of quasar data, providing unique test beds for their study.

In the current paradigm, a quasar consists of a massive black hole accreting matter from a disk-like structure which powers the luminosity, surrounded by material commonly referred to as a dusty torus, which leads to viewing angle-dependent effects by obscuring parts of the central engine (Antonucci, 1993; Urry & Padovani, 1995; Krolik, 1999). However, there is much debate regarding the detailed structure of this central engine, such as the standard α–disk model, and the relationship between emission in different wavebands. For example, quasar spectra at UV wavelengths are brighter than predicted by the standard thin disk model (e.g., Blaes et al., 2001). Fortunately, one of the defining characteristics of quasars is that their output from γ-rays to radio waves is variable over time, and this time dependence

provides a tool to probe the physics associated with regions very close to the black hole. For the purpose of this dissertation, a quasar is a Type-I broad line AGN, where in the current paradigm, we are viewing the central engine at a relatively low inclination angle so that the central engine is not obscured by the dusty torus.

The variability of quasars is fundamental in that it is related to the main energy generation process (Gaskell & Klimek, 2003). Nevertheless, the cause of the stochastic variations in quasar flux over a large range of timescales and wavelengths has eluded astronomers for $\sim$ 40 years. It became clear early on that optical variability is common to mostly all AGNs (e.g., Lyutyi 1972), and this characteristic is still observed in modern sky surveys (Sesar et al., 2006; Butler & Bloom 2011). The variations are seemingly chaotic or "stochastic," on the order 20%, and on timescales of months to years in the optical. Traditionally, quasar variability has been quantified using a structure function (SF) analysis, where the SF is essentially the rms variability as a function of time between observations. The slope of the SF is able to distinguish between simple physical models behind the variability such as starbursts or gravitational lensing, and accretion disk instabilities offer a promising explanation in this regard (Kawaguchi et al., 1998; Lyubarskii 1997). Although the variability has been shown to be intrinsic to AGNs through reverberation mapping (Peterson, 1993), where the BELs are seen to respond to the continuum variations, they have been difficult to understand because of limited data and the lack of a detailed theoretical context.

1.1 Variability as a Probe into the Physics of Quasars

Much of what we know about accretion disks stems from the studies of microquasars, or binary systems in which a normal star orbits around and loses matter to a nearby compact object, either a black hole or a neutron star. Microquasars exhibit many similar characteristics to quasars such as rapid variability in the X-rays and radio jets, but they are much easier to study due to their close proximity and shorter variability timescales. Due to their smaller masses, microquasars emit their peak energy in the X-rays rather than in the UV as in quasars (which complicates matters slightly because the emission from the corona is also in the X-rays). By obtaining X-ray light curves of microquasars, we have determined that the variability exhibits a red noise power spectrum, where the power declines as a function

of frequency with a power-law index of -2. Below a characteristic frequency, the power spectrum begins to flatten to white noise as the variations become uncorrelated. Studies of variability in the X-rays show that the characteristic frequency is correlated with the mass of the black hole, and furthermore that an additional break to an index of -1 is present in the power spectra for microquasars (McHardy et al., 2006). It has been a primary goal for theorists to relate what we know about microquasar variability to that in accretion disks around supermassive black holes, but the progress has been hampered by the lack of quality data for AGNs. The leading physical model used in magneto-hydrodynamic (MHD) simulations of both types of accretion disks is the magneto-rotational instability (MRI). MRI-based simulations predict a smooth red-noise power spectrum (with index -2) and SF slopes in agreement with observations (Hawley & Krolik, 2001).

Along with comparing SF slopes, another way to use variability to deduce physical characteristics of the source is to analyze the characteristic timescales of variability and compare them with physical timescales associated with the quasar. For example, variability on the light-crossing timescale has been valuable in determining the extent of the variable region for sources such as 3C279, a well-studied blazar. The fastest variability timescale detected in the γ-rays is around 1–3 days for 3C279, indicating that the size of the flaring region is roughly 8 times the size of the Solar System (Kniffen et al., 1993), and even shorter variability timescales are detected in some X-ray AGN studies (Kelly et al., 2011). The light-crossing timescale is also used in optical reverberation mapping to deduce the extent of the broad line region (BLR), as the broad emission lines respond to changes in the continuum flux originating from the disk. However, it is well known that quasars vary on much longer timescales of months to years, suggesting that an additional physical timescale is at work.

The second shortest physical timescale is the dynamical timescale, which encompasses the free-fall, sound-crossing, and orbital timescales. For a $10^8 M_\odot$ black hole, the dynamical timescale is about 100 days. Quasar variations are known to be aperiodic, indicating that the variability timescale should not be associated with the orbital timescale. Slightly longer is the thermal timescale for the accretion disk, which is the natural choice for explaining the observed variability timescales given that the emission from quasars is largely thermal.

Furthermore, it has previously been found that quasars get bluer as they brighten (Giveon et al., 1999), similar to a blackbody. For a $10^8 M_\odot$ black hole and a standard α viscosity parameter of 0.01, the thermal timescale is roughly 5 years. Lastly, there is the viscous timescale, which corresponds to the timescale for radial drift of material inwards within the accretion disk. This timescale is of order 10^3–10^6 years, making it essentially immeasurable in real data.

While the theoretical context is still unclear, quasar variability has been a useful tool in reverberation mapping, for example, to determine the size of the BLR and thus calibrate black hole mass estimates (Zu et al., 2011), or in gravitational lens time delays and thus cosmology (Kundic et al., 1997; Koptelova et al., 2012). However, aside from "non-parametric" statistical characterizations such as structure functions, the variability itself has been of little direct use because we lacked a mathematical way to describe it. In this dissertation, I discuss the application of a quasar variability model in order to help bridge the gap between observations and theory.

1.2 Overview

This is an ideal time to study quasar variability. Whereas before, samples only included ∼100 quasar light curves, sample sizes have jumped by two orders of magnitude since the advent of the Sloan Digital Sky Survey (SDSS). The sample sizes will continue to increase dramatically with upcoming surveys such as the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS), the Dark Energy Survey (DES), and the Large Synoptic Survey Telescope (LSST). The success in extracting science from future time domain data will hinge upon having robust and efficient methods for analyzing light curves.

I began my thesis work by assembling a large and diverse photometric database consisting of two distinct observational sets:

1. A set of fairly well-sampled light curves with several hundred SDSS observations per object (summed over all bands), for ∼9,000 spectroscopically confirmed quasars from SDSS Stripe 82 (S82) which was originally designed to search for supernovae. These light curves span a time range of 10 years.

2. 30,000 pairs of photometric measurements per quasar, with observed timescales ranging from a couple of days to a decade, based on SDSS data. When including data from a unified SDSS-POSS catalog, this set is extended to 80,000 objects with timescales in the range 10-50 years.

In the next three subsections, I highlight the main results from analyzing each data set.

1.2.1 Modeling the Time Variability of SDSS S82 Quasars as a Damped Random Walk

The main advantage of the first set is that variability analysis can be done on an object-by-object basis. Using S82 quasar light curves, I develop a detailed quantitative description of quasar photometric variability using a damped random walk (DRW) model. This model was originally applied in economics (Morley et al., 2003), and it recently entered the field of astronomy to describe quasar variability (Kelly et al., 2009; Kozłowski et al., 2010b). The DRW model provides a simple, fast [O(N) for N data points], and powerful statistical description of quasar light curves by a characteristic timescale (τ) and an asymptotic variability amplitude (SF_∞). The model parameters show trends with physical parameters such as the luminosity, black hole mass, wavelength, and radio-loudness of the quasar, but they exhibit a substantial scatter around the mean trends. We find that the measured timescales are consistent with thermal timescales, but the observed scalings of τ with physical parameters are not consistent with naive theoretical scalings based on a thermal timescale.

Through a Monte Carlo analysis, we estimate the uncertainties in model parameters due to limited time sampling. We also determine that a lack of sufficient long-timescale information in the S82 survey prevents us from distinguishing between a $1/f^0$ or a $1/f$ power-spectral density (PSD) at frequencies $f < (2\pi\tau)^{-1}$. Our results provide a simple quantitative framework for generating mock quasar light curves which can be used in simulations of future time-domain surveys such as the Large Synoptic Sky Survey (LSST). The work presented in Chapter 2 is published in a journal article (MacLeod et al., 2010).

1.2.2 A Description of Quasar Variability Measured Using Repeated SDSS and POSS Imaging

The second, two-epoch data set provides only statistical constraints averaged over appropriately chosen subsamples, but for a much larger data set and over a better-sampled range of timescales. I adopt the DRW model in order to better understand the relationship between the two data sets, and have concluded that the available data cannot reject the DRW model as a viable description of the optical continuum variability of quasars.

The longer timescales probed by the two-epoch set provide additional constraints on quasar variability. In particular, there is evidence that the measured τ and SF_∞ distributions are biased low for the S82 quasars, which can either be attributed to the limited length of the S82 survey or an inaccurate model description on long timescales. This analysis has enabled robust predictions for the incidence of quasar contamination in transient surveys using detailed simulations of quasar light curves from a mock LSST catalog. A publication of the work presented in Chapter 3 is also available (MacLeod et al., 2011a).

1.2.3 Quasar Selection Based on Variability

The results from our S82 analysis enabled the development of selection criteria for the use in quasar identification based on photometric variability, which will play an important role in upcoming large sky surveys such as Pan-STARRS, DES, and LSST. With the aid of an SDSS spectroscopically confirmed quasar sample, we demonstrated that variability selection in typical extragalactic fields with low stellar density can deliver relatively complete samples with reasonable purity. In summary, we determined that given an adequate survey cadence, photometric variability provides an even better method than color selection for separating quasars from stars. The results from Chapter 4 are published in MacLeod et al. (2011b).

Finally, our selection criteria identified several hundred quasar candidates in S82 that vary on long timescales and lacked SDSS spectra. We obtained follow-up spectroscopy using the Dual Imaging Spectrograph on the ARC 3.5m at Apache Point Observatory for a subsample of these objects, targeting those that resemble quasars in other wavelength regimes (IR, UV, X-ray, and radio) with highest priority. The results, along with those from

an SDSS-III ancillary BOSS program in which we obtained follow-up spectra for 1,500 quasar candidates in S82, will provide an invaluable training sample for optimizing quasar selection algorithms based on photometric variability and for identifying the major contaminants in variability-based quasar selections.

Chapter 2

MODELING THE TIME VARIABILITY OF SDSS STRIPE 82 QUASARS AS A DAMPED RANDOM WALK

2.1 Abstract

This Chapter contains the text of the published paper, MacLeod et al. (2010), where we model the time variability of ∼9,000 spectroscopically confirmed quasars in SDSS Stripe 82 as a damped random walk. Using 2.7 million photometric measurements collected over 10 years, we confirm the results of Kelly et al. (2009) and Kozłowski et al. (2010b) that this model can explain quasar light curves at an impressive fidelity level (0.01-0.02 mag). The damped random walk model provides a simple, fast [O(N) for N data points], and powerful statistical description of quasar light curves by a characteristic timescale (τ) and an asymptotic rms variability on long timescales (SF_∞). We searched for correlations between these two variability parameters and physical parameters such as luminosity and black hole mass, and rest-frame wavelength. Our analysis shows SF_∞ to increase with decreasing luminosity and rest-frame wavelength as observed previously, and without a correlation with redshift. We find a correlation between SF_∞ and black hole mass with a power law index of 0.18±0.03, independent of the anti-correlation with luminosity. We find that τ increases with increasing wavelength with a power law index of 0.17, remains nearly constant with redshift and luminosity, and increases with increasing black hole mass with power law index of 0.21±0.07. The amplitude of variability is anti-correlated with the Eddington ratio, which suggests a scenario where optical fluctuations are tied to variations in the accretion rate. However, we find an additional dependence on luminosity and/or black hole mass that cannot be explained by the trend with Eddington ratio. The radio-loudest quasars have systematically larger variability amplitudes by about 30%, when corrected for the other observed trends, while the distribution of their characteristic timescale is indistinguishable from that of the full sample. We do not detect any statistically robust

differences in the characteristic timescale and variability amplitude between the full sample and the small subsample of quasars detected by ROSAT. Our results provide a simple quantitative framework for generating mock quasar light curves, such as currently used in LSST image simulations.

2.2 Introduction

The optical variability of quasars has been recognized since they were first identified (Matthews & Sandage 1963). Indeed, most quasars are variable (∼90% at the 0.03 mag rms level; Sesar et al. 2007), and the variations in brightness are aperiodic and on the order of 20% on timescales of months to years (e.g., Hook et al. 1994; Vanden Berk et al. 2004). Furthermore, the smooth power spectra suggest a chaotic, or stochastic, origin for the variability. A range of models have been advanced to describe quasar variability, including supernova bursts, microlensing, and accretion disk instabilities (Aretxaga et al. 1997; Hawkins 1993; Kawaguchi et al. 1998; Trèvese & Vagnetti 2002). These models are discussed and compared in Hawkins (2007). Reverberation mapping studies (e.g., Peterson et al. 2005) show that the broad emission lines respond to continuum fluctuations, therefore providing strong evidence that the variability is intrinsic to the quasars. A number of studies have utilized standard accretion disk models to demonstrate that the optical-UV variability of quasars could be driven by a variable accretion rate (e.g., Pereyra et al. 2006; Li & Cao 2008; Liu et al. 2008). Blackburne & Kochanek (2010) find evidence in the light curves of microlensed quasars that the optical variability is caused by a change in the effective area of the accretion disk.

Recently, Kelly et al. (2009, hereafter KBS09) proposed a model where the optical variability is described by a damped random walk (a self-correcting term added to a random walk model that acts to push any deviations back towards the mean value). They proposed that the variability timescale might be identified with the thermal timescale of accretion disks, as also proposed by Collier & Peterson (2001). A thermal origin of the variability would explain why quasars become bluer as they brighten (e.g., Giveon et al. 1999; Trèvese et al. 2001; Geha et al. 2003).

Although the physical causes have yet to be proven, it has been established by KBS09 and Kozłowski et al. (2010b, hereafter Kozł10) that a damped random walk can statistically

explain the observed light curves of quasars. Using 100 well-sampled single-band light curves compiled from the literature, KBS09 show that this stochastic process is capable of modeling complex quasar light curves at an impressive fidelity level (0.01-0.02 mag). Kozł10 applied the model to the OGLE light curves (Udalski et al. 1997; Udalski et al. 2008) of mid-infrared-selected quasars behind the Magellanic Clouds from Kozłowski & Kochanek (2009). Their analysis shows that this stochastic model is robust enough to efficiently select quasars from other variable sources (see Schmidt et al. 2010 for a different method of selecting quasars based on variability). The model has only three free parameters: the mean value of the light curve, the driving amplitude of the stochastic process, and the damping timescale. The predictions are only statistical, and the random nature reflects our uncertainty about the details of the physical processes.

Instead of applying a model to observed light curves for individual quasars, numerous studies have looked at the ensemble variability of quasars, particularly in samples where individual light curves are not available. Significant progress in the description of quasar variability has been made by employing the Sloan Digital Sky Survey (SDSS) data (Vanden Berk et al. 2004, hereafter VB04; Ivezić et al. 2004, hereafter I04; de Vries et al. 2005; Wilhite et al. 2005, 2006, 2008; Sesar et al. 2006). For example, the size and quality of the sample analyzed by VB04 (two-epoch photometry for 25,000 spectroscopically confirmed quasars) allowed them to constrain how quasar variability in the rest frame optical/UV regime depends upon rest-frame time lag (up to $\sim$2 years), luminosity, rest wavelength, redshift, the presence of radio and X-ray emission, and the presence of broad absorption line systems. Using repeated SDSS photometric observations, Wilhite et al. (2008) confirmed the result of Wold et al. (2007) that variability is correlated with black hole mass, and show that this is independent of the anti-correlation between variability and luminosity established by many studies. This led them to suggest that the amplitude of variability may be driven by the quasar's Eddington ratio, implying differences in accretion rate.

These studies typically quantify the observed optical variability of quasars using a structure function (SF) analysis (see also Hughes et al. 1992; Collier & Peterson 2001; Bauer et al. 2009; Kozłowski et al. 2010a), where the SF is the root-mean-square (rms) magnitude difference as a function of the time lag (Δt) between measurements. This autocorrelation-

like function is less sensitive to aliasing and other time-sampling problems than a power spectral distribution. By studying the magnitude difference distribution for appropriately chosen subsamples with fixed values of absolute i-band magnitude (M_i), rest-frame time lag (Δt_{RF}, in days), and wavelength (λ_{RF}, in Å), the mean dependence of the SF on these quantities was inferred by I04 to be

$$SF_{model} = A[1 + B\,M_i]\left(\frac{\Delta t_{RF}}{\lambda_{RF}}\right)^{C}\,\mathrm{mag}, \tag{2.1}$$

with $A = 1.00 \pm 0.03$, $B = 0.024 \pm 0.04$, and $C = 0.30 \pm 0.05$. A qualitatively similar result was obtained by VB04. Kozłowski et al. (2010a), in the first large study of the mid-IR structure functions of quasars, also found lower variability for higher luminosities and longer wavelengths, but the temporal slope of the ensemble structure functions were significantly steeper than in the optical. In addition, there is evidence for a turnover in the SF on long time lags (I04; Rengstorf et al. 2006; Wold et al. 2007). Studies by de Vries et al. (2005) and Sesar et al. (2006) using SDSS combined with earlier Palomar Observatory Sky Survey measurements for 40,000 SDSS quasars constrained quasar continuum variability on timescales of 10 to 50 years in the observer's frame. They report that the characteristic timescale, which in this context is the time lag above which the SF flattens to a constant value, is of order 1 year in the quasar rest frame. Using a shot-noise light curve model, de Vries et al. (2005) found evidence for multiple variability timescales in long-term ensemble variability measurements, while Collier & Peterson (2001) found a wide range of different timescales in their analysis of individual light curves, and even evidence for multiple timescales in a single active galactic nucleus (AGN).

These analyses of ensemble variability are based on a fundamental assumption that photometric observations at two epochs for a large number of quasars will reveal the same statistical properties as well-sampled light curves for individual objects. This assumption has been tested by MacLeod et al. (2008) using light curves for spectroscopically confirmed quasars observed roughly 50 times over 8 years in SDSS Stripe 82 (S82). They found that while the mean SF for individual sources is consistent with Eq. 2.1, the contribution of the mean trends to the observed dispersion in variability properties is minor compared to an intrinsic stochasticity of unknown origin. Further investigation of this stochastic behavior

is one of the main goals of this study.

In order to better understand the relationship between the two types of data analyses (individual versus ensemble quasar variability), and to begin linking to physical models, we apply the damped random walk model to the *ugriz* light curves of ∼9,000 spectroscopically confirmed SDSS S82 quasars. This large sample greatly benefits from the robust, accurate, five-band SDSS photometry. We estimate the variability parameters following Kozł10, who demonstrated that their approach is more statistically powerful than the forecasting methods used by KBS09. We also note that both the Kozł10 and KBS09 approaches are much faster than that used by Schmidt et al. (2010), requiring only O(N) rather than O(N^2) operations to determine the model parameters for a light curve with N data points.

In Section 2.3, we describe the model, define our variability parameters, and demonstrate their relationship to those utilized in previous studies. In Section 2.4, we introduce the S82 data set and outline our initial light curve selection. In Section 2.5, we present the best-fit variability parameters for our final sample of light curves and estimate their scatter due to the limited time sampling of SDSS. We also estimate the sensitivity of our results to variations in the slope of the model power spectrum on long timescales. In Section 2.6, we describe the relationship between the long-term variability parameters and physical parameters such as wavelength, absolute magnitude, black hole mass, and Eddington ratio. Using these results, we also provide a prescription for simulating mock quasar light curves. In Section 2.7, we explore the variability properties of subsamples detected at radio and X-ray wavelengths. Finally, we summarize our results in Section 2.8.

2.3 Methodology

We model the time variability of quasars as a stochastic process described by the exponential covariance matrix

$$S_{ij} = \sigma^2 \exp(-|t_i - t_j|/\tau) \tag{2.2}$$

between times t_i and t_j. As detailed by KBS09 and Kozł10, this corresponds to a damped random walk (more specifically, an Ornstein-Uhlenbeck process) with a damping timescale τ,

also called the characteristic timescale, and a long-term standard deviation of variability[1] σ. Following Kozł10, we model the light curves and estimate the parameters and their uncertainties using the method of Press et al. (1992), its generalization in Rybicki & Press (1992), and the fast computational implementation described in Rybicki & Press (1994). Kozł10 demonstrate that this approach is more statistically powerful than the forecasting methods used by KBS09, while still having computation times scaling linearly with the number of data points.

2.3.1 Structure Function for the Damped Random Walk Model

For our analysis, we express the long-term variability in terms of the structure function (SF) in order to relate to our previous studies and to those of two-epoch samples. The first order SF is $SF^2(\Delta t) = 2\sigma^2[1 - ACF(\Delta t)]$ (e.g., Hughes et al. 1992), where the autocorrelation function for a damped random walk is $ACF(\Delta t) = \exp(-|\Delta t|/\tau)$ (Eq. A6 in KBS09). This ACF results in the structure function[2]:

$$SF(\Delta t) = SF_\infty (1 - e^{-|\Delta t|/\tau})^{1/2}. \tag{2.3}$$

Asymptotic values of the SF at large and small Δt are

$$\begin{aligned} SF(\Delta t >> \tau) &\equiv SF_\infty = \sqrt{2}\sigma \\ SF(\Delta t << \tau) &= \sigma\sqrt{\frac{2|\Delta t|}{\tau}} = SF_\infty \sqrt{\frac{|\Delta t|}{\tau}} \end{aligned} \tag{2.4}$$

The form $SF \propto |\Delta t|^\beta$ is equivalent to a power spectral distribution $PSD \propto f^\alpha$, where $\alpha = -2\beta - 1$ (see Appendix of Bauer et al. 2009; KBS09). The SF at small time lags is therefore equivalent to a power spectral distribution $PSD \propto f^{-2}$. We adopt SF_∞ and τ as our two main variability model parameters.

---

[1]The σ used here is related to the σ used in KBS09 (σ_{KBS}) and the parameter $\hat{\sigma}$ used in Kozł10 as $\sigma_{KBS} = \hat{\sigma} = \sigma\sqrt{2/\tau}$.

[2]The functional form of SF(Δt) fit to the long-term SDSS-POSS data in Sesar et al. (2006), see their Eq. 5, is similar but not identical to the functional form given by Eq. 2.3.

2.3.2 Model Light Curves

Equipped with a statistical description of quasar variability, we generate well-sampled light curves in order to 1) demonstrate the relationship between our variability parameters and the traditional SF analyses of many previous works, and 2) to estimate the systematic effects that the sampling rate and light curve length have on the fitted parameters. The latter is especially important because the S82 data are fairly sparse. As shown below and in Section 2.5.2, these indeed have a large impact.

A light curve is generated using only three input parameters: SF_∞, τ, and the mean value of the light curve, μ. The magnitude $X(t)$ at a given timestep Δt from a previous value $X(t-\Delta t)$ is drawn from a normal distribution with a mean and variance given by

$$E(X(t)|X(t-\Delta t)) = e^{-\Delta t/\tau}X(t-\Delta t) + \mu(1-e^{-\Delta t/\tau})$$
$$Var(X(t)|X(t-\Delta t)) = 0.5(SF_\infty)^2(1-e^{-2\Delta t/\tau}) \quad (2.5)$$

(Eqs. A4 and A5 in KBS09). The asymptotic variance of the time series is then $0.5(SF_\infty)^2$. The top panel of Figure 2.1 shows a segment of a well-sampled light curve generated using $\tau = 20$ days, $SF_\infty = 0.14$ mag, and a time sampling of $0.1(\Delta t/\tau)$. The structure function, $SF(\Delta t)$, is computed by collecting the differences in magnitude for all points in the light curve separated by a given time lag, Δt. The distribution of magnitude differences (Δm) is Gaussian by construction, and the rms of the Δm distribution is the SF value for that time lag.

When fitting values for SF_∞ and τ for a given light curve, the length of time that it spans plays an important role. For example, in Figure 2.1, the SF computed for the full light curve length of 1500τ (82 years) is much smoother than that computed for three equal sections, each spanning 27 years. Therefore, the determination of variability parameters for S82 quasars will be affected by their light curve lengths, which are typically 10τ. When the light curve is too short, it is easy to overestimate τ because there is no information on the timescale of the break (where $SF(\Delta t)$ flattens to SF_∞). The model will reproduce the observed variance in the light curve by overestimating $SF_\infty \simeq (SF_\infty)_{true}(\tau/\Delta t)^{1/2}$, and therefore $\hat{\sigma} = SF_\infty/\sqrt{\tau}$ is the more robustly estimated model parameter when τ cannot be well-determined.

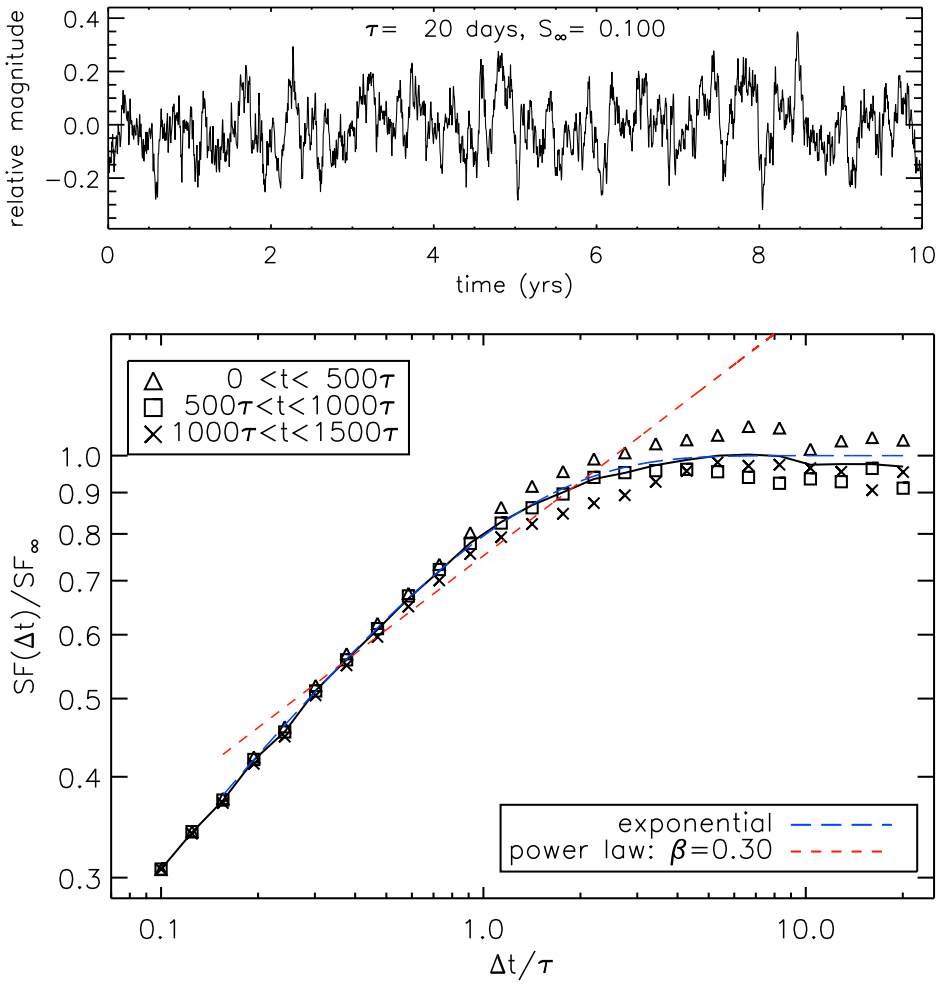

Figure 2.1 The top panel shows a segment of a simulated light curve with $\tau = 20$ days and $SF_\infty = 0.10$ mag. In the bottom panel, triangles, squares, and crosses represent the SF computed for the first, second, and third sections of the total light curve, respectively (each has a length of 500τ). The increased scatter in $SF(\Delta t)$ at large time lags is due to the finite length of each light curve section – the scatter decreases when using the entire light curve length of 1500τ (as shown by the solid line). The short-dashed line is a power law fit, $SF(\Delta t) \propto \Delta t^\beta$ with $\beta = 0.3$, to the data points with $0.15 < \Delta t/\tau < 3$. The long-dashed line is the true function $SF = SF_\infty[1 - e^{-\Delta t/\tau}]^{1/2}$.

2.3.3 Comparison with Published Work

Before interpreting the form of $SF(\Delta t)$, we summarize the major differences between our S82 analysis and previous studies based on ensemble structure functions. The ensemble

structure function is computed using only a few observations of many quasars, combining all magnitude differences to find $SF(\Delta t)$. Using an ensemble approach is beneficial because it enables one to constrain the average variability properties when it is difficult to constrain such quantities for individual quasars. Indeed, even with well-sampled light curves, spurious breaks in the individual SFs are common (Emmanoulopoulos, McHardy, & Uttley 2010). However, in previous works (e.g., I04; de Vries et al. 2005), the characteristic timescale is defined as the time lag at which the ensemble $SF(\Delta t)$ flattens to a constant value, and thus it represents some complex average over the intrinsic τ distribution. In contrast, by applying a stochastic model to the individual S82 light curves, we are relatively insensitive to time sampling issues (for details, see KBS09), and we obtain a model fit for every quasar in each filter described by the parameters SF_∞ and τ.

A power law fit to $SF(\Delta t)$ has previously been a common way to describe $SF(\Delta t)$, and even to reject certain classes of models (e.g., Kawaguchi et al. 1998; VB04; I04). However, Figure 2.1 shows that the best-fit power law index is extremely sensitive to the fitted range of $\Delta t/\tau$. For example, for $0.1 < \Delta t/\tau < 1$, $SF(\Delta t)$ is well fit by a power law with an index of $\sim$0.5, while for $0.15 < \Delta t/\tau < 3$, we obtain a strongly biased power law index of 0.3. Furthermore, each quasar has its own values of τ and SF_∞, and the ensemble structure function is a convolution of the individual structure functions with their distribution in parameters,

$$SF(\Delta t) = \int \mathrm{d}\tau \mathrm{d}SF_\infty \frac{\mathrm{d}^2 n}{\mathrm{d}\tau \mathrm{d}SF_\infty} SF(\Delta t|\tau, SF_\infty), \tag{2.6}$$

where $SF(\Delta t|\tau, SF_\infty)$ (Eq. 2.3) is the structure function at time Δt for a quasar with variability parameters τ and SF_∞. Hence, the ensemble structure function is only indirectly related to the structure function for any particular quasar, and results based on fitting a power law to observed ensemble structure functions should be interpreted with caution.

2.4 The SDSS Stripe 82 Quasar Data Set

The Sloan Digital Sky Survey (SDSS, York et al. 2000) provides homogeneous and deep ($r < 22.5$) photometry in five passbands (*ugriz*, Fukugita et al. 1996; Gunn et al. 1998; Smith et al. 2002) accurate to 0.02 mag, of almost 12,000 deg^2 in the Northern galactic cap, and a smaller, but deeper, survey of 290 deg^2 in the Southern galactic hemisphere. For this

290 deg^2 area known as Stripe 82 (S82), there are on average more than 60 available epochs of observations. These data were obtained in yearly "seasons" about 2-3 months long over the last decade, and the cadence effectively samples timescales from days to years. The light curve lengths are effectively shorter than the actual period of the survey because the better-sampled supernova observations begin about 5 years into the survey. Because some observations were obtained in non-photometric conditions, improved calibration techniques have been applied to SDSS S82 data by Ivezić et al. (2007) and Sesar et al. (2007), and we use their results. For these data, photometric zero-point errors are 0.01–0.02 mag.

We have compiled a sample of 9,275 spectroscopically confirmed quasars in S82 with re-calibrated *ugriz* light curves (see also Bhatti et al. 2010). Most (8,974) of these are in the SDSS Data Release 5 (DR5) Quasar Catalog (Schneider et al. 2007), and the remaining are newly confirmed DR7 (Abazajian et al. 2009) quasars. Summed over all bands and epochs, the data set includes 2.7 million photometric measurements. For 41% of the sample, the random photometric errors are smaller than 0.03 mag. Only 1% have errors ≥ 0.1 mag in g, r, and i, and 2.4% have errors exceeding 0.25 mag in u and z filters. In MacLeod et al. (2011a), these light curves, as well as a much larger sample of quasars with two SDSS epochs selected from 12,000 deg^2 of the sky, are made publicly available. We adopt the K-corrected i-band absolute magnitudes from Schneider et al. (2007), and virial black hole masses and bolometric luminosities where available from Shen et al. (2008). The Shen et al. masses were estimated from emission line widths (Hβ for $z < 0.7$, MgII for $0.7 < z < 1.9$, and CIV for $z > 1.9$). However, we note that at low spectroscopic signal-to-noise, black hole masses tend to be overestimated (Denney et al. 2009).

2.4.1 Initial Light Curve Selection

The damped random walk model was fit to all available *ugriz* light curves for $9,275$ S82 quasars. Summed over 5 bands, there are $46,375$ best-fit values of the characteristic (damping) timescale τ and long-term structure function SF_∞. For further analysis, we select light curves that satisfy the following criteria:

1. First, we remove light curves with fewer than 10 observations. The top-left panel of

Figure 2.2 shows the distribution of the number of observations (N_{obs}) per light curve before this cut, which reduces our sample to 45,814. At a given N_{obs}, the distribution of the ratio of light curve length to τ is similar to that at all other values of N_{obs}. Therefore, any systematic effects of N_{obs} on derived parameter distributions should be small.

2. We then require that the stochastic model must provide a better fit than uncorrelated noise. Following Kozł10, we select light curves with a likelihood improvement over simply broadening the measurement errors of $\Delta L_{noise} \equiv \ln(L_{best}/L_{noise}) > 2$, where L_{best} is the likelihood of the stochastic model and L_{noise} is that for the noise solution where $\tau \to 0$. The top-right panel of Figure 2.2 shows the distribution of ΔL_{noise} before this cut, which removes 14% of u and z light curves, whose photometric errors are larger. About 7% of our light curves (82% of which are u or z band) are removed in this step, reducing our total to 42,623.

3. Finally, we remove cases where τ is merely a lower limit due to the length of the light curve. We define $\Delta L_{\infty} \equiv \ln(L_{best}/L_{\infty})$, where L_{∞} is the likelihood that $\tau \to \infty$, indicating that the light curve length is too short to accurately measure τ. The bottom-left panel shows a peak at $\Delta L_{\infty} = 0$, and we exclude these objects by requiring that $\Delta L_{\infty} > 0.05$. Most (95%) of the rejected light curves have lengths $< \tau$; 76% have $\tau \geq 10^4$ days, and 64% have $SF_{\infty} \geq 1$ mag. The latter is due to the fact that as the τ value becomes long and uncertain, the model will necessarily overestimate SF_{∞} in order to keep the overall light curve rms fixed (see Section 2.3.2).

   The rejected light curves tend to be higher redshift quasars because stronger time dilation leads to shorter rest-frame light curve lengths, making it increasingly difficult to constrain long rest-frame τ. This criterion removes 22% of our sample, leaving a total of 33,112 values of τ and SF_{∞}. Because we are limited by the duration of the S82 survey, this is a significant loss, and therefore our final τ distribution is biased low, but the bias should not be significant considering our results in Section 2.5.2.

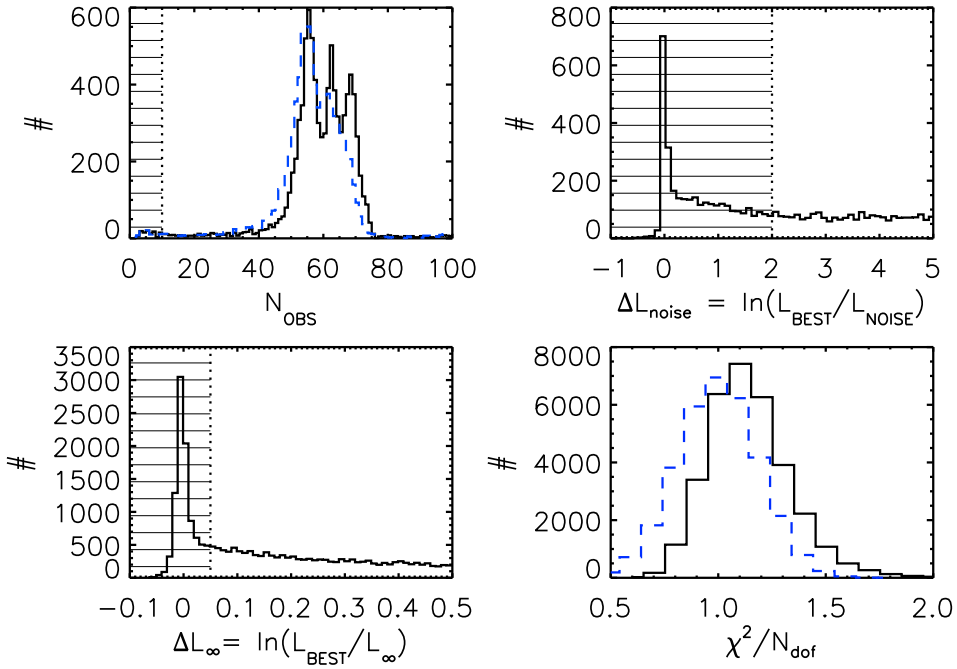

Figure 2.2 Initial light curve selection. *Top-left*: the distribution of the number of observations per light curve for the r (solid) and u (dashed) bands. *Top-right*: distribution of ΔL_{noise}. We define light curves with $\Delta L_{noise} \leq 2$ to be more consistent with uncorrelated noise rather than our model. *Bottom-left*: distribution of ΔL_{∞}; light curves with $\Delta L_{\infty} \leq 0.05$ likely have run-away timescales. In this panel and the previous panel, the x-axes are truncated at 0.5 and 5 respectively, but the histograms continue to greater values. *Bottom-right*: distribution of χ^2 per degrees of freedom (N_{dof}) for the damped random walk model (solid line). The expected Gaussian distribution based on N_{dof} is also shown (dashed). The hashed region in each panel shows the values rejected from our final sample.

The resulting light curves are well-fit by the stochastic model, as can be seen from the distribution of χ^2/N_{dof} shown in the bottom-right panel in Figure 2.2, where N_{dof} is the number of degrees of freedom. The expected Gaussian distribution with rms $= \sqrt{2/N_{dof}}$ is also shown in the panel, where we have averaged over the N_{dof} distribution of the light curves. The observed distribution is centered at $\chi^2/N_{dof} = 1.1$. This difference is some combination of uncertainty in the estimated errors, outliers in the light curves, and any poorly modeled physics. Kozł10 noted a similar difference in their analysis of OGLE light curves. Only 5% of the light curves have $\chi^2/N_{dof} > 1.5$, confirming that most quasars are variable (at the $\Delta L_{noise} > 2$ level), and that a damped random walk is a good description

of quasar variability.

2.5 Variability Parameters for Stripe 82 Quasars

2.5.1 Observed Distributions

We assume the variability to be intrinsic to the quasars and convert the timescales to the rest frame (dividing by $(1 + z)$) before further analysis. The long-term structure function should be independent of redshift other than through evolution in physical parameters, and this view is confirmed by KBS09. Figure 2.3 shows the distributions in SF_∞ and rest-frame τ found for the S82 quasars. If we consider only the brighter ($i < 19$) quasars, based on the best-fit mean magnitude, the distributions are generally similar but biased towards lower asymptotic amplitudes (peaked at 0.12 mag). The distributions are consistent with what was found in KBS09 and Kozł10; however, in these studies, there are not as many objects with runaway ($\tau \to \infty$) timescales because of the improved time sampling of the OGLE survey and many of the light curves in the KBS09 sample. Also, the observed-frame τ distribution lacks many of the short timescales observed in Kozł10; this is likely due to either the better time sampling of the OGLE light curves or unrecognized stellar contamination in their sample.

The bottom panel of Figure 2.3 shows that the best-fit variability parameters are highly correlated with each other, indicating that quasars with larger asymptotic amplitudes of variability also have longer characteristic timescales. We fit a power law slope of 1.3 ± 0.01 dex/dex for all 33,112 data points, and this trend persists within each *ugriz* band as well. Note that a correlation between τ and SF_∞ is expected even if τ is independent of the driving amplitude of short-term variations, $\hat{\sigma} = SF_\infty/\sqrt{\tau}$. Since the power law slope of 1.3 is steeper than that expected if τ and $\hat{\sigma}$ are independent ($\sim$0.2, see below), the timescale must also be correlated with the amplitude of short-term variability. We confirm that this correlation is intrinsic, rather than an artifact of short light curve lengths, in Section 2.5.2.

The approximate values for τ and SF_∞ can often be guessed from light curves by simple visual inspection. Figure 2.4 shows representative light curves from several regions of $\tau - SF_\infty$ space, indicated by stars in Figure 2.3. The weighted average of all model light curves

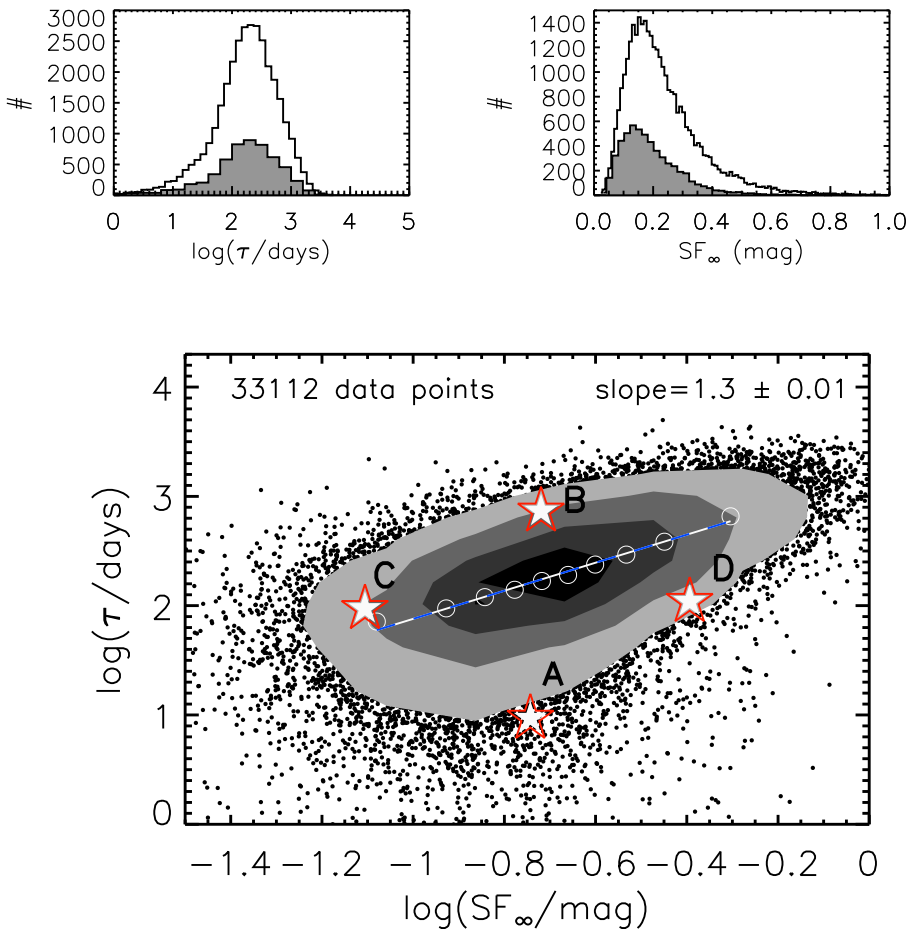

Figure 2.3 *Top*: Distributions of the best-fit rest-frame timescale τ (*left*) and long-term structure function SF_∞ (*right*) for the S82 quasars (33,112 light curves). The filled histograms show the fits for quasars with absolute i-band magnitude in the range $-27 < M_i < -26$ (the median M_i for the whole sample is -25.46). *Bottom*: Relationship between τ and SF_∞. A power law shown by the dashed line is fit to the medians (open white circles), with the slope listed in the top-right corner. Contours show regions containing 90%, 70%, 50%, and 20% of the total number of points. The star symbols label the four regions in parameter space from which the sample light curves in Fig. 2.4 (labeled A through D) were chosen. The observed-frame τ distribution lacks many of the short timescales observed in Koz10. This is likely due to either the better time sampling of the OGLE light curves or stellar contamination in their sample.

consistent with the data is also shown, and the "error snake" is the $\pm 1\sigma$ range of those light curves about this mean. The top panel shows a light curve with a relatively short τ – this is

due to the large amount of variability within each season (i.e. each grouping of points), as compared to that in the second panel, which shows a curve with a much larger τ. The third panel shows a light curve with a relatively low SF_∞, while the bottom panel shows one with a larger SF_∞, as can be seen by the larger difference in median r band magnitude between about 10 and 11 years. Note the outlier in the third panel: outliers such as these are generally responsible for higher values of χ^2/N_{dof}. The (data−model) brightness difference provides a convenient way to identify outliers, and they may be an indicator of variability behavior not captured by the damped random walk model. However, they may also be caused by occasional non-Gaussian photometric errors and thus their analysis requires a detailed and careful study (e.g., by utilizing control samples of appropriately chosen nearby non-variable stars). Since the model provides satisfactory fits for the overwhelming majority of objects (see Section 2.4.1), we do not further investigate such outliers in this work. We have also searched for periodic signals in observed light curves and did not find any convincing cases. This analysis is summarized in the Appendix.

2.5.2 The Effect of S82 Time Sampling and Estimate of Fitting Errors

There are three contributions to the scatter in the best-fit variability parameters τ and SF_∞. First, the fitting errors, including those due to insufficient time sampling and light curve length, will introduce some scatter. Second, trends with physical parameters (see Section 2.6) result in a certain distribution width. There may also be scatter due to other sources of variability that are not captured by the model, such as flares, or other activity related to radio emission, for example. Here, we carry out two tests in order to understand how the best-fit parameters are affected by the limited data sampling of S82.

Since the correlation between τ and SF_∞ seen in Figure 2.3 might be expected if the light curves are not sampled over sufficiently long periods of time (see Section 2.3.2), we first test whether the correlation is real or simply an artifact. We generated light curves as described in Section 2.3.2 using the time sampling and photometric uncertainties of the S82 data. For each object, the input parameters τ_{in} and $SF(\infty)_{in}$ are randomly drawn from a uniform distribution in $\log\tau$ and $\log SF_\infty$ limited by the dotted rectangle in the left panel

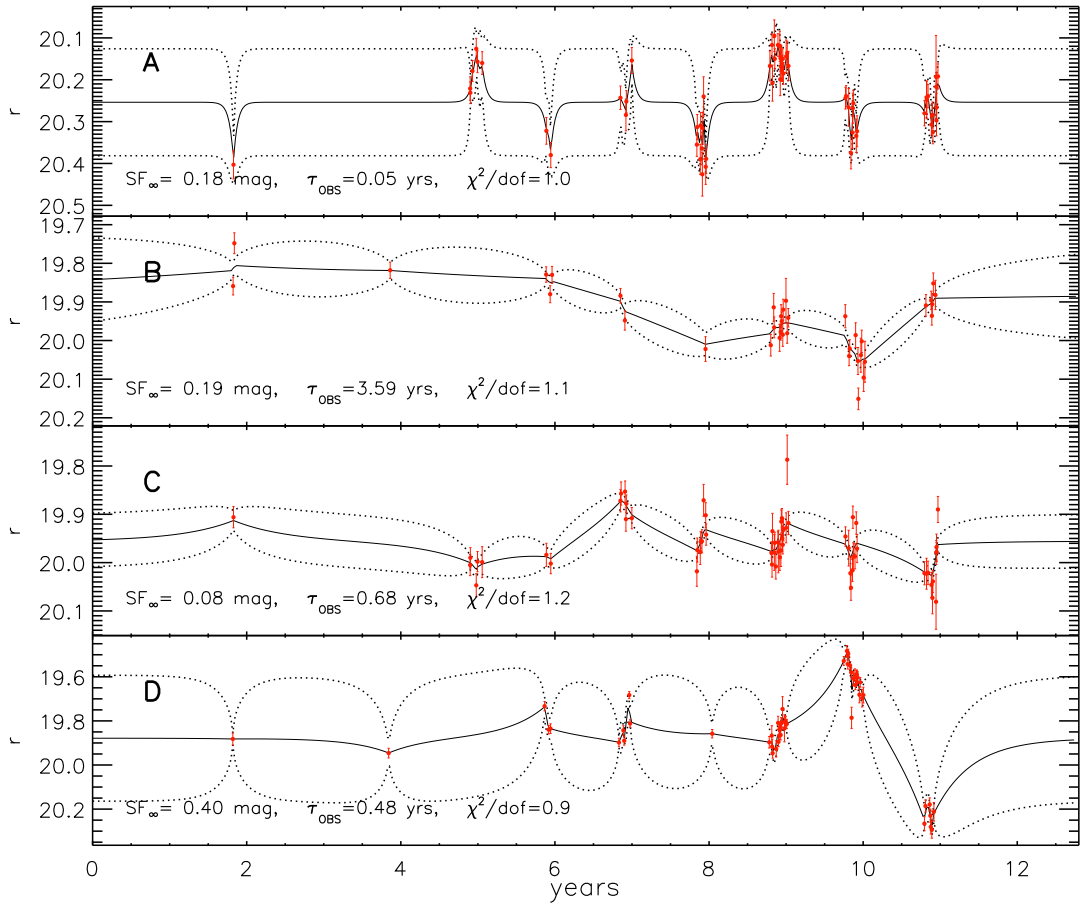

Figure 2.4 Each panel shows a sample light curve from a certain region of τ, SF_∞ space (the four regions A–D are indicated with stars in the bottom panel of Figure 2.3). Data points with error bars show the observed S82 data. Solid lines show the weighted average of the damped random walk model light curves that are consistent with the data (see Section 2.3). Dotted lines show the $\pm 1\sigma$ range of these possible stochastic models about the average.

of Figure 2.5. These artificial light curves are then fit to obtain τ_{out} and $SF(\infty)_{out}$, and the resulting distribution is shown by the contours in Figure 2.5. The open circles show data points that do not satisfy $\Delta L_{noise} > 2$ (3% of all input points) – these are concentrated at small τ_{out} and $SF(\infty)_{out}$. The closed circles show those that do not satisfy $\Delta L_\infty > 0.05$ (21% of all input points) – these have been smeared to large values of τ_{out} and $SF(\infty)_{out}$. After omitting points with $\Delta L_{noise} \leq 2$ and $\Delta L_\infty \leq 0.05$, the distribution of the output

estimates is similar to the input distribution and shows no strong correlation between τ_{out} and $SF(\infty)_{out}$. This suggests that the correlation in Figure 2.3 is largely real, and not an artifact of sampling and fitting procedures. These results also justify the selection cuts outlined in Section 2.4.1. The right panel of Figure 2.5 shows the expected correlation that the larger the overestimate of τ, the larger the overestimate of SF_∞, with a slope between them following that expected by Eq. 4.3 (for $\Delta t << \tau$). We repeated the test using a uniform input distribution in $\log \hat{\sigma}$ (the driving amplitude of short-term variations) rather than in $\log SF_\infty$, where $\hat{\sigma} = SF_\infty/\sqrt{\tau}$. In this case, we find the output τ and SF_∞ are correlated with a power law slope of 0.18 ± 0.01. Since this slope is smaller than that for the observed distribution in Figure 2.3, the $\hat{\sigma}$ and τ must be intrinsically correlated for the S82 sample as well.

For our second test, we used the best-fit τ, SF_∞, and μ for the S82 sample to generate new light curves with the same time sampling and photometric uncertainties as the S82 data. By comparing the output and input parameter distributions, we can estimate how much the intrinsic stochasticity and the time sampling issues affect the results. Figure 2.6 compares an observed and "regenerated" light curve, where the differences are due to the stochastic nature of the process. The fit parameters for the "regenerated" light curve can be very different because of how the particular realization is affected by the time sampling, as illustrated in Figure 2.6. Figure 2.7 shows the ratio of output to input distributions for both τ and SF_∞. The input distributions normalized by their median values are also shown to illustrate their dynamic range. These two distributions are compared to each other in order to estimate the effect of fitting errors (the ratio of output to input should be a delta function centered at 1 for perfect time sampling). The bottom-right panel shows that the correlation between τ and SF_∞ becomes slightly weaker than that seen in Figure 2.3. Based on these results, we conclude that the uncertainties due to sparse sampling and limited lengths of the S82 light curves can account for 71% of the spread in SF_∞ and 57% of that in τ. As shown in Section 2.3.2, very long light curve lengths are needed to estimate accurate timescales and asymptotic amplitudes. Nevertheless, the observed distributions indicate that the underlying intrinsic distributions of τ and SF_∞ have finite widths that are similar to the observed widths.

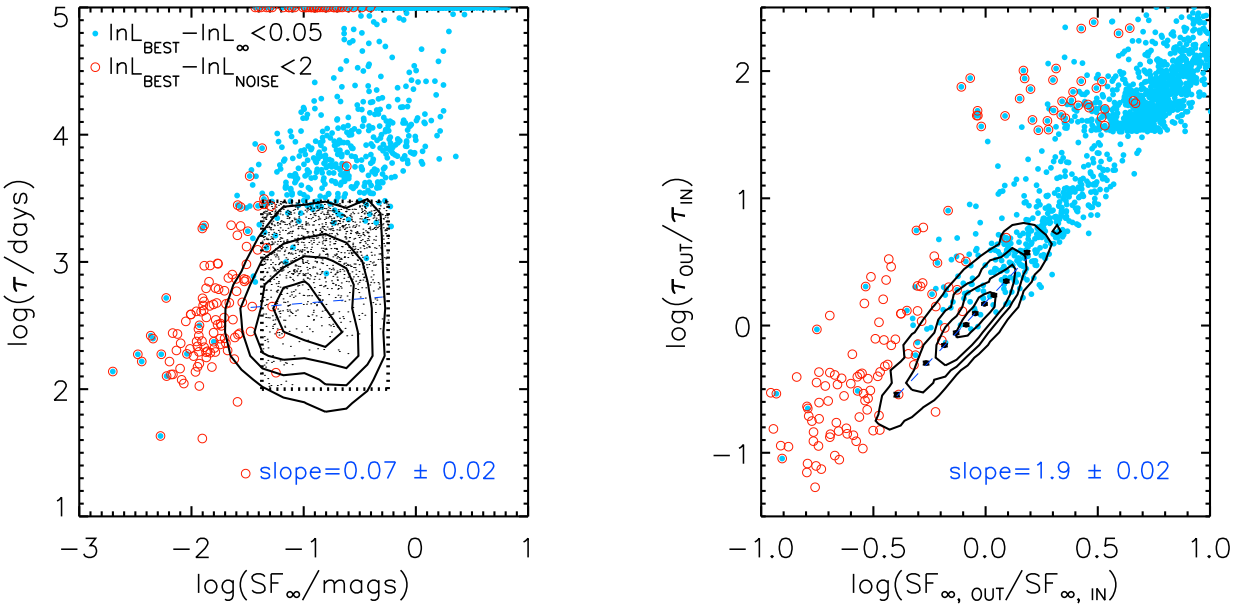

Figure 2.5 Test of biases in best-fit parameters. *Left*: The input variability parameters τ and SF_∞ were drawn from a uniform distribution in $\log \tau$ and $\log SF_\infty$ limited by the dotted rectangle. The contours show the output distribution after applying the cuts described in Section 2.4.1, and the dashed line is a linear regression between the output parameters (with the slope listed on the panel). Solid, blue circles show the objects which do not satisfy $\Delta L_\infty > 0.05$ (these make up 21% of the starting sample, where 13% are saturated at $\tau = 10^5$ days). The small black dots show the input values for these rejected points. Open, red circles show the 3% that do not satisfy $\Delta L_{noise} > 2$. *Right:* Relationship between τ_{out}/τ_{in} and SF_{out}/SF_{in}. The slope of the cleaned output distribution is listed. The sharp edge is due to the saturation limit of $\tau = 10^5$ days. The contours show the 90%, 70%, 50%, and 20% levels.

2.5.3 Relationship between the Individual SFs and the Ensemble SF

The distribution of τ is important to consider when relating the ensemble SF, such as those determined using two-epoch datasets, to the SFs for individual light curves. The analysis based on two-epoch data measures *the mean value* of the SF, but provides *no information* about the SF variance among individual objects. To measure the latter, individual light curves must be available.

MacLeod et al. (2008) analyzed individual light curves for S82 quasars in order to test

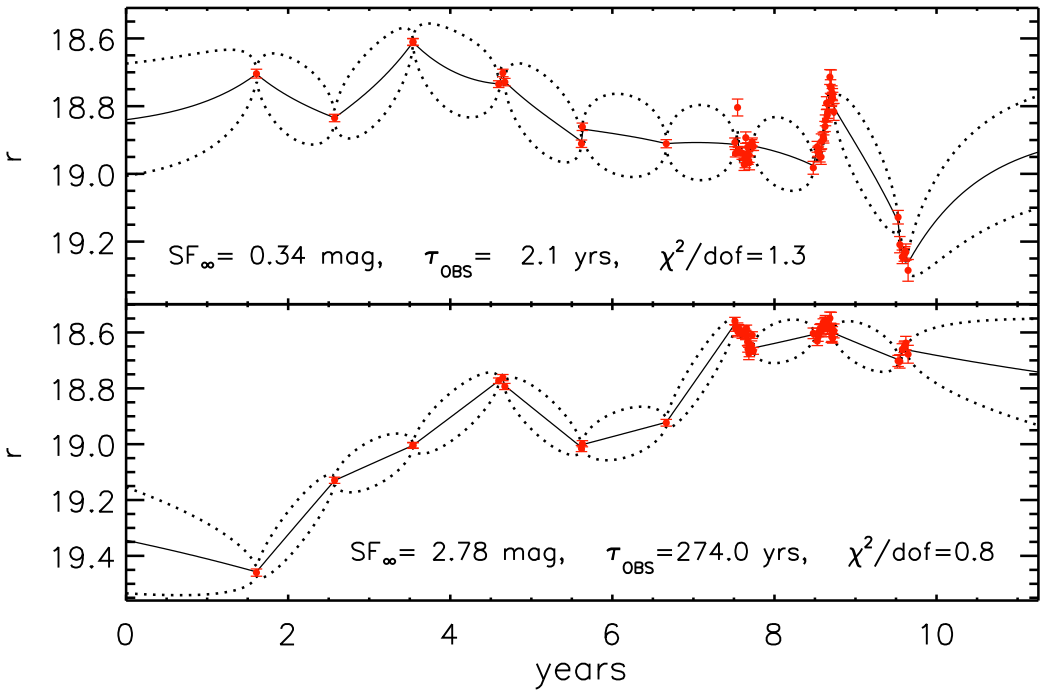

Figure 2.6 *Top panel*: Data points with error bars show the observed light curve for a quasar in our final sample, the solid line shows the weighted average of all consistent model light curves, and the dotted lines show the $\pm 1\sigma$ range of possible stochastic models (see Section 2.3). *Bottom panel*: A "regenerated" light curve using Eq. 2.5 and the estimated values of τ and SF_∞ listed in the top panel. Due to the poor time sampling of the light curve, the best fit to the "regenerated" light curve has a run-away timescale (listed at the bottom), with $\Delta L_\infty = 0.003$.

the common assumption that photometric observations at two epochs for a large number of quasars will reveal the same statistical properties seen in light curves for individual objects. They found that the dependence of the mean SF computed using SFs for individual light curves on luminosity, rest-frame wavelength and time lag is indeed qualitatively and quantitatively similar to that derived from two-epoch observations of a much larger sample. However, they also found that the scatter in the light-curve based SFs for fixed values of M_i, λ_{RF} and Δt_{RF} is very large, and in fact, similar to the scatter for the whole sample (see Figure 2.8). This large scatter was attributed to an intrinsic stochasticity of unknown origin. The new model-based analysis discussed here allows us to explain this puzzling result

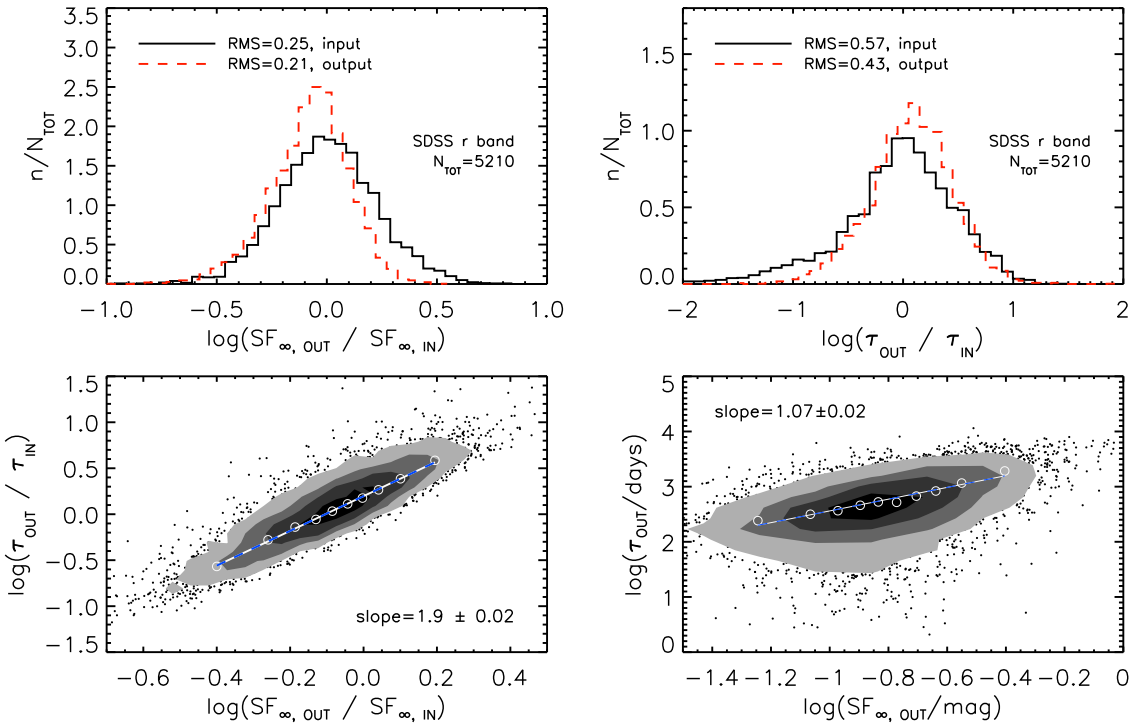

Figure 2.7 Results from "regenerating" light curves derived from the observed τ and SF_∞ (see Section 2.5.2). The solid, black lines in the *top panels* show the distribution of input SF_∞ (*left*) and τ (*right*) normalized by the median input value, and the ratio of output to input values are shown by the red, dashed lines. The *bottom panels* show the scatter in the output-to-input ratios (*left*), and the relationship between the output parameters (*right*). Listed in each bottom panel is the slope of a linear fit (dashed line) to the median values (open circles). Contours show the 90%, 70%, 50%, and 20% levels.

as a consequence of the finite width of the τ distribution.

An important piece of information missing from the MacLeod et al. (2008) analysis is the existence of a characteristic timescale τ. In MacLeod et al. (2008), the individual structure functions (SF) were computed following the standard approach with a fixed observer's frame time lag of 1 year (and with rest-frame time lags spanning 100 to 300 days). However, Eq. 4.3 indicates that the SF at small time lags should be proportional to $1/\sqrt{\tau}$, indicating that the observed structure functions will vary between quasars even if they have similar SF_∞. Indeed, the distribution of $\sqrt{\text{median}(\tau)/\tau}$ has an almost identical shape and width as that of SF/SF_{model} (see Figure 2.8). We therefore conclude that variations in τ are

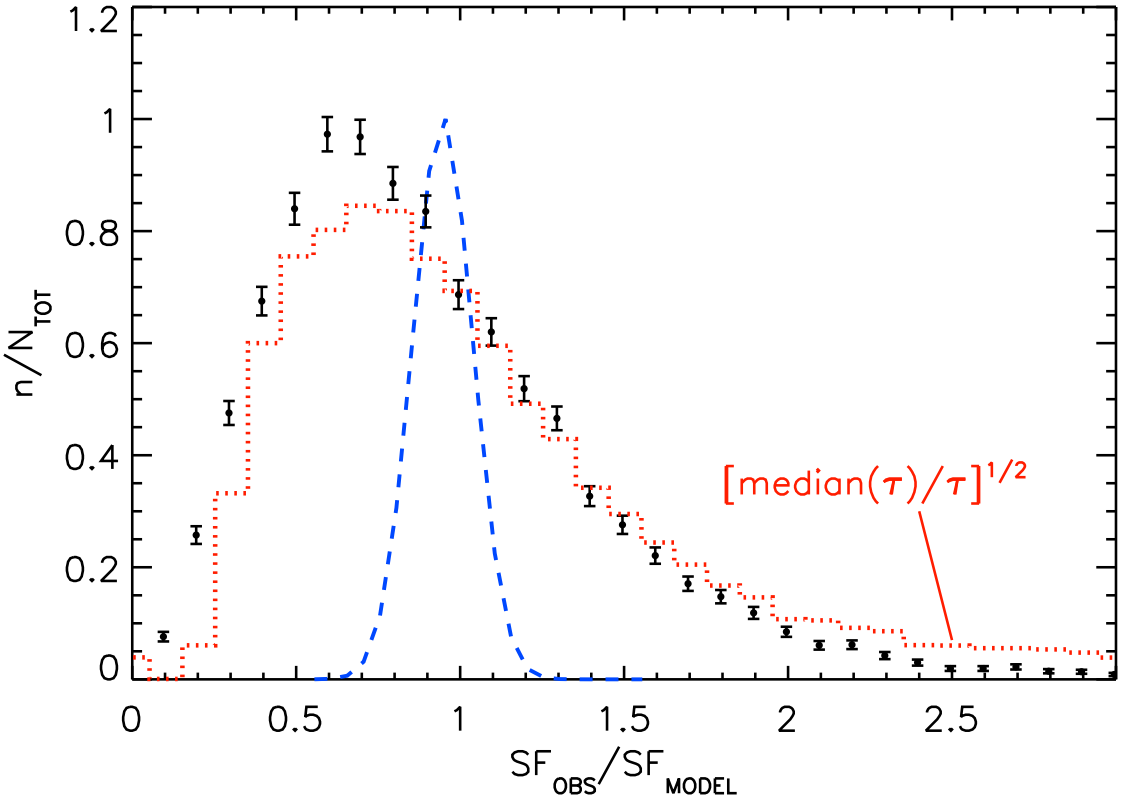

Figure 2.8 The symbols with error bars show the distribution of SF_{obs}/SF_{model} from Figure 2 in MacLeod et al. (2008), where SF_{obs} is the observed structure function for individual S82 light curves for a fixed observed-frame time lag of 1 year, and $SF_{model} = 1(1 + 0.03M_i)(\Delta t_{RF}/\lambda_{RF})^{0.47}$. The distribution from the ensemble analysis by I04 (based on two-epoch SDSS data) is shown by the dashed curve, scaled by a factor of 0.2. The dotted histogram is the distribution of $\sqrt{\text{median}(\tau)/\tau}$ from this work, where τ is the characteristic timescale for each light curve. Its agreement with the SF_{obs}/SF_{model} distribution for individual light curves shows that most of the scatter pointed out by MacLeod et al. (2008) is due to a finite width of the τ distribution.

responsible for most of the scatter in $SF(\Delta t << \tau)$ for quasars with similar luminosity, rest wavelength, and time lag. In addition, it is likely that the τ distribution is responsible for the exponential tails of the magnitude difference distribution for quasars reported by I04 and Sesar et al. (2006; see MacLeod et al. (2011a) for further discussion). Therefore, the published structure function results based on two-epoch datasets can only be interpreted in the context of Eq. 3.4.

2.5.4 Dependence on the Underlying PSD

The analysis throughout this paper assumes that the variability is described by a damped random walk, which has a power spectral distribution (PSD) described by $PSD \propto f^{-2}$ at frequencies $f > (2\pi\tau)^{-1}$, flattening to a constant at lower frequencies. In this Section, we investigate the sensitivity of our resulting parameter distributions to the possibility that the low frequency part of the PSD is not flat. For example, the X-ray variability of Seyfert galaxies is well-described by a broken power law with a slope of -2 at high frequencies, breaking to a shallower slope (-1) at low frequencies (Arévalo et al. 2008a). Since the optical and X-ray variations are correlated (Arévalo et al. 2008b), it is plausible that the optical variability might have a similar underlying PSD. On the other hand, as reviewed by McHardy (2010), the optical and X-ray fluctuations are only correlated on timescales which are shorter than the typical characteristic optical timescale (i.e., on timescales corresponding to the $1/f^2$ part of the optical power spectra). If the damped random walk is a good description for the optical variability, then we might expect that the optical and X-ray fluctuations are no longer correlated on timescales longer than τ, as the optical fluctuations resemble white noise on these time scales.

We consider three cases for a broken, or bending, PSD described by $PSD \propto f^{-2}$ at high frequencies and $PSD \propto f^{\alpha}$ at low frequencies. The first case is where $\alpha = 0$, which is a damped random walk. The second case is where $\alpha = -1$, and the third is where $\alpha = -1.9$, which is nearly a constant power-law slope. In each case, $\sim$7000 light curves are simulated using the algorithm from Timmer & Koenig (1995) with the chosen PSD. For each realization, the break frequency is set to $(2\pi\tau)^{-1}$ and the total rms is fixed, using the observed r-band τ and rms values for the S82 quasars that satisfy the selection criteria in Section 2.4.1. Therefore, the observed distributions in Figure 2.3 (for the r-band) are the "input" values. Figure 2.9 shows 3 example light curves simulated using $\alpha = -1$ (red), $\alpha = 0$ (blue), or $\alpha = -1.9$ (green). The lines in the last panel show each PSD from which the light curves are generated, with the black dotted line indicating the break frequency. As a check, the PSD was computed for each simulated light curve, and as seen by the colored dots, the shapes match the input.

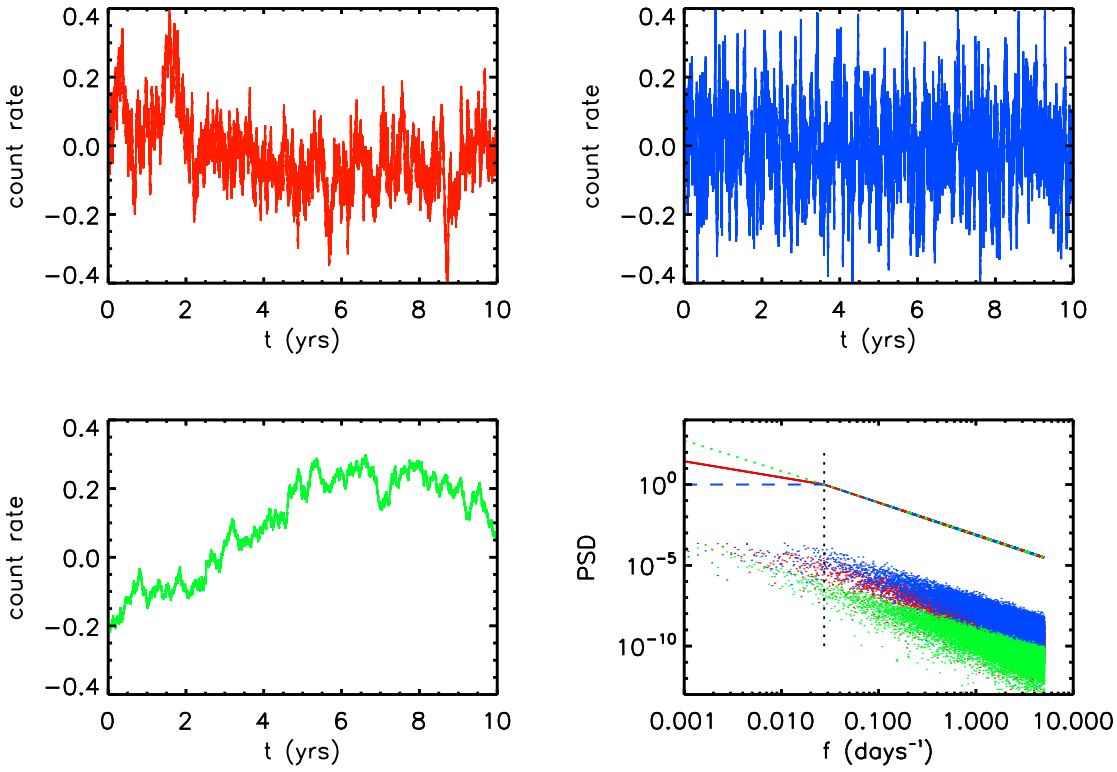

Figure 2.9 *First three panels*: Simulated light curves using $\tau = 5.8$ days, $rms = 0.124$ mag, and a PSD described by $PSD \propto f^{\alpha}$, with $\alpha = -2$ for $f > (2\pi\tau)^{-1}$, and $\alpha = -1$ (*red; top-left*), $\alpha = 0$ (*blue; top-right*), or $\alpha = -1.9$ (*green; bottom-left*) for $f < (2\pi\tau)^{-1}$. The lines in the *last panel* show each PSD from which the light curves are generated, with the *black dotted line* indicating the break frequency. As a check, the PSD was then computed for the simulated light curves, shown with *colored dots*. The y-axis in the bottom-right panel is in arbitrary units.

Each light curve was simulated over 100 years and then truncated to the 30–40 year segment in order to account for the additional variability and bias in the inferred SF_{∞}, which may result from a red noise leak or an incorrect specification of the PSD model (see Uttley et al. 2002). The simulated light curves were then modeled as a damped random walk to obtain the ("output") parameter distributions shown in Figures 2.10 and 2.11. For Figure 2.10, the simulated observations are spaced every 5 days over 10 years with typical errors of 0.01 mag, and for Figure 2.11, the S82 window function is imposed (i.e., all light curves have the S82 time sampling and photometric accuracy). The filled histograms show

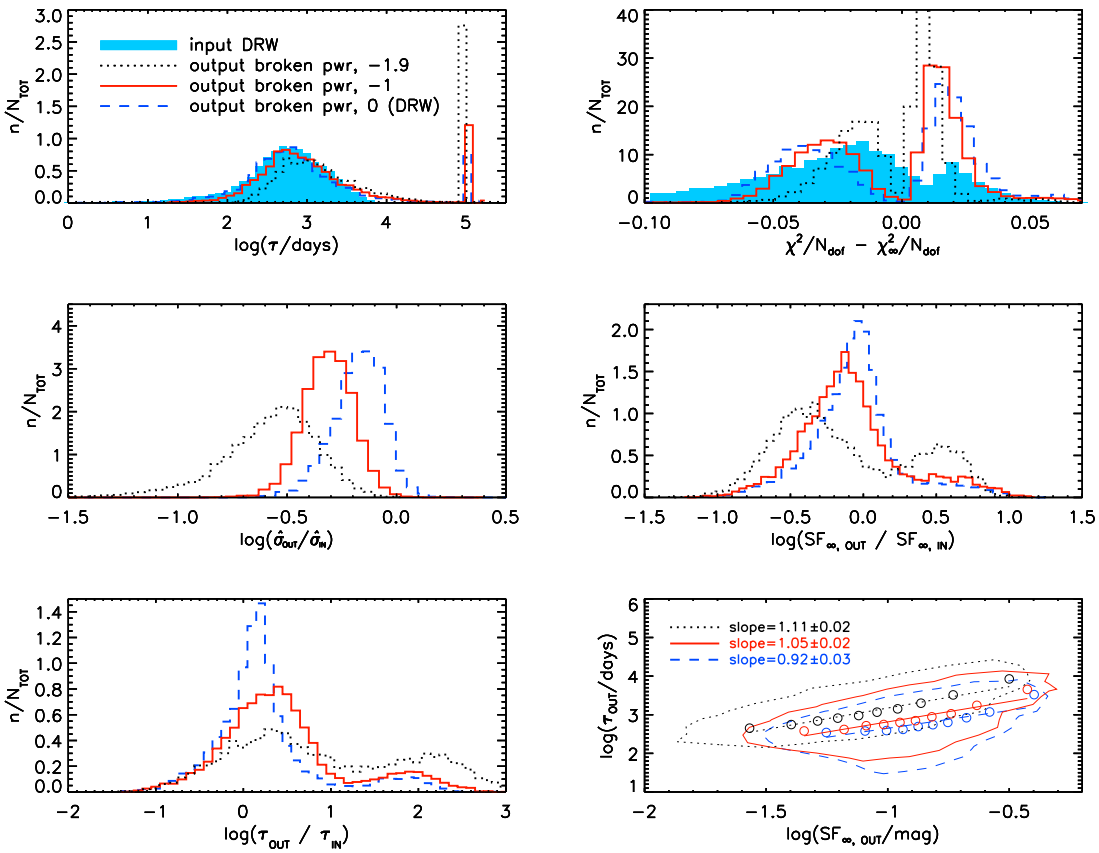

Figure 2.10 In the top panels, the *filled histogram* shows the ("input") distributions of τ and $\Delta\chi^2_\infty = \chi^2/N_{dof} - \chi^2_\infty/N_{dof}$ for r-band S82 light curves (∼7000 total). This distribution of τ, along with the r-band rms values, are used to generate ∼7000 realizations of noise processes with $\alpha = -2$ for $f > (2\pi\tau)^{-1}$ and $\alpha = 0$, $\alpha = -1$, or $\alpha = -1.9$ for $f < (2\pi\tau)^{-1}$. The *dashed blue*, *solid red*, and *dotted* lines show the best-fit ("output") distributions when modeling these three processes, respectively, as a damped random walk (DRW). Here, the simulated observations are spaced every 5 days over 10 years with typical errors of 0.01 mag. The output $\hat{\sigma}$, SF_∞, and τ are compared to the input values in the next three panels. In the last panel, the correlation between the output $log(SF_\infty)$ and $log(\tau)$ is shown using a linear fit to the median values (*open circles*), with the slopes listed in the legend for each case. Contours show regions containing 90% of the data points.

the input distributions and are the same for both Figures.

It is clear from comparing the input and output parameter distributions that τ and SF_∞ map simply onto the values of the power-spectral break timescale and amplitude. Indeed, the correlation between τ and SF_∞ is preserved in the output distributions, as seen in the bottom-right panels of Figures 2.10 and 2.11. However, as seen in the top- and bottom-left panels, as the PSD slope at low f steepens to $\alpha = -1.9$, the number of "run-away" time

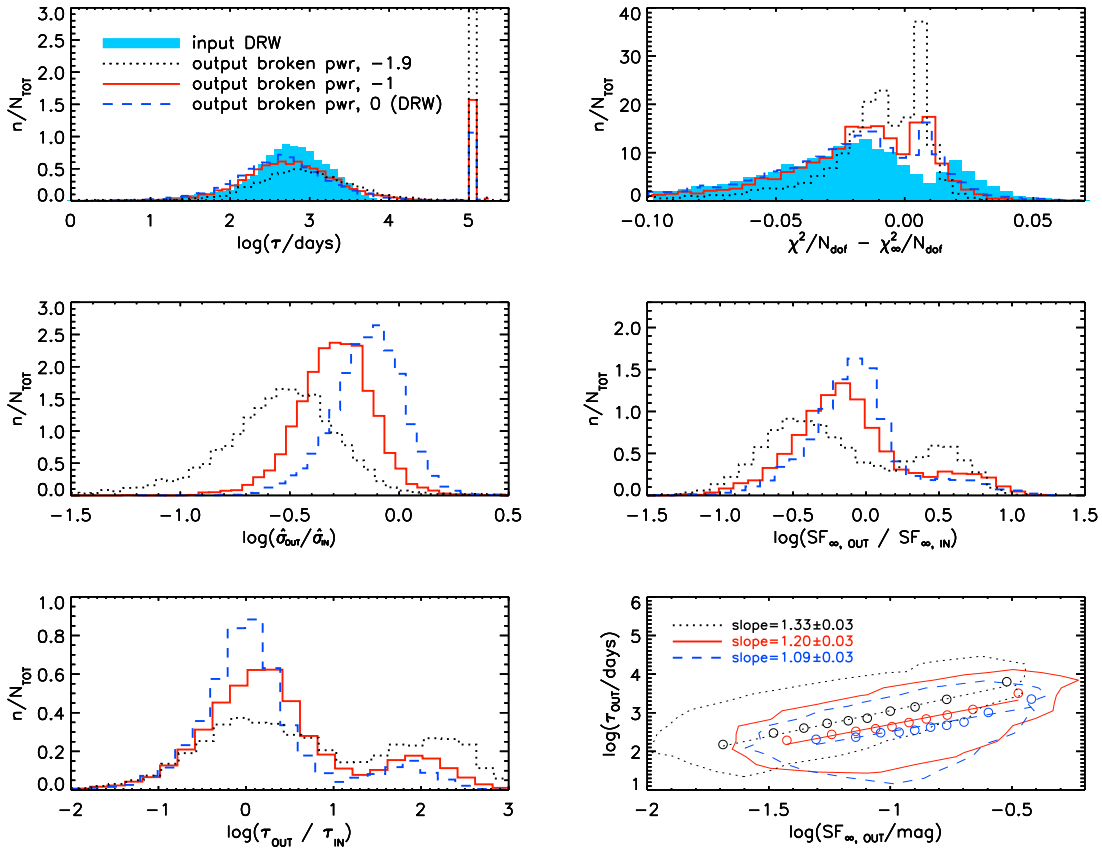

Figure 2.11 As in Figure 2.10, except with the Stripe 82 time sampling and photometric accuracy imposed on the simulated light curves.

scales (where τ saturates at 10^5 days) increase due to that fact that τ can no longer be constrained. Whereas for the S82 data, $\sim$20% of light curves were rejected as run-away cases (see step 3 of Section 2.4.1), for an $\alpha = -1.9$ PSD, the run-away fraction is 30%. This significant increase rules out the $\alpha = -1.9$ PSD as the correct model, as we do not see this large run-away fraction in the data. The fractions are similar for the $\alpha = -1$ and $\alpha = 0$ cases, suggesting that the correct model has $-1 < \alpha < 0$. The increasing fraction of run-away τ with decreasing α allows one to distinguish between each PSD in the well-sampled case, as seen in the top-right panel of Figure 2.10. In this panel, the difference between the χ^2_{pdf} for the best-fit damped random walk and that for a $\tau \to \infty$ solution, $\Delta\chi^2_\infty = \chi^2/N_{dof} - \chi^2_\infty/N_{dof}$, is shown for each input PSD. The total distribution shows

two peaks; that on the left corresponds to cases where the model is able to constrain τ. For well-sampled light curves (Figure 2.10), the case with $\alpha = 0$ (dashed line) yields an overall lower $\Delta\chi^2_\infty$ and therefore a more likely damped random walk fit, as expected. However, for the S82 sampling, it is difficult to distinguish between the red and blue lines ($\alpha = 0$ and -1), and the (input) observed distribution is similar to both. Therefore, within the limited S82 sampling, we are unable to distinguish reliably between a damped random walk and a $1/f$ PSD on long timescales.

2.6 Dependence of Variability Parameters on Luminosity, Wavelength, Redshift, and Black Hole Mass

Next, we discuss correlations between each of the variability parameters, τ and SF_∞, and the four available physical parameters: rest-frame wavelength (λ_{RF}), redshift (z), absolute i-band magnitude (M_i), and black hole mass (M_{BH}). It is important to fit a multiple regression to each of the physical parameters, functions of the form $\tau(\lambda_{RF}, M_i, M_{BH}, z)$, because of the correlations between physical parameters. For example, when searching for a correlation between SF_∞ and M_{BH}, one must take into account that a more luminous quasar hosts a more massive black hole (e.g., Kollmeier et al. 2006), or else a trend with luminosity may be mistaken for a trend with black hole mass. A similar example is the well-known luminosity-redshift (L–z) degeneracy seen in flux-limited samples of quasars. Magnitude limits result in the illusion that only the most luminous quasars are seen at high z, and therefore the observed L increases with z independent of reality. Having a large sample size helps to alleviate these degeneracies. For example, one can look for trends using 2-dimensional grids in any two physical parameters of interest. The numbers of data points per two-dimensional bin of M_i and redshift, and similarly for M_i and M_{BH}, are shown in the bottom panels of Figure 2.12. The two left panels of Figure 2.12 show the selection effect that quasars at higher redshift must have higher luminosities to be included in the survey. From the two-dimensional distribution in the top-left panel, we can see that this L–z degeneracy is essentially independent of M_{BH}. Furthermore, shorter rest-frame wavelengths are probed at higher redshifts for two reasons: first, quasars emitting at shorter rest wavelengths must be at higher redshifts in order to be observed within the *ugriz* filters, and second, quasars

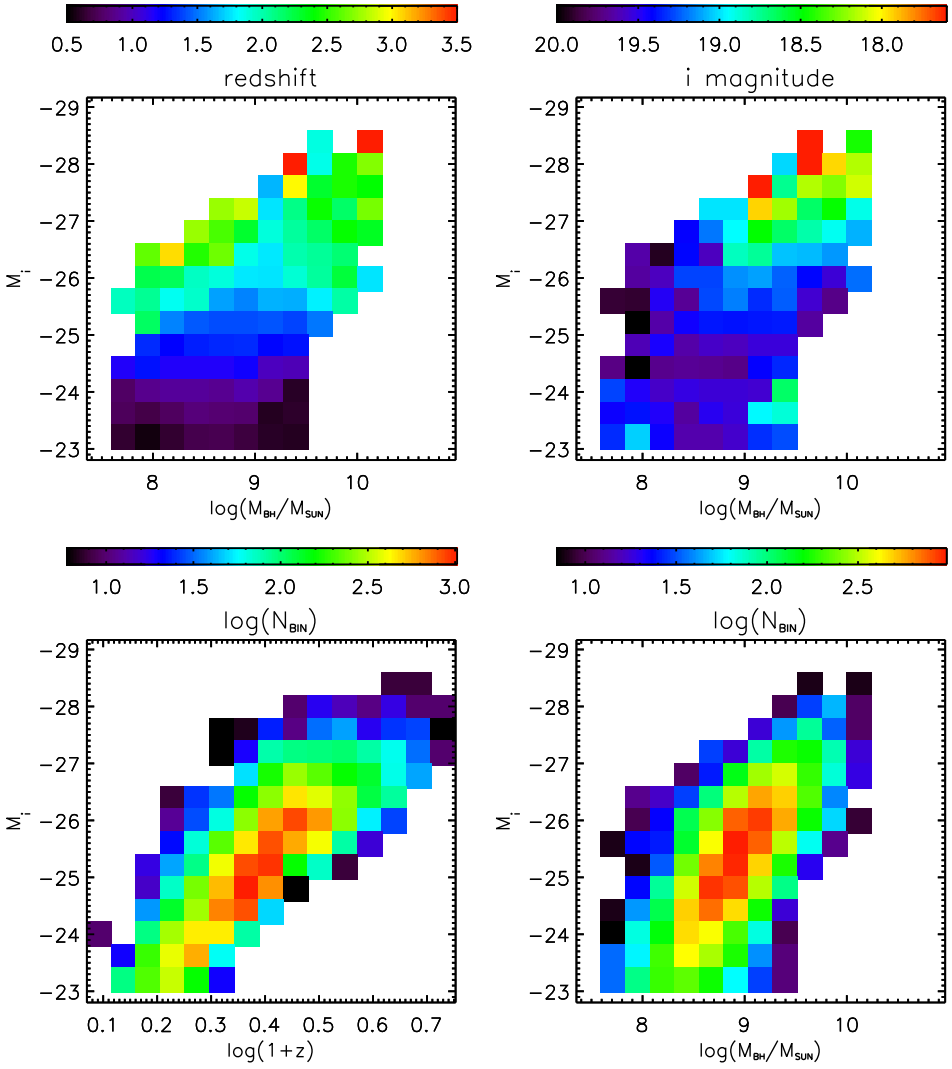

Figure 2.12 The *top panels* show the distribution of quasars in redshift (*left*) and i magnitude (*right*) on a grid of M_{BH} vs. M_i. The *bottom panels* show the number of data points in bins of M_{BH} (or redshift) vs. M_i.

at high redshifts are closer to their Eddington limit as a result of the L–z degeneracy, and possibly cosmological downsizing (e.g., Kollmeier et al. 2006). Therefore, any dependence of variability on wavelength must be accounted for when considering a dependence with redshift (or luminosity).

When fitting power laws to a large number of data points throughout this paper, we fit

to the median values in each bin of the independent variable, where all bins have the same number of data points N. The errors in the medians are computed as $0.93(\mathrm{IQR})/\sqrt{N-1}$, where IQR is the 25% – 75% interquartile range, and these constrain our formal errors in the power law slopes.

2.6.1 Trends with Rest-frame Wavelength

We start by examining the wavelength dependence of the variability parameters. Since there are multiple bands for each quasar, this dependence can be determined for individual sources for which z, M_i, and M_{BH} are fixed. The rest-frame wavelength is found by dividing the observed bandpasses (3520, 4800, 6250, 7690, and 9110 Å for *ugriz*, respectively) by $(1+z)$. We fit a power law $f \propto (\lambda_{RF}/4000\mathring{A})^B$ to the estimates of τ and SF_∞ for every quasar observed in at least two filters (∼8,000 quasars). The median values are $B = 0.17 \pm 0.02$ and -0.479 ± 0.005 for τ and SF_∞, respectively. We searched for significant correlations between B and other physical parameters and did not find any. We use these median values to correct for the wavelength dependence of τ and SF_∞ before searching for their correlations with other physical parameters in Sections 2.6.2, 2.6.3, and 2.7.

This method of fixing B before investigating other correlations naturally eliminates any degeneracies between rest wavelength and the other physical parameters. This is especially important in the case of the variability amplitude, SF_∞. Figure 2.13 shows how the two variability parameters vary with λ_{RF} in each of the *ugriz* filters. The median power law from the above analysis, shown as a straight line stretching across all wavelengths, accurately traces the overall trend. However, when the data are fit to ensembles of quasars in each band separately, the slopes for SF_∞, shown by the short lines for each filter, are very different (∼0.4). This difference is a consequence of the correlation between L and z: for a given band, shorter λ_{RF} corresponds to higher z, but at higher z, quasars have higher L and thus smaller SF_∞ (see below), creating a bias in the inferred slope of the wavelength dependence.

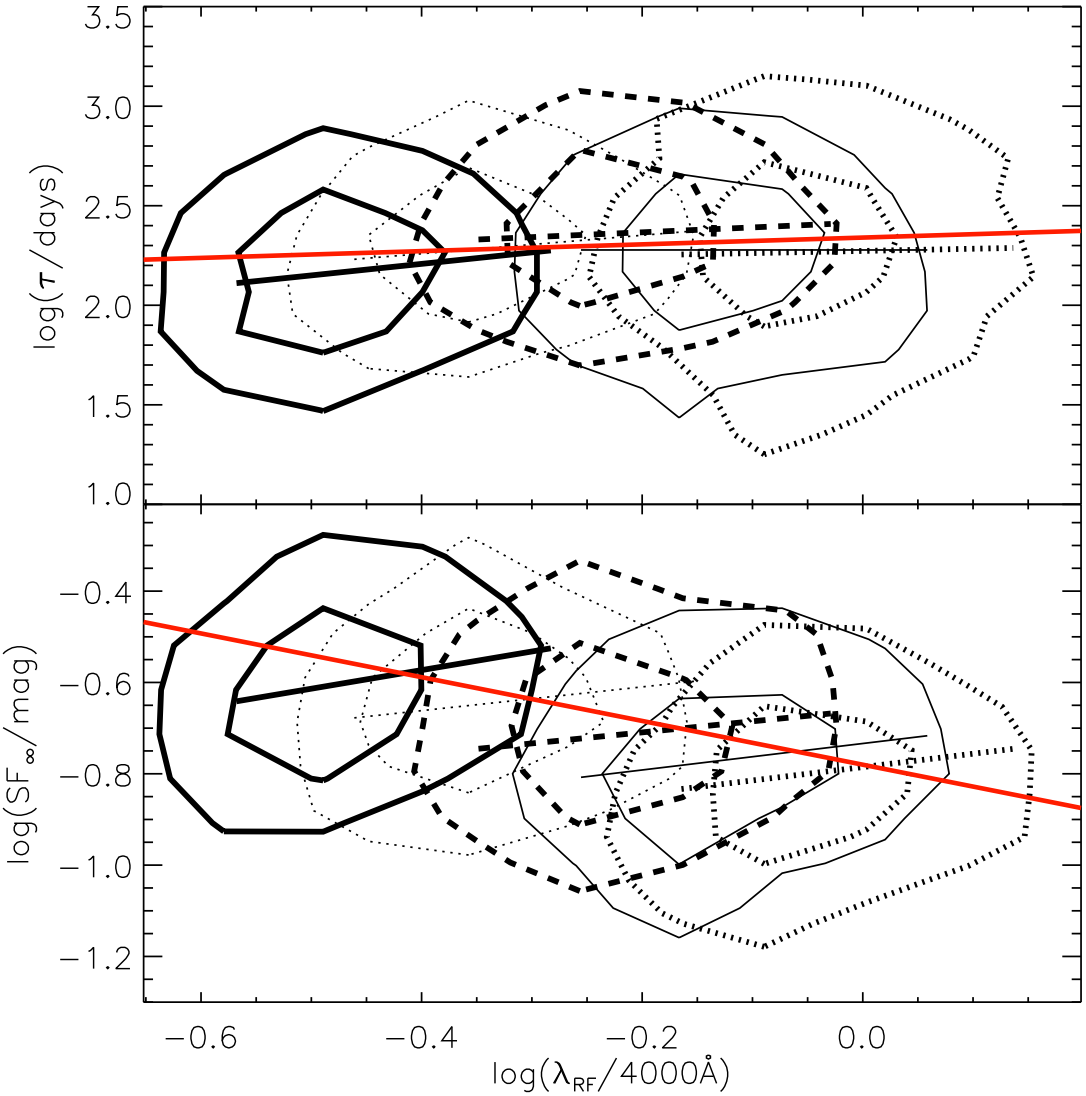

Figure 2.13 Distribution of the rest-frame timescale τ (*top panel*) and long-term structure function SF_∞ (*bottom panel*) as a function of rest-frame wavelength λ_{RF}. The different contours show the 70% and 30% levels for the u, g, r, i, and z (from left to right) bands. The best-fit power law for each subsample is shown with the same line style as the contours. The thick red line connecting the left and right axes shows the median power law slope derived by fitting individual quasars (see text), and has a value of $B = 0.17$ and -0.479 for τ and SF_∞, respectively. The slope within each band from fitting ensembles of quasars can be quite different from the overall slope because of the L–z degeneracy (see text).

2.6.2 Trends with Luminosity, Redshift, and Black Hole Mass

In the upper panels of Figure 2.14, the median values of SF_∞ and τ are shown as a function of absolute magnitude (M_i) and redshift. The structure function parameters are normalized to a fixed rest wavelength of 4000Å using the fitted power law dependencies of $(\lambda_{RF}/4000\mathring{A})^B$ with $B = -0.479$ and 0.17 for SF_∞ and τ, respectively, from Section 2.6.1, before finding the median in each pixel. For SF_∞, the anti-correlation with luminosity clearly dominates any trend with redshift. The τ distribution also shows negligible correlation with redshift. The bottom panels show the dependence on M_{BH}. Using a grid of M_i vs. M_{BH} allows one to search for trends in the variability parameters while accounting for the selection effect that more luminous galaxies host more massive black holes. A positive correlation between SF_∞ and M_{BH} is apparent, independent of the correlation with M_i. This is in agreement with the result from Wilhite et al. (2008), who used ensemble structure functions. The τ parameter shows a clear correlation with M_{BH}, which dominates any trend with M_i.

Motivated by these qualitative results, we fit a power law of the form

$$\log f = A + B \log\left(\frac{\lambda_{RF}}{4000\text{Å}}\right) + C(M_i + 23) + D \log\left(\frac{M_{BH}}{10^9 M_\odot}\right) + E \log(1+z), \qquad (2.7)$$

to all SF_∞ and τ data points in each *ugriz* band separately, keeping B fixed to -0.479 and 0.17, respectively, in order to avoid the bias discussed in Section 2.6.1. While there is a lot of scatter in the variability parameters for fixed M_i, M_{BH}, and λ_{RF} (see Section 2.5.2), Eq. 2.7 describes the mean trends well. The best-fit coefficients, averaged over the five bands, are reported in the first and sixth rows of Table 2.1. The best-fit coefficients when simultaneously fitting all bands are consistent with these averages. The reported error bars are computed from the variation of the best-fit parameters over the five bands.

The dependence of τ on redshift and M_i is only marginally detected, and can be attributed to the L–z degeneracy: as redshift increases, the best-fit τ decreases, while the coefficient for M_i indicates that as luminosity increases, τ increases. When E is fixed to zero, the dependence on M_i has low significance, while the correlation with M_{BH} remains the same (seventh row of Table 2.1). For illustration, when E is fixed to 1 [so $\tau \propto (1+z)$, third and eighth rows], a spurious dependence on M_i emerges. Therefore, the controlling variable for the characteristic timescale is clearly M_{BH}, suggesting that more massive black

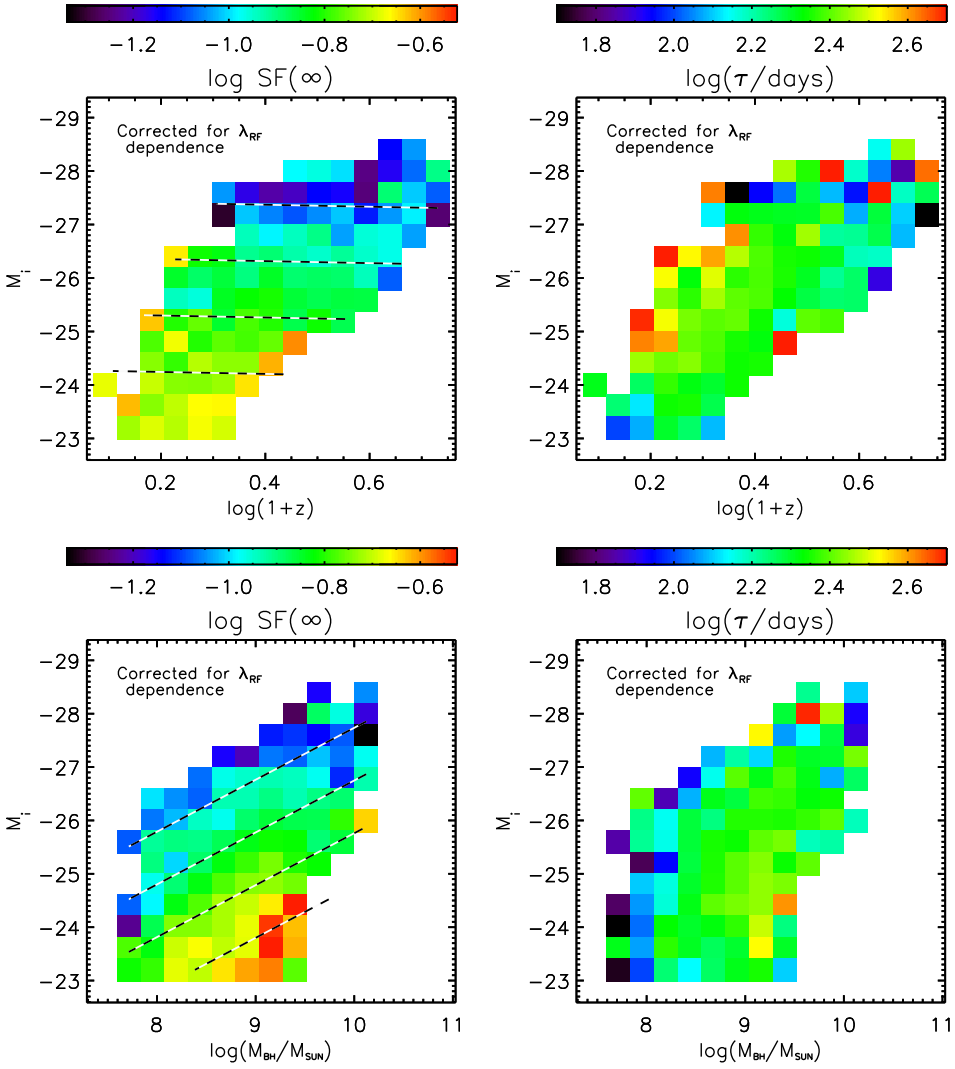

Figure 2.14 *Top panels:* The long-term rms variability SF_∞ (*left*) and characteristic timescale τ (*right*) are shown as colors on a grid of redshift and absolute i-band magnitude M_i. The structure function parameters are normalized to a fixed rest wavelength using the fitted power law dependencies of $(\lambda_{RF}/4000\mathring{A})^B$ with $B = -0.479$ and 0.17 for SF_∞ and τ, respectively. The lines of constant variability (dashed) show that SF_∞ is independent of redshift. *Bottom panels:* As in the top panels but with black hole mass M_{BH} on the x-axis.

holes vary on longer timescales. Similarly, with E fixed to 0 for SF_∞ (second row), the C (M_i) and D (M_{BH}) coefficients remain largely unchanged, confirming that there is no significant correlation between amplitude and redshift. Therefore, we force the final adopted

model to have no redshift dependence ($E = 0$). Table 2.1 also provides fits for when D is fixed to zero so that SF_∞ and τ can be estimated in the absence of black hole mass information (fourth and ninth rows).

Of the analyzed parameters, the probable errors for black hole mass estimates are the largest: of order 0.2 – 0.4 dex (Marconi et al. 2008; Vestergaard & Peterson 2006). When ignored, these large statistical uncertainties result in underestimated values for D (Kelly 2007). With an assumed random uncertainty of 0.2 dex in the black hole mass measurements, we find the best-fit coefficients reported in the fifth and tenth rows of Table 2.1 (continuing to assume that $E = 0$). It can be seen that including mass uncertainties of 0.2 dex increases D from $\sim$0.1 to $\sim$0.2 for both τ and SF_∞ (a bias of about 2-3 formal statistical fitting errors). *The coefficients in the fifth and tenth rows represent our best-fit model for the variability parameters, and we assume these values in the remaining Sections.*

The best-fit model shows a smaller slope for the correlation between τ and M_{BH} of 0.21 ± 0.07 than the value of 1.0 ± 0.4 found by KBS09. The two results are still marginally statistically consistent given the large uncertainties in KBS09. Moreover, their sample contains relatively lower mass and luminosity quasars than those analyzed in this study, and this difference might result in additional biases. Kozł10 also noted that the KBS09 results may be affected by contamination from host galaxy emission, a problem which will be far smaller for the generally more luminous SDSS quasars.

2.6.3 The Eddington Ratio as the Driver of Variability?

Since the Eddington ratio, L/L_{Edd}, is dependent on luminosity and black hole mass, the trends for SF_∞ in Table 2.1 might be explained if SF_∞ is simply driven by L/L_{Edd}. To test this, we estimated L/L_{Edd} as the ratio of the bolometric luminosities from Shen et al. (2008) to the Eddington luminosity, $L_{Edd} = 1.5 \times 10^{38}(M_{BH}/M_\odot)$ erg/s. Figure 2.15 shows L/L_{Edd} as a function of M_{BH} and M_i, and demonstrates that lines of constant L/L_{Edd} are similar in slope to those of constant SF_∞ (see Figure 2.14). The median SF_∞ versus the median L/L_{Edd} for each bin is shown in the right panel, where we find a power law slope of -0.23 ± 0.03. This significant anti-correlation was also found by Wilhite et al. (2008), Bauer

Table 2.1. Best-Fit Coefficients for Eq. 2.7

f	A	B (λ_{RF})	C (M_i)	D (M_{BH})	E ($1+z$)
SF_∞	-0.57 ± 0.01	-0.479 ± 0.005	0.117 ± 0.009	0.11 ± 0.02	0.07 ± 0.05
SF_∞	-0.56 ± 0.01	-0.479 ± 0.005	0.111 ± 0.005	0.11 ± 0.02	$\equiv 0$
SF_∞	-0.760 ± 0.009	-0.479 ± 0.005	0.193 ± 0.006	0.12 ± 0.02	$\equiv 1$
SF_∞	-0.618 ± 0.007	-0.479 ± 0.005	0.090 ± 0.003[a]	$\equiv 0$	$\equiv 0$
SF_∞[c]	-0.51 ± 0.02	-0.479 ± 0.005	0.131 ± 0.008[b]	0.18 ± 0.03	$\equiv 0$
τ	2.4 ± 0.2	0.17 ± 0.02	-0.05 ± 0.03	0.12 ± 0.04	-0.7 ± 0.5
τ	2.3 ± 0.1	0.17 ± 0.02	0.01 ± 0.03	0.12 ± 0.04	$\equiv 0$
τ	2.1 ± 0.1	0.17 ± 0.02	0.09 ± 0.03	0.13 ± 0.04	$\equiv 1$
τ	2.2 ± 0.1	0.17 ± 0.02	-0.01 ± 0.02	$\equiv 0$	$\equiv 0$
τ[c]	2.4 ± 0.2	0.17 ± 0.02	0.03 ± 0.04	0.21 ± 0.07	$\equiv 0$

[a]Based on K-corrected magnitudes. Without the K-correction, this coefficent changes to 0.079 ± 0.003.

[b]Based on K-corrected magnitudes. Without the K-correction, this coefficent changes to 0.113 ± 0.006.

[c]Measurement errors in M_{BH} of 0.2 dex have been included in the fitting. These coefficients are recommended when applying the model.

Note. — In each row, the B coefficient was determined and fixed before fitting a multiple regression in all other parameters (see Section 2.6.1). The cosmology used for determining M_i is $\Omega_M = 0.30$, $\Omega_\Lambda = 0.70$, and $h = 0.70$ (Schneider et al. 2007), whereas that used in the M_{BH} estimates is $\Omega_M = 0.26$, $\Omega_\Lambda = 0.74$, and $h = 0.71$ (Shen et al. 2008). This difference should only have a 1% effect on the best-fit coefficients.

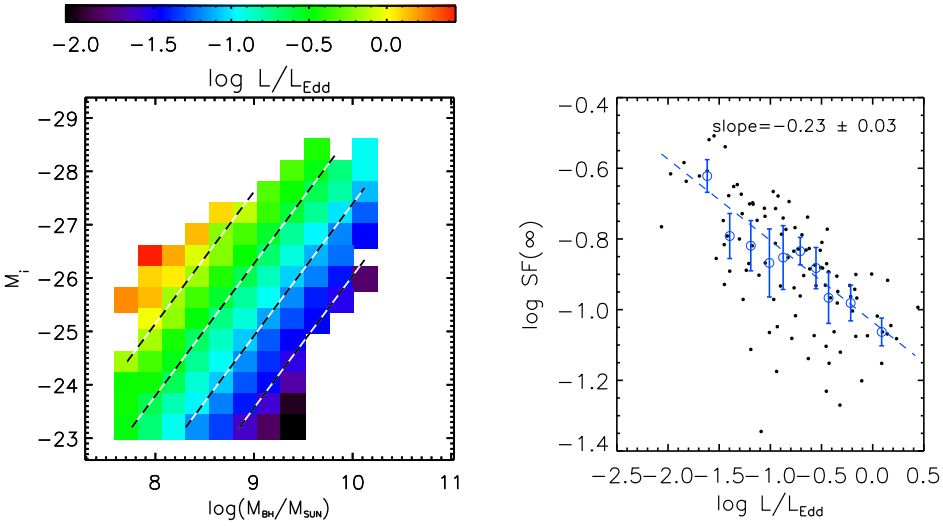

Figure 2.15 *Left:* The Eddington ratio for S82 quasars (estimated using masses and bolometric luminosities from Shen et al. 2008) is shown as colors on a grid of M_i vs. M_{BH}, with dashed lines of constant L/L_{Edd} over-plotted. *Right:* Long-term rms variability (corrected for wavelength dependence) is shown as a function of L/L_{Edd} (open circles are medians in each bin). The slope of the linear fit to the medians is listed on the panel.

et al. (2009), and Ai et al. (2010). KBS09 did not report such an anti-correlation; however, they did not compare L/L_{Edd} with SF_∞ but rather with $SF_\infty/\sqrt{\tau}$, the driving amplitude of short term variability. If L/L_{Edd} is the sole driver of the quasar variability amplitude SF_∞, we would expect that the coefficients for M_i and M_{BH} are related as $D = 2.5C$. However, we find $D = (1.37 \pm 0.23)C$, 4.7σ away from the presumed slope of 2.5. This suggests an additional source of variability, such as a dependence on luminosity or on M_{BH} in addition to the dependence on L/L_{Edd}. Moreover, if we ignore this additional source, the result that SF_∞ depends on L/L_{Edd} supports a scenario where the amplitude of the optical variability is determined by the accretion rate (see discussion in Wilhite et al. 2008).

If the Eddington ratio is a proxy for the quasar's age (e.g., Martini & Schneider 2003), then a lower Eddington ratio, and thus a larger amplitude of variability, could indicate a dwindling fuel supply and a more variable rate at which it is supplied to the black hole. However, it is unlikely that the observed variability is due to fluctuations in the external fuel supply to the disk, because the fluctuation timescales of 10 – 10,000 days are very short

compared to the viscous timescales of 10^5 – 10^7 days that should control large scale changes in the accretion rate. Moreover, these fluctuations in the fueling rate will also be smoothed out and damped as they travel inward, and will likely have effectively disappeared by the time they reach the optical emitting region (e.g., see discussion in Churazov et al. 2001). Instead, the origin of the fluctuations is probably more local.

Another possibility is that the dependence on L/L_{Edd} may simply be a reflection of the dependence on wavelength, which in turn depends on the disk radius. If a higher L/L_{Edd} means a hotter disk, then the optical flux originates at a larger radius. Assuming that longer wavelengths are emitted further out in the disk, the lower SF_∞ at longer wavelengths would lead to the anti-correlation between SF_∞ and L/L_{Edd}. In thin disk theory (e.g., see Frank, King, & Raine 2002), the characteristic radius for emission at wavelength λ_{RF} scales as $R_\lambda \propto M_{BH}^{2/3}(L/L_{Edd})^{1/3}\lambda_{RF}^{4/3}$, and the thermal timescale is related to the orbital timescale as $t_{th} \propto t_{orb} \propto R^{3/2}/\sqrt{M_{BH}}$. Therefore, under the assumption that τ is related to the thermal time scale, and that variability at wavelength λ_{RF} is dominated by the scale R_λ, τ scales with λ_{RF}, M_{BH}, and L as:

$$\tau \propto M_{BH}^{1/2}(L/L_{Edd})^{1/2}\lambda_{RF}^{2}. \tag{2.8}$$

Since $L_{Edd} \propto M_{BH}$, this means τ scales as

$$\tau \propto L^{1/2}\lambda_{RF}^{2}, \tag{2.9}$$

which does not match the observed scaling of $\tau \propto L^{-0.075}\lambda_{RF}^{0.17}M_{BH}^{0.21}$. If we assume that τ is related to the viscous timescale,

$$\tau \propto L^{7/60}\lambda_{RF}^{5/3}M_{BH}^{2/3}, \tag{2.10}$$

which is still in conflict with the measured values.

The variability timescale simply does not show the strong dependence on λ_{RF}, M_{BH}, and L expected from these simple scalings. Therefore, using this naive scaling, we are not able to relate the observed τ to either a thermal or viscous timescale of the radius associated with the wavelength of the variability. However, in reality there is a range of radii, which corresponds to a range in timescales, contributing to the observed flux in each band, and this will cause some degree of smoothing. Also, the radial regions might overlap for each

band, causing a single radius to contribute flux in multiple bandpasses, and this might attenuate some of the dependencies on λ_{RF}, M_{BH}, and L. Finally, our observed power law indices may be biased due to the uncertainty in the bolometric correction.

2.6.4 A "Recipe" for Generating Mock Light Curves

Given the success of the damped random walk model in explaining a large body of observations, it represents a simple quantitative framework for generating mock quasar light curves. The optical light curves can be simulated for quasars at different redshifts and for a wide range of luminosity and black hole mass, which provides the basis for quantitative modeling and optimization of quasar variability surveys. The model presented has already been implemented as a part of the simulation effort in support of the Large Synoptic Survey Telescope (Ivezić et al. 2008).

At a given redshift, a simulated value for absolute magnitude can be readily drawn from an adopted luminosity function (e.g., Croom et al. 2009 and references therein). An estimate for the black hole mass is also required in order to apply our model, and it can be generated using the adopted absolute magnitude as follows. According to the Shen et al. (2008) results, the quasar luminosity and black hole mass are strongly correlated (see the bottom-right panel of Figure 2.12). Using Shen et al. (2008) values, we quantify this correlation in Figure 2.16. At a given M_i, the M_{BH} distribution has a finite width due to the distribution of Eddington ratios and measurement errors. Both effects can be well-described using a Gaussian distribution:

$$p(\log M_{BH}|M_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\log M_{BH} - \mu)^2}{2\sigma^2}\right], \tag{2.11}$$

where $\mu = 2.0 - 0.27M_i$ and $\sigma = 0.58 + 0.011M_i$ (the black hole mass is expressed in solar mass units).

With adopted values for M_i and M_{BH}, the variability amplitude and the characteristic timescale can be estimated using Eq. 2.7. We note that choosing a timescale from an adopted value for SF_∞ by utilizing the correlation seen in Figure 2.3 (and drawing from a Gaussian distribution similar to Eq. 2.11) is not as accurate as using Eq. 2.7 and the adopted values

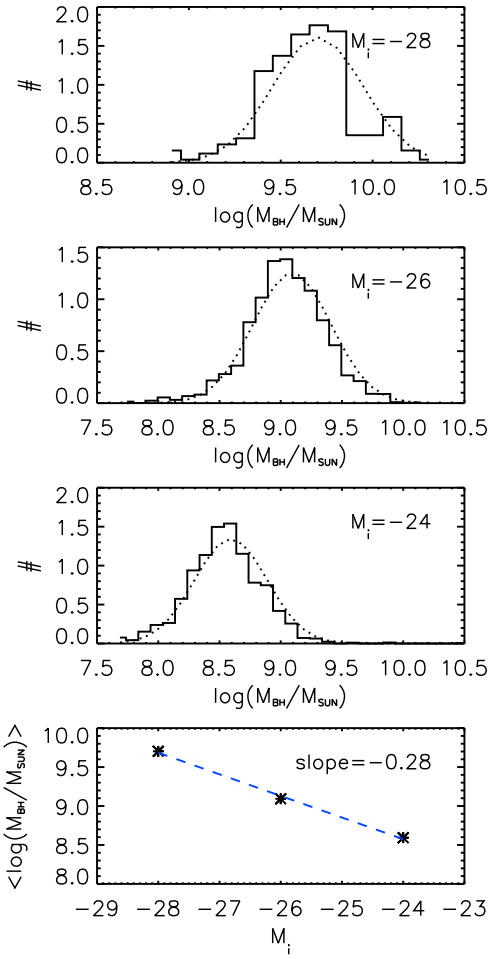

Figure 2.16 Determination of $p(M_{BH}|M_i)$ (see Eq. 2.11). The histograms in the top three panels show the distribution of black hole masses for S82 quasars with absolute magnitudes within 0.5 mags of the value listed in the upper right corners. In the bottom panel, the mean value of $\log(M_{BH}/M_\odot)$ from each histogram is plotted against M_i.

of M_{BH} and λ_{RF}. Finally, given a mean magnitude and wavelength, a quasar light curve can be easily generated using Eq. 2.5.

2.7 Variability Properties of Radio- and X-ray–detected Quasars

Previous studies, based mostly on two-epoch datasets, found that radio-loud quasars are marginally more variable than radio-quiet quasars for rest-frame time lags in the range 50–400 days (e.g., VB04 and references therein). It was also found that X-ray–detected quasars

are significantly more variable than X-ray–undetected quasars at rest-frame time lags up to 250 days.

The large size and the availability of light curves for our sample allows us to revisit these results with more statistical power. In addition, the damped random walk model and our best-fit results from Table 2.1 provide a convenient method to account for various selection effects. For example, it is possible that subsamples selected by various means, such as the requirement of radio or X-ray detections, have different distributions of luminosity and black hole mass. In this case, they would display different variability behavior not because of intrinsic differences in the variability mechanism, but rather because of the trends captured by Eq. 2.7. A comparison of the *ratio* of observed and predicted variability parameters for any subsample and the full sample will automatically take into account these sampling effects, and this is the main statistical method we use in this section. We first discuss a radio subsample, and then analyze a sample of quasars with X-ray detections.

2.7.1 Radio-detected Subsample

We use the unified radio catalog of Kimball & Ivezić (2008) to access the FIRST and NVSS data (20 cm continuum data) for our objects. We refer the reader to this paper and references therein for details about these radio surveys and object association. Figure 2.17 shows the radio-detected fraction of quasars in S82 as a function of magnitude. The overall fraction (5%) is considerably lower than that in the Northern galactic cap footprint of the SDSS for the magnitude range $19 < i < 20$ (White et al. 1997; Ivezić et al. 2002). This is due to a difference in targeting algorithms between the two surveys, and results in nearly 300 quasars with both optical light curves and radio data. When available, we use NVSS values for radio flux, but otherwise we adopt the FIRST integrated flux. The radio-loudness parameter R_g is calculated exactly as the R_i in Ivezić et al. (2002), but using g magnitudes instead. We assume a spectral index of -0.5 for the optical and 0 for the radio (because this sample is dominated by the core radio emission; see Kimball & Ivezić 2008) when computing the K-correction.

Table 2.7.2 compares the mean properties for various subsamples detected in the radio

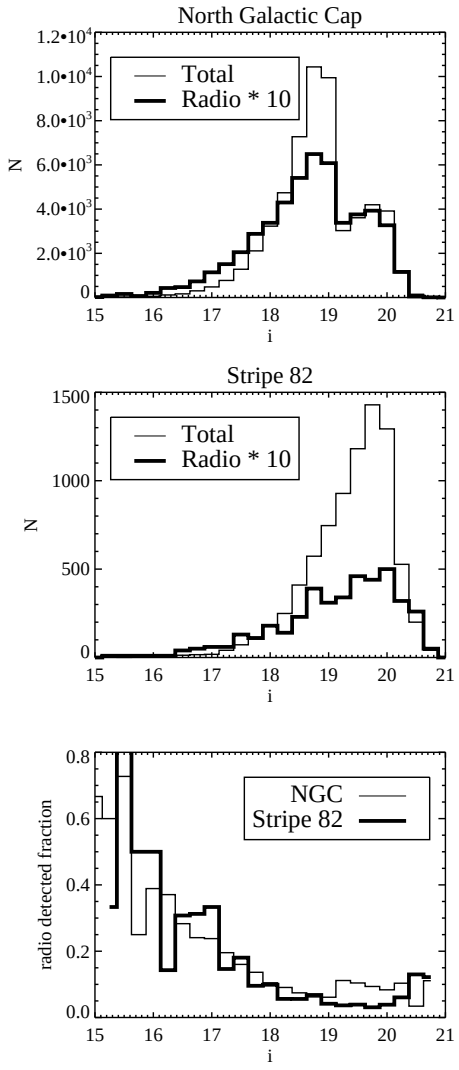

Figure 2.17 *Top:* The number of FIRST radio detections (thick line, multiplied by 10) as compared to the total number of quasars (thin line) as a function of i magnitude in the Northern galactic cap (NGC) footprint of the SDSS. *Middle:* As in the top panel but for Stripe 82. *Bottom:* Fraction of quasars in the NGC (thin line) and in S82 (thick line) samples that have radio detections. Between $i = 19$ and 20, the fraction of radio-detected quasars is considerably lower in S82 than in the NGC.

to the radio-undetected subsample (essentially the full sample due to small fraction of radio-detected objects). For each subsample, the mean τ and SF_∞ are listed, as well as the mean ratio of the observed values to those predicted using Eq. 2.7 and the measured values of λ_{RF}, M_i, and M_{BH}. The mean values are computed iteratively with $\pm 3\sigma$ outliers excluded.

Errors are reported as the (clipped) rms divided by $\sqrt{N-1}$, where N is the number of data points in the distribution. Numbers of objects are lower for columns involving model quantities because black hole mass estimates are not available for all objects.

Table 2.7.2 shows that only the variability amplitude of the radio-loudest quasar subsample ($R_g \geq 3$) is significantly different ($> 3\sigma$ deviation) from the behavior of the full sample. The radio-loudest quasars have systematically larger variability amplitudes, when corrected for trends described by Eq. 2.7, by about 30%, compared to the full sample dominated by radio-quiet objects, in agreement with VB04.

2.7.2 X-ray Detected Subsample

For the analysis of variability properties of quasars detected at X-ray wavelengths, we use data from the ROSAT All-Sky Survey (RASS; Voges et al. 1999). The X-ray subsample consists of 82 quasars with RASS full-band count rates greater than 10^{-3} ct/s, taken from the Schneider et al. (2007) catalog. As can be seen from Table 2.7.2, the variability properties of this subsample are statistically indistinguishable from the full sample. VB04 detected a significant increase in structure function at rest time lags below 250 days for their X-ray subsample, but on long timescales there was no significant difference. It is plausible that their result for short timescales was influenced by small sample size and τ effects discussed in Section 2.5.3.

2.8 Summary and Conclusions

We have used the damped random walk model of KBS09 and Kozł10 to model the optical/UV variability of ∼9000 SDSS Stripe 82 quasars with the *ugriz* light curves. The dataset includes 2.7 million photometric measurements collected over 10 years. We confirm that this is a good model of quasar variability, and quantify the dependence of two variability parameters, the long-term rms variability SF_∞ and the damping timescale τ, on physical parameters such as wavelength, luminosity, black hole mass and Eddington ratio. Our main results are the following:

1. A stochastic process with an exponential covariance function characterized by an

Table 2.2 Mean Variability Properties of Radio and X-ray Subsamples

	$<\log\tau>$	N	$<\log(\tau/\tau_p)>$	N	$<\log SF_\infty>$	N	$<\log(SF_\infty/SF_{\infty,p})>$	N
no radio	2.305 ± 0.005	6989	-0.003 ± 0.007	4467	-0.634 ± 0.003	7067	0.018 ± 0.003	4508
no radio $i < 19$	2.327 ± 0.013	1519	-0.013 ± 0.013	1279	-0.721 ± 0.006	1527	0.003 ± 0.006	1287
radio	2.26 ± 0.03	277	-0.07 ± 0.05	154	-0.70 ± 0.02	283	-0.03 ± 0.02	154
radio $i < 19$	2.24 ± 0.06	105	-0.08 ± 0.06	73	-0.81 ± 0.03	108	-0.08 ± 0.03	74
$R_g \geq 3$	2.30 ± 0.05	135	-0.01 ± 0.07	68	-0.57 ± 0.02	137	$\mathbf{0.11 \pm 0.03}$	**68**
$R_g < 3$	2.28 ± 0.03	211	0.00 ± 0.04	133	-0.73 ± 0.02	219	-0.06 ± 0.02	140
$R_g < 2$	1.99 ± 0.14	30	-0.21 ± 0.14	19	-0.88 ± 0.07	30	-0.12 ± 0.06	19
resolved[1]	2.29 ± 0.05	111	0.00 ± 0.06	63	-0.63 ± 0.02	111	0.04 ± 0.03	63
unresolved[1]	2.24 ± 0.04	167	-0.13 ± 0.07	91	-0.73 ± 0.03	172	-0.06 ± 0.03	91
x-ray	2.41 ± 0.05	81	0.03 ± 0.06	59	-0.60 ± 0.03	82	0.01 ± 0.03	58
no x-ray	2.307 ± 0.005	6950	-0.004 ± 0.007	4559	-0.638 ± 0.003	7020	0.017 ± 0.003	4598

[1] The morphological radio classes are defined using the integrated and peak FIRST fluxes as in Kimball & Ivezić (2008).

Note. – τ and SF_∞ are the observed time scales and asymptotic amplitudes of optical variability (Section 2.5), while τ_p and $SF_{\infty,p}$ refer to those predicted from Eq. 2.7, using the coefficients in the fifth and the tenth rows of Table 2.1. Errors are reported as the rms divided by $\sqrt{N-1}$, where N is the number of data points, listed after each column. Numbers of objects are lower for columns involving model estimates because black hole mass estimates are not available for all objects. Only the variability amplitude for the radio-loudest quasar subsample ($R_g \geq 3$) is significantly different ($> 3\sigma$ deviation) from the behavior of the full sample.

amplitude and timescale provides a good fit to observed quasar light curves, as shown by KBS09 and Kozł10, using smaller samples with less wavelength coverage, but better time sampling.

2. The long-term rms variability, SF_∞, has a mode at $\sim$0.2 mag and characteristic timescales, τ, are roughly 200 days in the rest frame, as found previously by KBS09 and Kozł10. These timescales are consistent with thermal timescales, but simple accretion disk models fail to reproduce the observed scaling of τ with physical parameters.

3. Quasars with similar physical parameters can have different characteristic timescales for variability. It is now clear that the distribution of τ accounts for most of the scatter in the structure function on short timescales for quasars with similar luminosity, rest wavelength, and time lag, which explains the puzzling results from MacLeod et al. (2008). Results from fitting a power law to observed ensemble structure functions should be interpreted with caution.

4. The variability timescale is correlated with the long-term rms variability with a slope of 1.30$\pm$0.01 dex/dex. Quasars that have large long-term variability amplitudes generally vary on longer characteristic timescales. The amplitude of short term variations is also correlated with τ. This conclusion is unaffected by any time sampling issues in the S82 dataset.

5. The damped random walk corresponds to a PSD proportional to $1/f^2$ at frequencies $f > (2\pi\tau)^{-1}$, flattening to a constant at lower frequencies. At large f, the data are in great agreement with $PSD \propto 1/f^2$. In terms of the structure function, this means that $SF \propto (\Delta t)^{1/2}$. Whereas previous analyses of the SF obtained a power law slope of $\beta = 0.3$, here we demonstrated that this would be a consequence of fitting the data around the "knee" (turn-over) of the SF. Our constraints for small f are much weaker. As discussed in Section 2.5.4, due to a lack of sufficient long-timescale information, we are unable to distinguish between a $1/f^0$ or a $1/f$ PSD at frequencies $f < (2\pi\tau)^{-1}$ using the data and computational technique described here.

6. The rest-frame variability parameters show a negligible trend with redshift, suggesting that they are intrinsic to the quasars, and these properties do not evolve over cosmic time for fixed physical parameters of the quasar (M_{BH}, M_i, and λ_{RF}).

7. For fixed luminosity and black hole mass, τ increases with increasing rest-frame wavelength with a power law index of 0.17, and SF_∞ decreases with a power law index of -0.48. The latter result is similar to previous findings (e.g., MacLeod et al. 2008, VB04). Kozł10 also observed that the variability increases to shorter λ, but they kept τ fixed in their fits. If wavelength is a proxy for radius in the accretion disk, this implies that the characteristic timescales are longer and the variability amplitudes are smaller in the outer regions than in the inner regions.

8. The long-term variability SF_∞ is strongly anti-correlated with luminosity as found in previous studies such as VB04, Wilhite et al. (2008), and references therein. By studying the median SF_∞ in the plane of absolute magnitude and black hole mass, we can separate the anti-correlation of amplitude with luminosity from the positive correlation with black hole mass. As suggested in Wilhite et al. (2008), these trends may be largely explained if the amplitude of variability is tied to changes in the accretion rate in the disk, and is simply related to the Eddington ratio. However, despite the strong anti-correlation between SF_∞ and L/L_{Edd} (which accounts for most of the dependence on M_i and M_{BH}), the exact dependence with M_i and M_{BH} is not consistent with L/L_{Edd} as the sole driver of quasar variability.

9. The damping timescale τ appears to be nearly independent of luminosity and correlated with M_{BH} with a power law index of 0.21 ± 0.07. The mild discrepancy with the KBS09 result (1.0 ± 0.4) may be due to the different range of sampled luminosity and black hole mass, as well as contamination by host galaxy emission in many of the very low luminosity systems they consider (see Kozł10).

10. While the mean variability parameters can be related to physical parameters, for fixed values of M_i, λ_{RF}, and M_{BH}, there is still a large scatter around the mean values,

similar to the variance of the observed distributions. Some of that scatter can be attributed to measurement and fitting errors ($\sim$60% for τ and $\sim$70% for SF_∞), but there is enough evidence for residual stochastic nature of quasar variability. Therefore, it cannot be assumed that quasars with similar M_i, λ_{RF}, and M_{BH} will necessarily have similar variability properties.

11. The radio-loudest quasars have systematically larger variability amplitude by about 30%, while the distribution of their characteristic timescale is indistinguishable from that of the full sample. There are no statistically robust differences in the characteristic timescale and variability amplitude between the full sample and a small subsample of quasars detected by ROSAT.

With this paper, and results from KBS09 and Kozł10, the ability of the damped random walk model to quantitatively describe quasar variability is well-established. As emphasized by Kozł10, this means that variability studies can become fully quantitative because the entire process of identifying and assigning parameters to quasars can be simulated to allow estimates of completeness and parameter biases. In particular, an important next step is to determine the variability equivalents of luminosity functions, i.e., the intrinsic distributions of the variability parameters. While our results represent a good first step in this direction, we caution that a non-negligible fraction of the S82 quasars have indeterminately long timescales. If, following KBS09, we identify the characteristic time scale with the thermal timescale, then the next question is whether there are additional timescales (such as the dynamical or viscous timescale), or sources of variability. For example, Blackburne & Kochanek (2010) recently found evidence for changes in disk size with changes in luminosity using gravitational microlensing. Such searches for additional sources of variability will likely require better sampled light curves and over a longer timescale.

The prospect of advancing these studies of quasar variability now faces a bottleneck. The S82 quasars have the advantage of sample size, wavelength coverage, and spectroscopy, but the light curves have poor sampling and modest overall lengths, leading to significant problems for accurately estimating τ when it is long. The quasars behind the Magellanic Clouds (Kozłowski & Kochanek 2009) are a smaller sample without spectroscopic confirmation, but

have superb, long term light curves that continue to be extended because of the continuing microlensing projects. Improving on our present results in the short term depends on either reviving the monitoring of S82 or spectroscopically confirming the Magellanic Cloud quasars.

Resuming the monitoring of Stripe 82 is challenging with the decommissioning of the SDSS imaging system. The best short-term prospects are the Pan-STARRS project (Kaiser et al. 2002), the Palomar-QUEST project (Schweitzer et al. 2006), or using the DECam being built for the Dark Energy Survey (Honscheid et al. 2008). Since the challenge is to constrain long timescales, the presence of a multiple-year gap is mainly a complication for ensuring that problems in matching photometric bands are not interpreted as a form of long-term variability. Obtaining spectra of the Magellanic Cloud quasars is in some ways easier, because the quasar magnitudes and densities are well-suited to the AAOmega fiber spectrograph on the AAT and, to a lesser degree, the IMACS spectrograph on Magellan.

The more general, Northern monitoring projects such as Pan-STARRS and Palomar-QUEST will slowly build light curves for essentially all the SDSS quasars, but at present their cadences are not ideal (see Schmidt et al. 2010) and it will take nearly a decade to build the long duration light curves needed for the analysis. In the long term, observations will be significantly improved with the advent of next-generation sky surveys. Most notably, the Large Synoptic Survey Telescope (LSST, Ivezić et al. 2008) will obtain accurate, well-sampled light curves for millions of AGN. The observed distribution of rest-frame characteristic timescales for S82 quasars spans the range from about 10 days to 1000 days (c.f. Figure 2.3). To probe the timescales as short as 0.1τ, and assuming a characteristic redshift of 2, the light curves should be sampled every 3 days in the observer's frame, which is in good agreement with the baseline cadence of LSST. With a 10 year-long survey, the length of the light curves will be in the range (1-200)τ. A combination of the SDSS, Pan-STARRS, DES, and LSST data for $\sim$10,000 Stripe 82 quasars would span well over two decades, with multi-band photometry obtained for hundreds of epochs, and would represent the best sample to date for studying the optical continuum variability of quasars. In particular, such a dataset would enable a robust measurement of the low-frequency behavior of their PSD (c.f. point 5 above). For illustration, the LSST photometric errors in the r band will be

< 0.02 mag for $r < 22$, and there are roughly 2-3 million AGN with $r < 22$ in the 20,000 sq. deg. covered by the main LSST survey (see Table 10.2 in the LSST Science Book; Abell et al. 2009). Each of these objects will be observed about 1000 times, yielding a database of over 2 billion photometric measurements. This data set, roughly a thousand times larger than that analyzed here, will enable a significant improvement in our understanding of quasar variability.

2.9 Appendix: Search for Periodic Light Curves

Although a stochastic process has proven to be an accurate statistical description of quasar light curves, any discovery of periodic behavior (even of a single source) would have interesting physical implications. However, a periodogram can only be used as an indicator of significant periodicity in a signal as compared to pure white noise (i.e., having no signal at all). Since quasars are genuinely variable, as described by a damped random walk (i.e., a red noise process; see KBS09), one can only evaluate the significance of the periodicity when also allowing for the signal covariance that is also present (see Markwardt et al. 2009; Gotz et al. 2009; Cenko et al. 2009, and references therein). Nevertheless, as a quick test for outstanding cases of periodicity among the S82 light curves, we analyzed 8,863 light curves for evidence of periodicity using the Lomb-Scargle periodogram (Lomb 1976; Scargle 1982). The threshold for considering the strongest peak in the periodogram as candidate evidence for periodicity was set following Horne & Baliunas (1986), with an adopted false alarm probability of 0.05. However, this is the threshold for ruling out white noise in favor of periodicity, and therefore provides no information on how a periodic description compares to a colored noise process such as a damped random walk. To determine the latter, a different threshold is needed (see Kozł10). When adopting the threshold for ruling out white noise in favor of periodic variability, we identify 88 light curves as good candidates.

A close inspection of the period distribution for the full sample and the selected 88 candidates shows important differences: while the full sample displays a fairly flat distribution ranging from 100 days to values exceeding 10,000 days, the period distribution for 88 candidates is bimodal. The first peak with 22 objects corresponds to aliasing at roughly one-year sampling cadence, while the second peak is centered on periods of about 6-7 years, similar

to the total length of observations. We have visually inspected the light curves and phased light curves for all 88 candidates. It turns out that candidates with proposed periods of the order one year have light curves consistent with aliasing, while those with longer periods typically have only one observed "oscillation" that might not be used as robust evidence for periodicity (six examples are shown in Figure 2.18). Therefore, our search for periodic light curves in the S82 quasar sample has not yielded any convincing cases. Again, it is important to note that in cases such as in Figure 2.18, the periodogram indicates that these sources with long τ are likely to be periodic, not because they show true periodicity, but because it computes the likelihood relative to the wrong null hypothesis (white noise rather than colored noise). Moreover, the fact that $<5\%$ of the light curves exceed the threshold for periodicity being a better description than pure white noise is a further indication that the periodogram is not a very powerful statistic for poorly sampled light curves, since it is clear from the previous Sections that quasar light curves are not white noise, but rather are well described by a damped random walk. In fact, the single "oscillation" observed for the best candidate long period objects is entirely expected from colored noise processes such as a damped random walk (see right panels of Figure 2.18). Therefore, this analysis is further evidence that a damped random walk is a good description of quasar light curves.

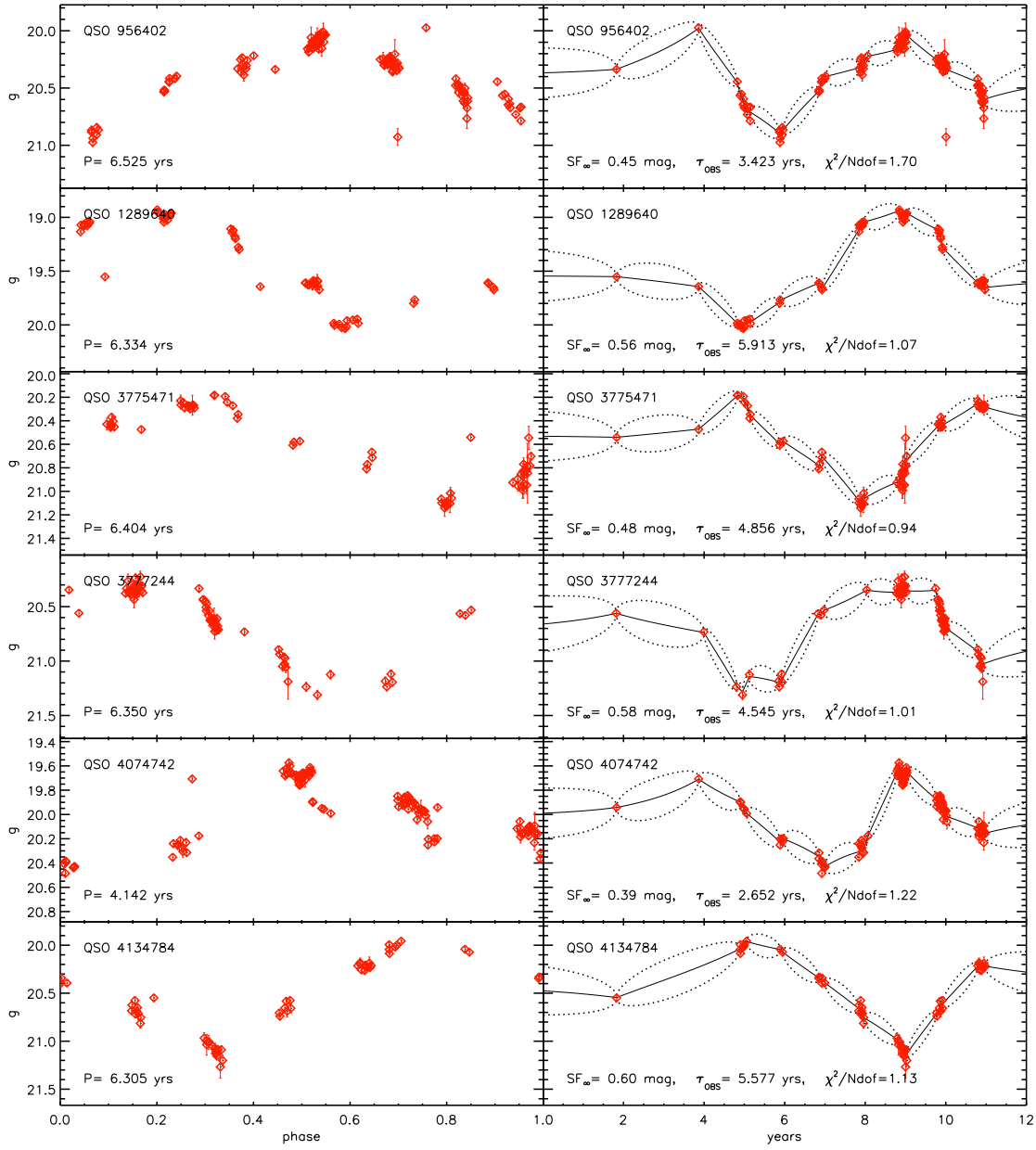

Figure 2.18 Six examples of candidate light curves showing significant periodicity as compared to pure noise (see Appendix). The phased light curves are shown on the *left*, where the phase is the fractional part of the ratio of time to the period (i.e., the phase zero-point is arbitrary). The best-fit period is listed on the bottom of each panel. The full light curves are shown on the *right*, where the solid lines show the weighted average of all consistent damped random walk model light curves, and the dotted lines show the $\pm 1\sigma$ range of possible stochastic models (see Section 2.3). The best-fit damped random walk parameters are listed at the bottom of the right panels, along with the reduced χ^2.

Chapter 3

A DESCRIPTION OF QUASAR VARIABILITY MEASURED USING REPEATED SDSS AND POSS IMAGING

3.1 Abstract

The contents of this Chapter, as well as the Appendix which describes the catalog format, are published in MacLeod et al. (2011a).

We provide a quantitative description and statistical interpretation of the optical continuum variability of quasars. The Sloan Digital Sky Survey (SDSS) has obtained repeated imaging in five UV-to-IR photometric bands for 33,881 spectroscopically confirmed quasars. About 10,000 quasars have an average of 60 observations in each band obtained over a decade along Stripe 82 (S82), whereas the remaining ∼25,000 have 2–3 observations due to scan overlaps. The observed time lags span the range from a day to almost 10 years, and constrain quasar variability at rest-frame time lags of up to 4 years, and at rest-frame wavelengths from 1000Å to 6000Å. We publicly release a user-friendly catalog of quasars from the SDSS Data Release 7 that have been observed at least twice in SDSS or once in both SDSS and the Palomar Observatory Sky Survey, and we use it to analyze the ensemble properties of quasar variability. Based on a damped random walk (DRW) model defined by a characteristic timescale and an asymptotic variability amplitude that scale with the luminosity, black hole mass, and rest wavelength for individual quasars calibrated in S82, we can fully explain the ensemble variability statistics of the non-S82 quasars such as the exponential distribution of large magnitude changes. All available data are consistent with the DRW model as a viable description of the optical continuum variability of quasars on timescales of ∼5–2000 days in the rest frame. We use these models to predict the incidence of quasar contamination in transient surveys such as those from PTF and LSST.

3.2 Introduction

The optical continuum variability of quasars has been recognized since their first optical identification (Matthews & Sandage 1963), and it has been proposed and utilized as an efficient method for their discovery (van den Bergh, Herbst, Pritchet 1973; Hawkins 1983; Hawkins & Veron 1995; Ivezić et al. 2004a; Rengstorf et al. 2006). The observed characteristics of the variability can then be used to constrain the origin of their emission (e.g., Kawaguchi et al. 1998; Trevese, Kron & Bunone 2001, and references therein). The variability of quasars has typically been quantified using a structure function (SF) analysis (e.g., Hughes et al. 1992; Collier & Peterson 2001; Bauer et al. 2009; Kozłowski et al. 2010b; Welsh, Wheatley, & Neil 2011), where the SF is the root-mean-square (rms) magnitude change (Δm) as a function of the time lag (Δt) between measurements (similar to an auto-correlation function). It is fairly well established that quasar variability properties depend on physical properties such as the quasar luminosity, wavelength, the timescale considered, and the presence of radio emission. However, despite the considerable observational effort invested over last few decades, many conflicting claims about various correlations exist in the literature (see Giveon et al. 1999 for a detailed discussion).

The traditional method for studying variability has been to monitor a small, select sample of quasars over a long time baseline (e.g., Hawkins 2002; Giveon et al. 1999; Rengstorf et al. 2004). In this case, it is possible to compute the SF for each quasar, which can later be sample-averaged or studied individually. An alternative, utilized in more recent studies based on Sloan Digital Sky Survey (SDSS; York et al. 2000) and Palomar Observatory Sky Survey (POSS; Minkowski & Abell 1963) data, is to compute a single SF for all quasars in a particular wavelength or luminosity range. This approach, mandated by the fact that typically only a few epochs were available per object, only measures *ensemble* properties and *assumes* that all quasars selected from a narrow range of physical properties vary in the same way. Nevertheless, this approach enabled studies of quasar optical variability based on tens of thousands of objects and several hundred thousand photometric observations, as well as explorations of the long-term variability (Vanden Berk et al. 2004, hereafter VB04; Ivezić et al. 2004c, hereafter I04; Wilhite et al. 2005; Mahabal et al. 2005; De Vries, Becker,

White, & Loomis 2005, hereafter dV05; Sesar et al. 2006, hereafter Ses06). For example, the size and quality of the sample analyzed by VB04 (two-epoch photometry for 25,000 spectroscopically confirmed quasars) allowed them to constrain how quasar variability in the rest frame optical/UV regime depends upon rest frame time lag (up to ∼2 years), luminosity, rest wavelength, redshift, the detection of radio or X-ray emission, and the presence of broad absorption line systems. By comparing SDSS and POSS measurements for ∼20,000 quasars spectroscopically confirmed by the SDSS, Ses06 constrained the optical quasar variability on timescales from 10 to 50 years (in the observer's frame). They report that there is a characteristic timescale of order 1 year in the quasar rest frame beyond which the SF flattens to a constant value. The SDSS has also facilitated both individual- and ensemble-based approaches by providing a large multi-epoch sample of quasars over the Northern Galactic Cap and well-sampled light curves in the Southern Stripe 82 (S82) survey. It is reassuring that the two approaches lead to similar SFs, as discussed by dV05. A test of this assumption is also described in MacLeod et al. (2008), who show that indeed the *mean* behavior is the same. With such large samples, the ensemble SF(Δt) slopes are well-constrained, and the values suggest that accretion disk instabilities are the most likely mechanism causing the observed optical variability (VB04; Kawaguchi et al. 1998; see also Lyubarskii 1997). However, attempts to constrain physical models using the ensemble SF are invalid as soon as one realizes that the ensemble SF(Δt) is a weighted sum of individual quasars with different SFs (MacLeod et al. 2008).

While studies have traditionally examined "non-parametric" statistical measures of variability such as the SF, a major challenge has been to describe the variability of individual quasars in a compact way. Recently, the introduction of a damped random walk (DRW) model has provided a way to mathematically characterize quasar light curves in terms of a characteristic timescale (τ) and an amplitude (SF_∞) which are then correlated with the physical properties such as luminosity and black hole mass. Kelly et al. (2009, hereafter KBS09) modeled a sample of 100 quasar light curves as a DRW and suggested that thermal fluctuations driven by an underlying stochastic process such as a turbulent magnetic field may be the dominant cause for the optical flux fluctuations. Kozłowski et al. (2010a; hereafter Kozł10) applied the DRW model to the well-sampled Optical Gravitational Lensing

Experiment (OGLE) light curves (Udalski et al. 1997; Udalski et al. 2008) of mid-infrared-selected quasars behind the Magellanic Clouds from Kozłowski & Kochanek (2009). Their analysis shows that the DRW model is robust enough to efficiently select quasars from other variable sources, despite the large surface density of foreground Magellanic Cloud stars (see also Butler & Bloom 2011; MacLeod et al. 2011; Kozłowski et al. 2011).

In MacLeod et al. (2010, hereafter Mac10), we applied the DRW model to the light curves of ∼10,000 quasars in S82 and found a correlation between SF_∞ and black hole mass which is independent of the anti-correlations with luminosity and wavelength (see also Ai et al. 2010; Meusinger et al. 2011). We also found that τ increases with increasing wavelength, remains nearly constant with redshift and luminosity, and increases with increasing black hole mass (see also KBS09; Kozł10). In Kelly et al. (2011), it was shown that a similar stochastic model but with multiple timescales for a single object can accurately reproduce the X-ray variability of active galactic nuclei (AGN) and microquasars. An inhomogeneous accretion disk model, where the temperature fluctuations throughout the disk are driven by a DRW process, can explain the disk sizes derived from microlensing light curves (see Morgan et al. 2010) while matching the observed level of optical variability, and predicts SEDs which are in better agreement with observations than standard thin disk models (Dexter & Agol 2011).

One defining feature of the DRW model for a *single* quasar is that it predicts a Gaussian distribution of magnitude differences Δm for a given Δt. On the other hand, the observed Δm distribution in the optical for an *ensemble* of quasars observed at two times separated by Δt deviates strongly from a Gaussian but is well fit by an exponential distribution (I04). This conflict represents an important puzzle for understanding the DRW model and its applicability to quasar light curves. Also, the high likelihood of extreme values of Δm has important implications for the interpretation of observations of transients. For example, Vanden Berk et al. (2002) reported the detection of an orphan gamma-ray burst afterglow based on the 2.5 magnitude decrease in optical flux. Such a large flux change was inconsistent with a quasar based on a Gaussian model for their variability, but the source was nevertheless confirmed to be a highly variable quasar (Gal-Yam et al. 2002). An accurate statistical description of the two-epoch photometry for ensembles of quasars

will be important for transient detection in large surveys, where quasars represent a major contaminant.

Our goal here is to produce a unified view of ensemble and individual optical variability in the context of the DRW model. In this study, we show that the differences in shape between the ensemble SF and the DRW SF are well-explained by averaging over the properties of individual quasars. We also show that the exponential distributions of magnitude changes for ensembles of quasars at fixed time lag are naturally constructed by summing the intrinsically Gaussian distributions of magnitude changes produced by individual quasars. There are several residual issues which we discuss as part of the comparison. An overview of the SDSS and POSS data used in this study is presented in Section 3.3. In Section 3.4, we describe the observed ensemble quasar variability in terms of the SF as a function of wavelength and time lag in the observer's frame. We then convert to rest-frame quantities and compare the data to a model ensemble SF based on our previous DRW analysis of S82 light curves. In Section 3.5, we combine the constraints on short-term quasar variability based on SDSS data with the constraints on long-term variability derived from matching the SDSS and POSS catalogs. In Section 3.6, we discuss the implications our results have on transient identification, with a focus on future time-domain surveys such as the Palomar Transient Factory and the Large Synoptic Survey Telescope. Our results are discussed and summarized in Section 3.7.

3.3 Data Overview

In this section, we briefly summarize the relevant SDSS and POSS data. We focus our description on the quasars with multiple observations (similar to MacLeod et al. 2010; 2011).

3.3.1 The Basic Characteristics of the SDSS Imaging Survey

The SDSS provides homogeneous and deep ($r < 22.5$) photometry in five passbands (u, g, r, i, and z, Fukugita et al. 1996; Gunn et al. 1998; Smith et al. 2002; Ivezić et al. 2004b) accurate to 0.02 mag, of up to 12,000 deg^2 in the Seventh Data Release (DR7, Abazajian et al. 2009). The DR7 sky coverage results in photometric measurements for about 357

million unique objects. Astrometric positions are accurate to better than 0.1 arcsec per coordinate (rms) for sources brighter than 20.5 mag (Pier et al. 2003), and the morphological information from the images allows robust star–galaxy separation to ∼21.5 mag (Lupton et al. 2001). A compendium of technical details about the SDSS can be found in Stoughton et al. (2002).

The SDSS offers an unprecedented photometric accuracy for such a large scale optical survey. Not only are the photometric errors generally small, but they are accurately determined by the photometric pipeline (*photo*; Lupton et al. 2001) and can be reliably used to estimate the statistical significance of measured magnitude differences. This ability is of paramount importance for a robust statistical study of variable objects. This error behavior is illustrated in Figure 2 in Ivezić et al. (2003). Throughout our analysis, we assume the SDSS photometric errors to be $\sigma_{phot} = 0.018$ mag in g, r, and i bands, and 0.04 mag in u and z (see also Figure 1 in Sesar et al. 2007). The photometric errors are assumed to be independent of magnitude due to the bright sample limit of $i < 19.1$ (see the next section), and for simplicity when accounting for photometric errors in the SF calculations.

3.3.2 The SDSS Multi-epoch Data

The SDSS imaging data are obtained by imaging the sky in six parallel scanlines, each 13.5 arcmin wide (a "strip" in SDSS terminology). The six scanlines from two adjacent scans are then interleaved to make a filled "stripe". Because of the scan overlaps, and because the scans converge near the survey poles, ∼40% of the sky in the northern survey is essentially surveyed twice. In addition, 290 deg^2 of the southern survey area lies along S82 and has been observed about 60 times on average to search for variable objects and, by stacking the frames, to go deeper (Frieman et al. 2008; Annis et al. 2011). This valuable subsample contains well-sampled light curves for 9258 spectroscopically confirmed quasars whose variability properties are analyzed in Mac10, and it can be used to verify some of the results inferred from the analysis of two repeated observations. Overall, the SDSS has obtained multi-epoch data for ∼4,000 deg^2 of sky, with timescales ranging from 2 hours to over 9 years, and with a wide range of Galactic latitudes extending all the way to the

Galactic plane.

We define a quasar as any object listed in the SDSS catalog of spectroscopically confirmed quasars (the "DR7 Quasar Catalog", Schneider et al. 2010). Its most recent fifth edition lists the SDSS DR7 BEST photometry for 105,783 quasars from 9380 deg^2. For a description of the spectroscopic target selection for quasars, see Richards et al. (2002). We note that the quality of photometry is good for the objects in the DR7 Quasar Catalog. In total, there are 33,881 spectroscopically confirmed quasars with at least two observations. We provide a catalog of all the SDSS repeated imaging of quasars online[1] (see the Appendix for a detailed catalog description). The sky distribution of repeatedly imaged quasars is shown in Figure 3.1.

When we compare pairs of observations, we use the secondary imaging observation with the largest time lag from the primary imaging observation listed in the DR7 Quasar Catalog, for a total of two observations per quasar. We define $\Delta m = m_2 - m_1$, where m_2 is the magnitude observed at a later epoch than m_1. We omit the S82 quasars since we do not wish to test our model on the same sample of objects from which the DRW model parameters were derived. The DR7 Quasar Catalog is spectroscopically complete to $i < 19.1$ in the quasar region of color space. Therefore, unless otherwise noted, we require that the primary observation for the remaining 24,627 quasars has $i \leq 19.1$ in order to be consistent with the flux limit of the quasar sample, for a total of 14,939 quasars. Given the SDSS limit of $\sim$22 mag, we further restrict the magnitude differences in each band to $|\Delta m| < 3$ mag in order to reduce the amount of contamination by poor photometry, for a total of $\sim$80,000 pairs of SDSS measurements (summed over all bands). Finally, we adopt the redshifts and absolute magnitudes listed in the Schneider et al. (2010) spectroscopic quasar catalog, and the black hole masses as measured from emission lines by Shen et al. (2011). All magnitudes are corrected for Galactic extinction using Schlegel et al. (1998).

[1] http://www.astro.washington.edu/users/ivezic/macleod/qso_dr7/

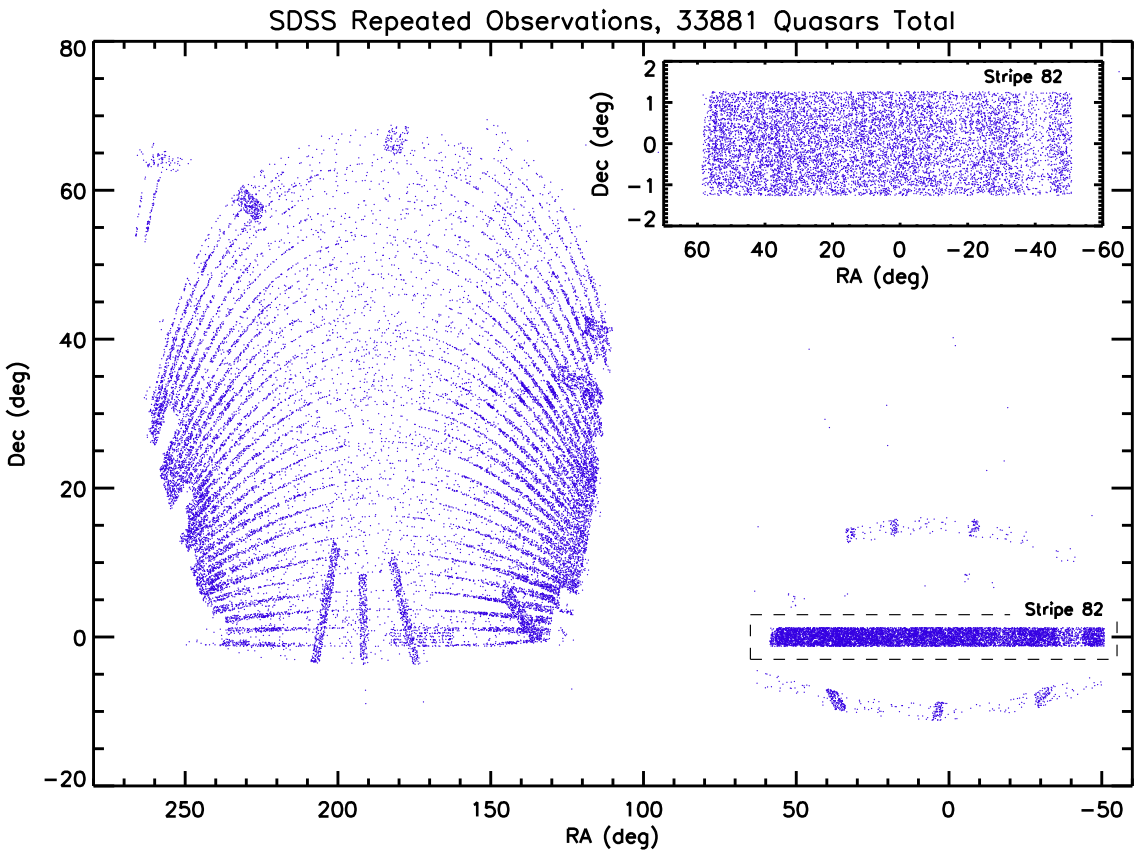

Figure 3.1 Distribution of 33,881 DR7 quasars with at least 2 observations on the sky in equatorial coordinates. The number of observations increases towards the survey poles where the "stripes" overlap. The inset shows a closer view of the SDSS Stripe 82, which has ∼60 observations per object (the density of quasars is non-uniform in S82 due to the increasing Galactic contribution toward negative R. A.).

3.3.3 SDSS–POSS Long-term Measurements

While SDSS S82 obtained measurements with time differences of up to 10 years, longer timescales of up to 50 years can be probed by comparing the SDSS and POSS catalogs. Ses06 have addressed the problem of large systematic errors in POSS photometry by recalibrating several publicly available POSS catalogs (USNO-A2.0, USNO-B1.0, DPOSS and GSC2.2). A piecewise recalibration of the POSS data in 100 arcmin^2 patches (one SDSS field) generally resulted in an improvement of photometric accuracy (rms) by nearly a factor of two compared to the original data (POSS-I magnitudes can be improved to ∼0.15 mag

accuracy, and POSS-II magnitudes to ∼0.10 mag accuracy). In addition to the smaller core width of the error distribution, the tails of the distribution become much steeper after the recalibration. These improvements are mostly due to the very dense grid of calibration stars provided by the SDSS, which rectifies the intrinsic inhomogeneities of Schmidt plates.

The much longer time lags between the Palomar Observatory Sky Surveys (POSS-I and POSS-II) and the SDSS (up to ∼50 years) than spanned by the available SDSS data make it easier to detect deviations of the SF from a simple power law. dV05 and Ses06 compared the SDSS and POSS data for over 10,000 quasars from the SDSS Data Release 2 in order to constrain the long-term quasar variability. As discussed by Ses06, the dV05 and Ses06 measurements of the SF agree within a 1σ uncertainty of the Ses06 measurements. Here, we use results from both the dV05 and Ses06 studies. We also make use of data from the Digitized Palomar Observatory Sky Survey (DPOSS; Djorgovski et al. 1998) that overlap 8,000 deg^2 of sky from the SDSS Data Release 5. These DPOSS data were recalibrated following the procedure outlined by Ses06. There are 81,189 SDSS DR7 quasars with DPOSS observations in the G, R, and I bands, and we provide their two-epoch photometry at the same Web site (see footnote 1). For our analysis in Section 3.5, we use the primary SDSS observations as listed in the DR7 quasar catalog when comparing SDSS–DPOSS magnitudes. We calculate synthetic POSS GRI magnitudes from the SDSS photometry, require $i < 19.1$, and impose various quality cuts following Ses06. In total, we have 56,732 SDSS–DPOSS Δm measurements.

3.4 Characteristics of Observed Variability

For our SF analysis, we examine the behavior in both the observer's and the quasar rest frames because the former is informative for data interpretation (e.g., transients in large surveys) and the latter constrains quasar physics. The SF behavior in the observer's frame is discussed in Section 3.4.1, and we convert to rest-frame quantities in Section 3.4.2. We handle various trends with observed properties by taking narrow ranges of the relevant quantities, in particular, the time lag and wavelength in each frame. The luminosity, mass, and redshift information are utilized in Section 3.4.2, where we test whether the scalings with these physical parameters derived from individual light curves in S82 can reproduce

the ensemble variability of the two-epoch sample.

3.4.1 Quasar Variability in the Observer's Frame

Assuming that the observed variability reflects the physics in the accretion disk, the behavior in the observer's reference frame will be a convolution of rest-frame variability over redshift, luminosity, and other parameters. While this convolution will obscure important physical scalings, quasar variability in the observer's reference frame is still of major interest when interpreting survey data, in particular when distinguishing between quasars and other variable sources such as transients, or in using variability to select quasars. Therefore, we start by considering quasar variability in the observer's frame.

Δm Distribution

Figure 3.2 shows the cumulative distribution of Δm in the SDSS u, r and z bands for four different Δt ranges. The thin lines show the predicted distributions based on Gaussian (dotted) and exponential (dashed) analytic functions with the same rms as the data. The photometric errors are taken into account by adding a Gaussian component of width $\sigma_{phot} =$ 0.018 mag in r and 0.04 mag in u and z. The exponential curves predict a much higher probability of large magnitude changes than the Gaussian curves, and the data follow this prediction (we only show the data for bins with more than 10 points). The data points become increasingly unreliable in the tails of each distribution due to small sample sizes, and this fact may cause the large discrepancy with the exponential curve for small Δt in the r and z bands.

Figure 3.3 shows the differential Δm distributions for each SDSS band and for two slices of observed-frame time lag. The distribution width (i.e., the SF value) increases with time lag and decreases with wavelength. The wings of the distributions are also closer to exponential (more accurately, a double-exponential or Laplace distribution) than to Gaussian. Figure 3.4 shows similar distributions for the changes in color for two slices in time lag. The color changes are smaller than for the individual magnitudes at the same time separation. In the bottom six panels, the color changes are plotted against

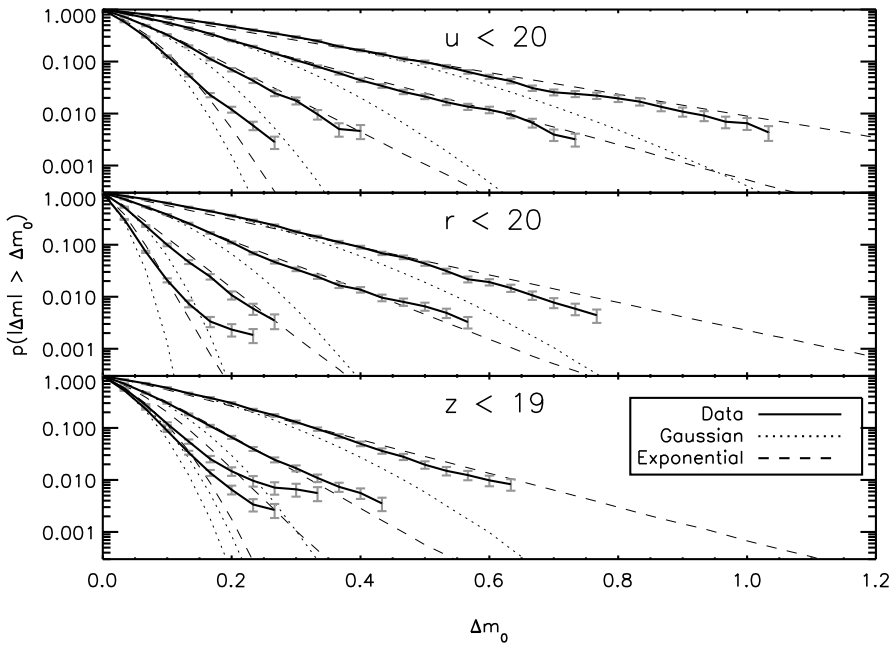

Figure 3.2 The observed cumulative distributions of urz magnitude differences as a function of time lag in the observer's frame. We define $\Delta m = m_2 - m_1$, where m_2 is the magnitude observed at a later epoch than m_1. The data (thick lines with Poissonian error bars) are only shown for sample sizes exceeding 10 data points (the probabilities become increasingly unreliable when below $\sim$0.003). Going from the narrowest distribution to the widest, the time lags span 1–30 days, 50–150 days, 200–400 days, and 1400–3000 days. The thin lines show the predicted distributions based on Gaussian (dotted) and exponential (dashed) analytic functions with the same rms as the data for each time lag range, where the photometric errors are taken into account by adding a Gaussian component of width σ_{phot} as defined in Section 3.3.1.

each other as well as against the r-band differences. Although the scatter around these relationships is quite large, quasars tend to get bluer as they brighten on average (bottom four panels), in accordance with previous results (e.g., Giveon et al. 1999). An analogous correlation was found recently by Schmidt et al. (2012), who analyzed individual light curves. Their study showed the slope s_{gr} between the r-band and g-band variations, as in $r - \langle r \rangle = (s_{gr} + 1)(g - \langle g \rangle) + b$, is on average around -0.2, indicating that quasars get bluer as they brighten. This value corresponds to a slope of 4 in the bottom-right panel of Figure 3.4 (equating $m - \langle m \rangle$ to Δm), shown by the solid line, which approximately follows the observed, two-epoch distribution.

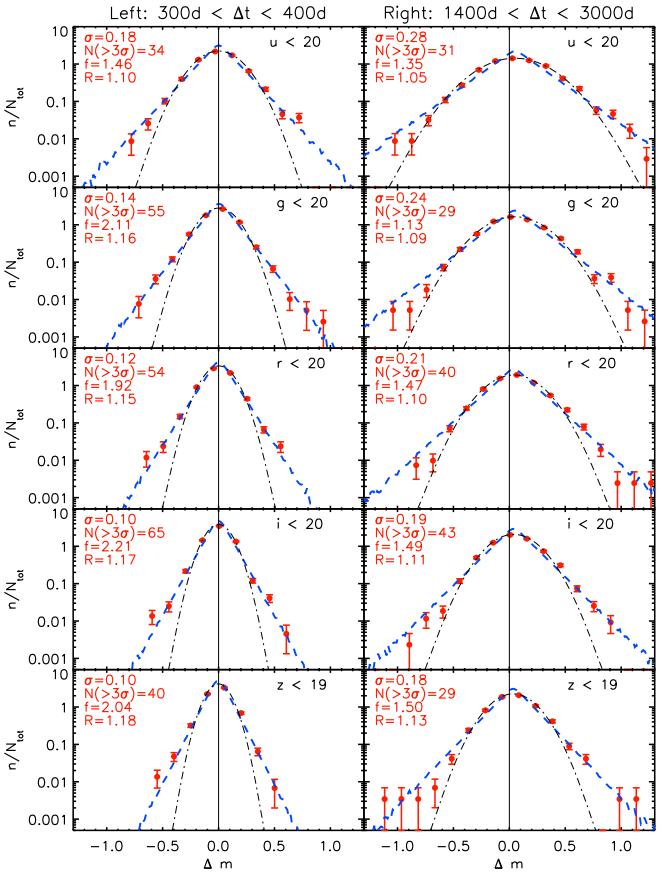

Figure 3.3 The symbols show the distribution of measured magnitude differences, Δm, in the five SDSS bands, and for two narrow ranges of time lag, Δt, as indicated at the top. n indicates the number of points in a bin divided by the bin width, and N_{TOT} is the total number of points used for each histogram. The distribution width (σ) calculated using Eq. 3.1, shown in each panel, is increasing with time and decreasing with wavelength. The dot-dashed lines show Gaussian distributions with a root-mean-square (rms) equal to σ, and the dashed lines show exponential distributions, $\exp(-\Delta m/\Delta_c)$, with $\Delta_c = \mathrm{rms}/\sqrt{2}$. Note that an exponential distribution provides a better fit to the wings of the observed distributions. Each panel also displays the number of measurements outside the $\pm 3\sigma$ range ($N > 3\sigma$), their corresponding fraction (f) expressed in percent (the expected value of this fraction for a Gaussian distribution is 0.3%, and for an exponential distribution 1.4%), and the ratio $R = \mathrm{rms}/\sigma$.

Structure Function

Next, we quantify the quasar variability using an SF analysis. Although several definitions can be found in the literature, the SF essentially measures the rms magnitude difference as

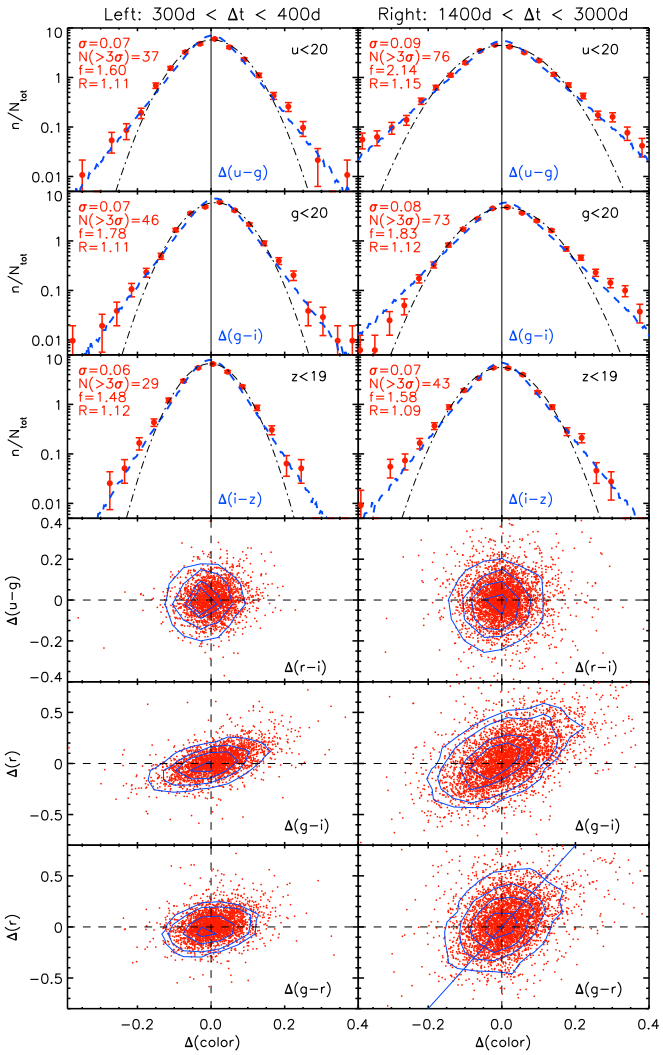

Figure 3.4 The top three rows are analogous to Fig. 3.3, except that we show the changes in the $u-g$, $g-i$, and $i-z$ colors for two time lags in the observer's frame (left: 300–400 days, right: 1400–1600 days). The fourth row shows the change of $u-g$ color as a function of the change of $r-i$ color, and the bottom two rows shows the change of the r band magnitude as a function of the change of $g-i$ or $g-r$ color. The contours show regions containing 25%, 50%, 75% and 90% of the data points. The solid line in the bottom-right panel shows the slope found by Schmidt et al. (2012) using individual quasar light curves.

a function of time lag between magnitude measurements. We compute the SF as

$$\mathrm{SF} = 0.74(\mathrm{IQR})/\sqrt{N-1}, \tag{3.1}$$

where IQR is the 25% – 75% interquartile range of the Δm distribution, and N is the number of Δm values. This approach is insensitive to outliers in the data which may result from poor data quality and is especially effective at short time lags where the SF is small. This value is equivalent to the rms if the distribution is Gaussian. When stated, the photometric errors are taken into account by subtracting $\sqrt{2}\sigma_{phot}$ in quadrature from the SF. Two other definitions of the SF are found in the literature. Following Bauer et al. (2009), we refer to them as $\mathrm{SF}^{(A)}$ and $\mathrm{SF}^{(B)}$:

$$\mathrm{SF}^{(A)}(\Delta t) = \sqrt{\langle \Delta m^2 \rangle} \tag{3.2}$$

$$\mathrm{SF}^{(B)}(\Delta t) = \sqrt{\frac{\pi}{2}\langle |\Delta m| \rangle^2}. \tag{3.3}$$

For a Gaussian distribution, the ratio of the form adopted here to the other two forms is $\mathrm{SF}/\mathrm{SF}^{(A)} = \mathrm{SF}/\mathrm{SF}^{(B)} = 1$, while for an exponential (Laplace) distribution, $\mathrm{SF}/\mathrm{SF}^{(A)} = 1.03/\sqrt{2}$ and $\mathrm{SF}/\mathrm{SF}^{(B)} = 0.82$.

Figure 3.5 compares the SF for the combined SDSS and POSS data over timescales ranging from 5 days to 50 years in the observer's frame. The SF values shown up to $\Delta t \simeq 2000$ days are computed using Eq. 3.1, with the photometric errors subtracted. For variability in the g band with time lags of 10 and 50 years, we adopt 0.35 and 0.43 mag to represent the dV05 results, and 0.21 and 0.40 mag for the Ses06 results, respectively. The errors in these estimates are ∼0.05-0.10 mag. The short-dashed line shows the best power law fit to the SDSS measurements alone ($\mathrm{SF} \propto \Delta t^{0.44}$). It overestimates the variability level for timescales longer than several years. The long-dashed line shows the best-fit asymptotic function to all data points given by $S_\infty\,[1-\exp(-\Delta t/\tau_{\mathrm{SF}})]^{1/2}$, a common parameterization of the SF (e.g., Hook et al. 1994), with S_∞ and τ_{SF} as free parameters. The limiting value of the best-fit overall SF, $S_\infty = 0.28 \pm 0.01$ mag, is probably systematically uncertain at a level of 0.02–0.03 mag due to the discrepancies between the SDSS–POSS-I/II measurements (see Ses06). The measured characteristic timescale, $\tau_{\mathrm{SF}} \sim 1400$ days, corresponds to ∼700 days in the rest frame and is uncertain by ∼10–20%. While this analysis and the earlier ones

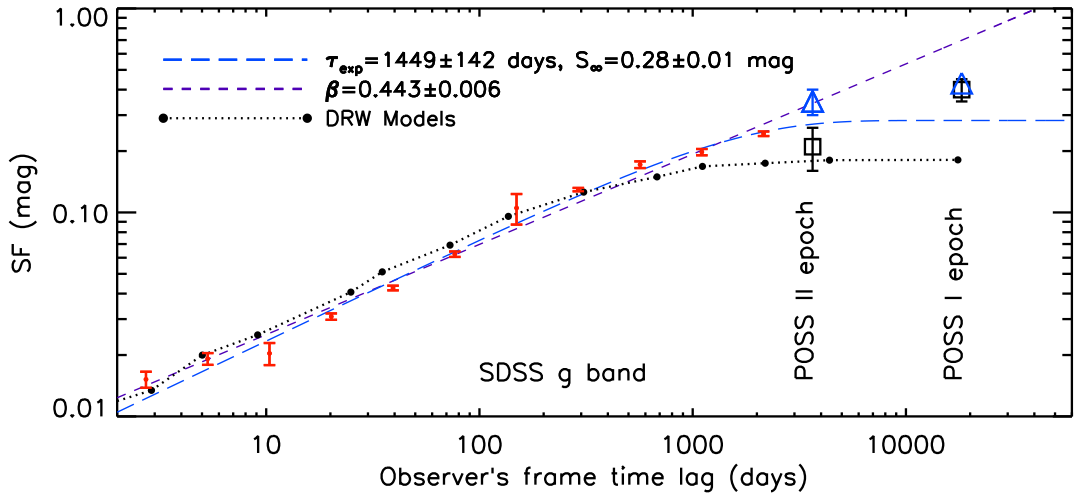

Figure 3.5 The structure function for quasar variability measured in the g band and in the observer's frame (corrected for errors using $SF = \sqrt{SF^2_{obs} - 2\sigma^2_{phot}}$). The red data points show the SDSS measurements, and the data points at $\Delta t > 3000$ days are inferred from comparing SDSS data and POSS I/II data, taken from Ses06 (open squares) and de Vries, Becker, and White (2003, blue triangles, which have similar errors as those from Ses06). The short-dashed line shows the best power law fit to the SDSS measurements alone, $SF \propto \Delta t^{\beta}$, and the long-dashed line shows a simultaneous fit to all data points, $SF = S_{\infty}[1 - \exp(-\Delta t/\tau_{\rm SF})]^{1/2}$. The best-fit parameters for these two fits are listed in the upper-left corner. The dotted line shows the prediction of the DRW model trained on S82, as described in Section 3.4.2.

by dV05 and Ses06 provide strong evidence that the SF levels out on long timescales, we cannot rule out a continuing but slower rise.

3.4.2 Quasar Variability in the Rest Frame

The redshift distribution of the quasars enables one to map a discrete distribution of wavelengths and time differences Δt in the observer's frame to a smoother distribution in the quasar rest frame. Figure 3.6 shows the distribution of our 79,787 Δm measurements for $i < 19.1$ objects in the rest-frame time difference (Δt_{RF}) and wavelength (λ_{RF}) plane. Most (54%) observations have $\Delta t_{RF} < 50$ days. The discrete distribution[2] of λ_o and Δt_o in the

[2]Note that a discretely-distributed observed Δt can lead to artificial correlations between the SF and Δt_{RF}. For fixed Δt_o, longer Δt_{RF} correspond to both lower redshift and, because of magnitude limits,

observer's frame is spread along the lines $\Delta t_{RF} = k\,\lambda_{RF}$, with $k = \Delta t_o/\lambda_o$, and according to the sample redshift distribution [$\Delta t_{RF} = \Delta t_o/(1+z)$, $\lambda_{RF} = \lambda_o/(1+z)$]. Here, we consider the shape of the Δm distribution as a function of these quantities, the form of SF(Δt_{RF}), and the dependence of the SF on all physical parameters simultaneously, including the i-band absolute magnitude M_i (K-corrected to the rest frame) and the black hole mass M_{BH}.

Application of a DRW model

We develop a model SF which reproduces the observed ensemble variability using information derived from individual quasar light curves in S82. Our model is motivated by the success of a DRW in describing quasar light curves (KBS09; Kozł10; Mac10), as well as the detection of a turnover in the ensemble SF on long timescales, suggesting a characteristic timescale for variability (e.g., Ses06; Welsh, Wheatley, & Neil 2011). We assume that the turnover is a consequence of the characteristic timescale distribution measured among individual quasars, and that the turnover in the ensemble SF corresponds to the average timescale, $\tau_{\rm SF}$. Suppose each quasar is described by its own DRW parameters, τ and SF$_\infty$. Then, the ensemble SF is a weighted contribution of all the individual SFs over their distribution in these parameters,

$$\mathrm{SF}(\Delta t) = \int d\tau d\mathrm{SF}_\infty \frac{d^2n}{d\tau d\mathrm{SF}_\infty} \mathrm{SF}(\Delta t|\tau, \mathrm{SF}_\infty)_{qso}, \tag{3.4}$$

where SF$(\Delta t|\tau, SF_\infty)_{qso}$ is the SF at time Δt for a quasar with DRW variability parameters τ and SF$_\infty$, given by

$$\mathrm{SF}(\Delta t|\tau, \mathrm{SF}_\infty)_{qso} = \mathrm{SF}_\infty(1 - e^{-|\Delta t|/\tau})^{1/2}. \tag{3.5}$$

In particular, SF$(\Delta t)_{qso}$ is the expected standard deviation of magnitude differences Δm for a given quasar at a time lag Δt. SF$(\Delta t)_{qso}$ is related to the standard deviation in

lower luminosities. Since there is also an anti-correlation between variability amplitude and luminosity, this leads to a SF that increases toward higher Δt_{RF}. Therefore, caution must be taken when binning in Δt_{RF} using very sparse Δt_o so that one does not mistake wiggles in the ensemble SF(Δt_{RF}) for multiple intrinsic timescales, for example.

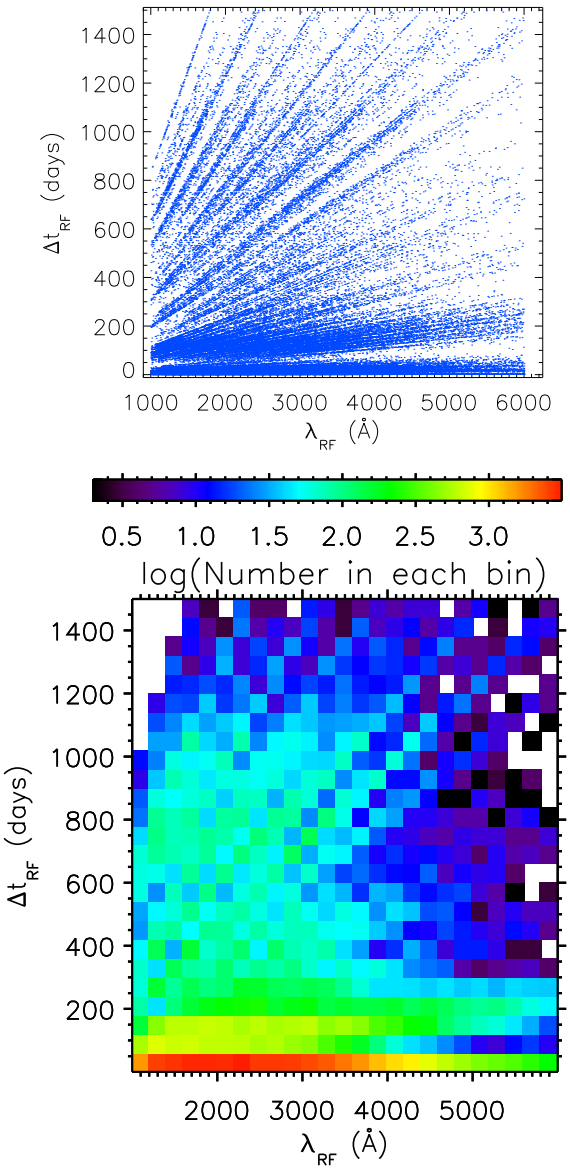

Figure 3.6 (Top) The distribution of SDSS magnitude difference measurements in the plane spanned by rest-frame wavelength, λ_{RF} and time lag, Δt_{RF}. The discrete distribution of λ_o and Δt_o in the observer's frame is spread along the lines $\Delta t_{RF} = k\,\lambda_{RF}$, with $k = \Delta t_o/\lambda_o$, and according to the sample redshift distribution [$\Delta t_{RF} = \Delta t_o/(1+z)$, $\lambda_{RF} = \lambda_o/(1+z)$]. The bottom panel shows the number of measurements in each bin on a log scale according to the legend at top. The white pixels contain fewer than 5 data points.

magnitudes (σ_m) at a given Δt by $\mathrm{SF}(\Delta t)_{qso} = \sqrt{2}\sigma_m$, where the factor of $\sqrt{2}$ results from subtracting two magnitudes.

To build a model for the ensemble variability, we follow these steps for each quasar in the two-epoch sample:

1. Predict τ and SF_∞ based on the quasar's physical parameters.

2. Include the intrinsic scatter in τ and SF_∞ for quasars with similar physical parameters.

3. Estimate the SF value at the measured time lag Δt_{RF} using Eq. 3.5.

4. Draw one model Δm value from a Gaussian distribution with a width set by $\mathrm{SF}(\Delta t_{RF})_{qso}$, adding photometric noise if necessary.

We generally average over 1000 of these Monte Carlo models for the distributions expected from the DRW model.

In the first step, τ and SF_∞ are estimated using the scalings of the DRW parameters with the quasar's physical parameters found in Mac10:

$$\log f = A + B \log\left(\frac{\lambda_{RF}}{4000\text{Å}}\right) + C(M_i + 23) + D \log\left(\frac{M_{BH}}{10^9 M_\odot}\right) + E \log(1+z), \qquad (3.6)$$

where $A = -0.51$, $B = -0.479$, $C = 0.131$, and $D = 0.18$ for $f = \mathrm{SF}_\infty$ (in mag), and $A = 2.4$, $B = 0.17$, $C = 0.03$, and $D = 0.21$ for $f = \tau$ (in days). Note that the dependence on redshift was found to be negligible in both cases ($E = 0$). By applying this model to the quasars in the large two-epoch sample, we are also testing the accuracy of the scalings derived for the S82 quasars.

In the second step, we account for the intrinsic scatter in τ and SF_∞ for quasars with similar physical parameters. When measured for individual quasar light curves, τ and SF_∞ show scatter about their mean trends (Eq. 3.6). The magnitudes of these residuals are too large to be fully attributed to measurement uncertainties (Mac10; Bauer et al. 2011). From simulations (see Section 4.2 of Mac10), we estimated that the latter statistical uncertainties account for only 70%, 60%, and 13% of the scatter in τ, SF_∞, and $\hat{\sigma} = \mathrm{SF}_\infty/\sqrt{\tau}$, respectively. The fitting errors are much smaller for $\hat{\sigma}$ because it is well constrained even for light

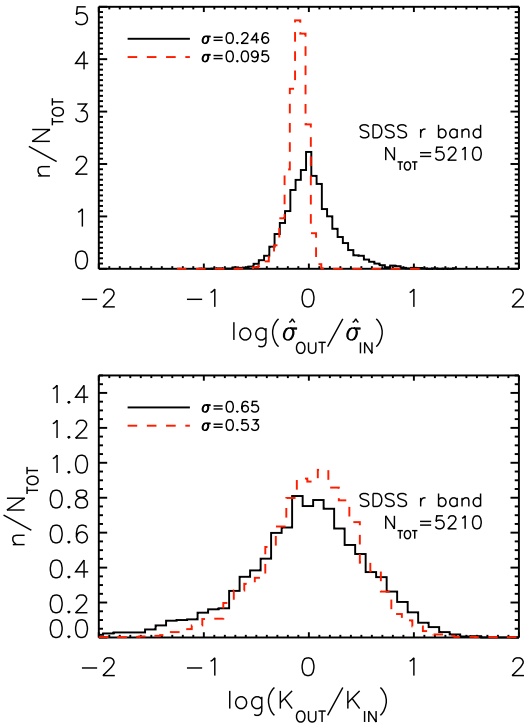

Figure 3.7 Determination of fitting errors. The distributions of DRW parameters $\hat{\sigma} = \mathrm{SF}_\infty/\sqrt{\tau}$ and $K = \tau\sqrt{\mathrm{SF}_\infty}$ for Monte Carlo models of S82 quasar light curves are shown in the top and bottom panels, respectively. The solid, black curve in each panel shows the distribution of observed $\hat{\sigma}$ (K) values, which are used as inputs to generate the light curves, normalized by the median value of 0.16 mag yr$^{-1/2}$ (200 mag$^{1/2}$ days). The red, dashed line shows the ratio between the best-fit parameters for the simulated light curves and their input values. The red histogram is narrower than the black histogram since the resulting best fit should be similar to the input value that generated the light curve. The contribution of fitting errors to the overall scatter of observed values is estimated to be the ratio of the variances (σ^2) of the red and black histograms.

curve lengths shorter than τ. If we define $K = \tau\sqrt{\mathrm{SF}_\infty}$ as a variable orthogonal to $\hat{\sigma}$ (in log space), its uncertainty due to fitting errors contributes ∼82% of its scatter. Figure 3.7 compares the scatters in observed $\hat{\sigma}$ (top panel) and K (bottom panel) for S82 quasars to those for Monte Carlo models of the light curves.

Figure 3.8 shows the distributions of $\hat{\sigma}$ and K measured for the S82 sample compared to their expected values from Eq. 3.6. We find that the differences between the observed $\log\hat{\sigma}$ ($\log K$) and the estimates from Eq. 3.6 peak near zero with an rms of 0.16 (0.62) dex.

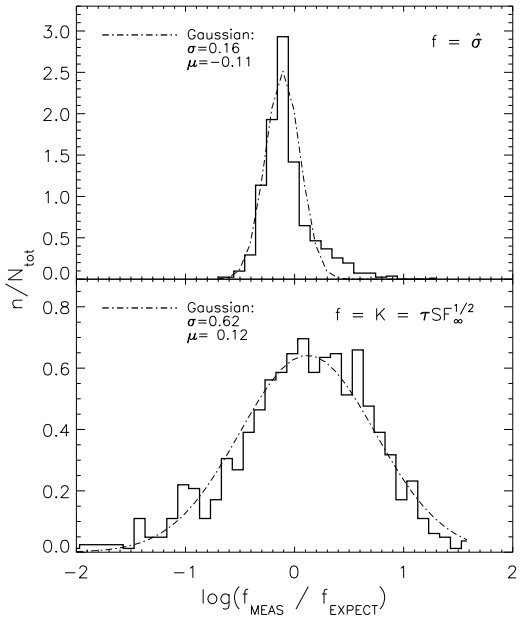

Figure 3.8 Distributions of $\hat{\sigma} = \mathrm{SF}_\infty/\sqrt{\tau}$ and $K = \tau\sqrt{\mathrm{SF}_\infty}$ for S82 quasar light curves with $r < 19$, normalized by the expected values from Eq. 3.6. Each dot-dashed curve shows a Gaussian with unit area, with a mean (μ) and rms (σ) as listed in the legends. After removing the contribution from fitting errors ($F = 13\%$ and 82% for $\hat{\sigma}$ and K, respectively), the intrinsic distribution widths are $\sigma_{int} = 0.149$ and 0.26 dex, where $\sigma_{int} = (\sigma^2 - F\sigma^2/100)^{1/2}$.

After taking the fitting errors into account, the rms is reduced to 0.149 (0.26) dex, which we take as the intrinsic stochasticity. Therefore, in the second step, we take the τ and SF_∞ estimated from Eq. 3.6, compute $\hat{\sigma}$ and K, add a random Gaussian deviate of width 0.149 (0.26) dex to each $\log\hat{\sigma}$ ($\log K$) value, and then convert back to τ and SF_∞. This process should provide a reasonable model for the intrinsic scatter.

In the third step, the SF for a particular quasar at a given Δt_{RF} is estimated using Eq. 3.5. In the fourth step, a model Δm value is drawn from a Gaussian distribution with a standard deviation of $\mathrm{SF}(\Delta t_{RF})_{qso}$, the Gaussian distribution expected from the DRW model. To account for photometric errors, we further add a random Gaussian deviate of width $\sqrt{2}\sigma_{phot}$ to each model Δm value (the $\sqrt{2}$ factor results from adding the photometric errors in quadrature). This procedure results in one model Δm value per quasar based on

the expectations from the DRW model. We then repeat this 1000 times for each quasar in the sample. The resulting model for the ensemble SF is then the rms width of this model Δm distribution otherwise calculated in the same way as for the data (using the interquartile range).

Explaining the Exponential Tails of the Δm Distribution

In Figure 3.9, the observed magnitude difference distribution is shown for a narrow slice in Δt_{RF} and λ_{RF}. The observed distribution is very similar to the model distribution (solid line), in which each quasar is assigned a model Δm value drawn from a particular $\mathrm{SF}(\Delta t_{RF})_{qso}$, as described above. The exponential distribution for large $|\Delta m|$ results from a superposition of many Gaussians (dashed lines) corresponding to different values of $\mathrm{SF}(\Delta t_{RF})_{qso}$. The range in $\mathrm{SF}(\Delta t_{RF})_{qso}$ is caused by differing τ and SF_∞ values, which can be attributed to a range in quasar luminosity and black hole mass plus intrinsic scatter.

Figure 3.10 shows the distribution of Δm in four bins of rest wavelength and three bins of rest-frame time lag. The distributions remain exponential at large $|\Delta m|$ for all 12 combinations of Δt_{RF} and λ_{RF}, as illustrated by the dashed lines. The model distributions, which carry information about each individual quasar's expected DRW parameters (and thus incorporate the M_i and M_{BH} information), are shown as solid curves. The observed and predicted $\mathrm{SF}(\Delta t_{RF})_{qso}$ values (σ) are listed to the right and left of each histogram, respectively. The model and data distributions agree well, showing that *the exponential distributions seen in the statistics of ensembles of quasars naturally result from averaging over quasars that are individually well-described by a Gaussian DRW process.* There are two systematic discrepancies, however. First, while the exponential tails can be reproduced at large rest-frame time lags ($500 < \Delta t_{RF} < 1500$ days), the value of the observed SF is systematically higher. Since there are no known intrinsic differences in the physical properties of the S82 quasars and the two-epoch sample studied here, the discrepancy is likely due to biases in estimates of the DRW variability parameters for the S82 quasars. Secondly, at small time lags ($50 < \Delta t_{RF} < 150$ days), the model over-predicts the data rms by 10%, while the statistical uncertainty is 1% assuming a perfect Gaussian distribution

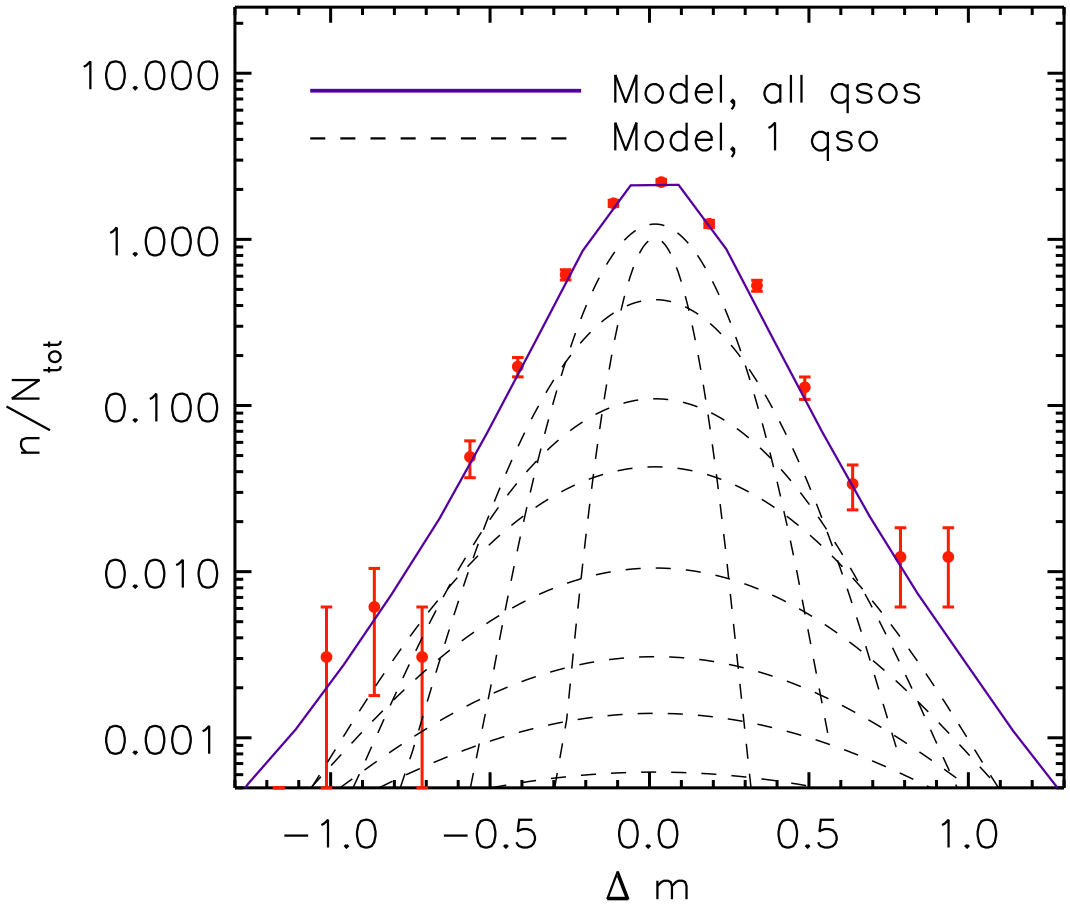

Figure 3.9 The red data points with errors bars show the observed magnitude difference distribution for $365 < \Delta t_{RF} < 730$ days and $2000\text{Å} < \lambda_{RF} < 3000\text{Å}$. The solid line shows the expected distribution based on the DRW model. The dashed Gaussian curves show the expected distributions for 9 individual quasars with different values of $\mathrm{SF}(\Delta t_{RF})_{qso}$ (and thus τ and SF_∞; see Eq. 3.5). Each curve has been convolved with a Gaussian noise component of width σ_{phot} to account for photometric errors, and shifted by the median of the observed distribution.

of Δm and no other systematic errors. This discrepancy may be due to some systematic effect that was not accounted for when computing the error bars (or when correcting for the SDSS photometric errors). Alternatively, the discrepancy could result if the DRW model is inaccurate on these timescales (1–200 days). These discrepancies are discussed further in

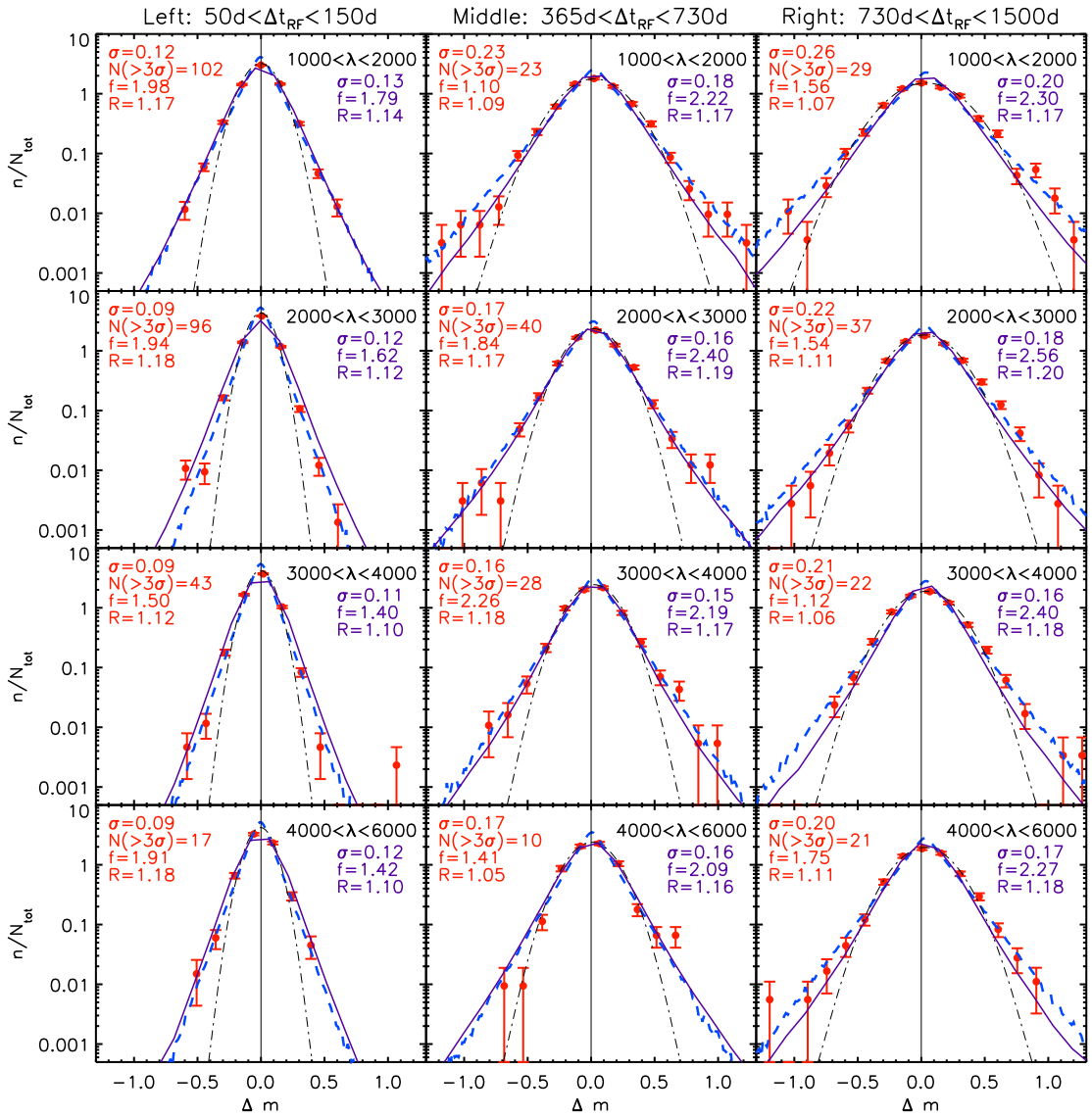

Figure 3.10 As in Fig. 3.3, except that the symbols show the distribution of measured magnitude differences, Δm, for subsamples selected using rest-frame quantities, as marked in the panels. Each row corresponds to the same wavelength range, and each column to the same Δt_{RF} range. The solid lines show the predicted DRW model distributions (see Section 3.4.2) with widths σ listed to the right of the histograms.

The SF as a Function of Time Lag and Wavelength

Now that we understand and can reproduce the shape of the distribution of magnitude changes Δm between two times, we can simply consider the SF (i.e., the width of this Δm distribution) as a function of physical quantities such as Δt_{RF} and λ_{RF} and test the model prediction. Figure 3.11 shows the SF as a function of Δt_{RF}. Here, the measured SF is not corrected for the SDSS photometric accuracy. Again, it is apparent that the model predicts a systematically lower SF for time lags less than 200 days. However, the stochastic model predicts a systematically higher SF for the longest time lags. This suggests the model SF may be biased low at long time lags, and the bias seems to be most prominent at shorter wavelengths (Figure 3.12). Note that the measurements at long rest-frame time lags and short rest wavelengths are dominated by the u band, so there may be biases simply from the fact that the 14% of u and z light curves which were dominated by noise were omitted from the analysis in Mac10. We investigate such biases further in Section 3.5.1.

In Figure 3.13, we show the observed ensemble SF without correction for the SDSS photometric errors. We also show the signal-to-noise ratio for each bin (top-right panel). In the bottom-left panel, we show the model expectation including the estimated photometric errors. The bottom-right panel shows the (data – model) residuals. Overall, the results are excellent, with a median difference of only 0.02 mag and a median absolute difference of 0.03 mag. We see a slight decrement at $\Delta t < 200$ days in the residuals at wavelengths coincident with the 2800Å MgII emission line. This decrement is expected based on the results of Reichert et al. (1994), where the MgII emission line is less variable and lags the continuum fluctuations by almost 10 days. In Figure 3.14, the time axis is collapsed onto the wavelength axis (using the weighted mean) so that the dip in the residuals is emphasized. Due to the excellent SDSS photometry and large sample size, we are able to resolve this feature. Note that the large width of the decrement may be due to the fact that the rest wavelengths are approximated from the fixed effective wavelengths of the SDSS bands. Not accounting for the variation of the effective wavelength with the shape and redshift of the quasar spectrum sampled limits the accuracy of the approximated rest wavelengths. Thus, any line feature will be weakened and smoothed by the broad bandpasses of the SDSS filters.

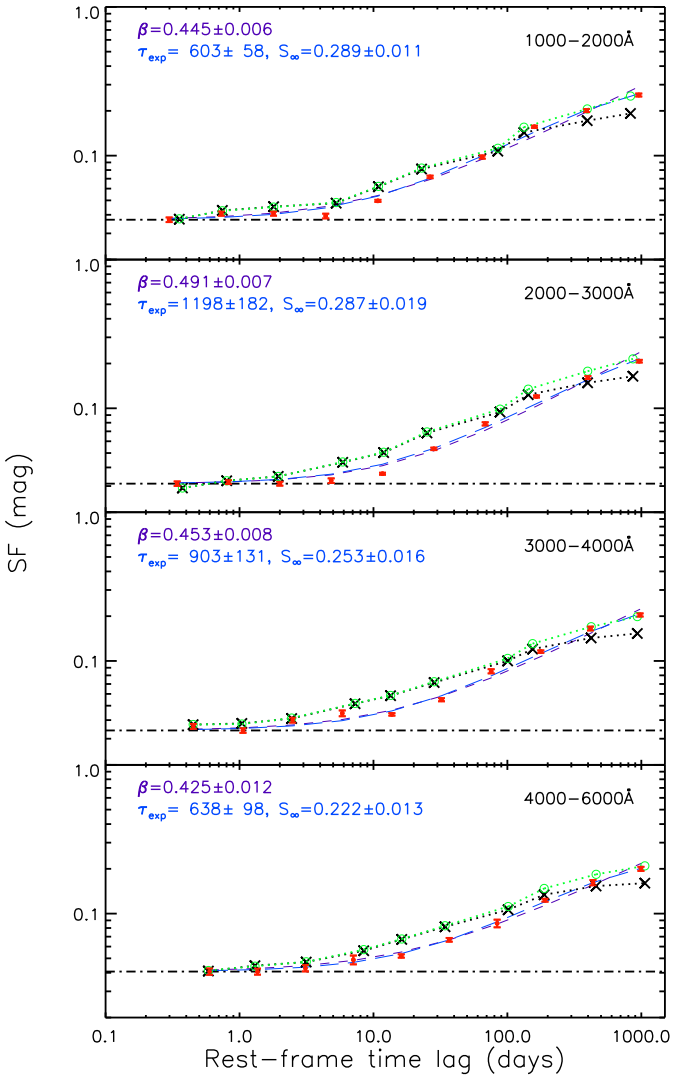

Figure 3.11 The red symbols show the structure function for quasar variability as a function of rest-frame time lag, for four ranges of rest-frame wavelength, as marked. The short-dashed line in each panel is the best-fit power law (SF $\propto \Delta t^{\beta}$), with β marked in each panel. The long-dashed lines are the best-fit exponential curves, SF $= S_{\infty}[1 - \exp(-\Delta t/\tau_{\rm SF})]^{1/2}$, with the parameters listed in each panel. Both curves include a photometric noise component equal to the minimum measured SF in each panel (marked by the horizontal dot-dashed line). The dotted lines with cross symbols show the DRW model prediction (for 100 realizations, see Section 3.4.2). The green dotted lines with circles show the prediction of the DRW model trained on S82 data with all τ and SF$_{\infty}$ values multiplied by 2 and $\sqrt{2}$, respectively.

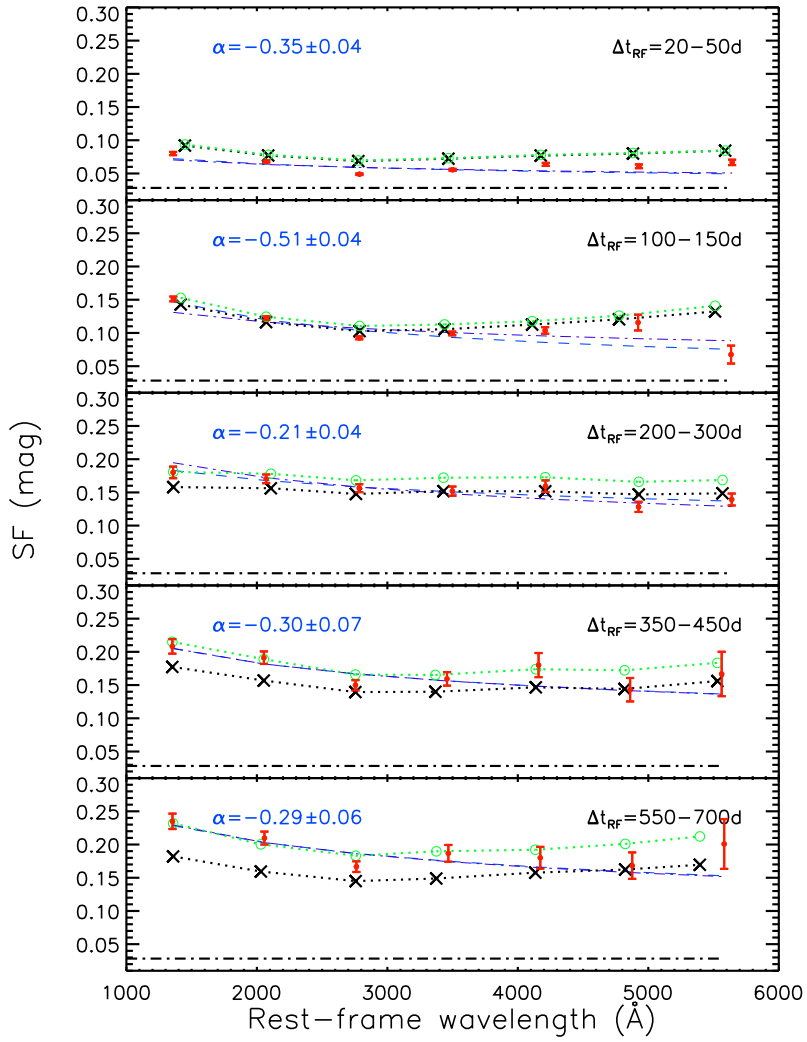

Figure 3.12 The red symbols show the structure function for quasar variability as a function of rest-frame wavelength, for five ranges of rest-frame time lag, as marked. The dashed lines are the best-fit power laws (SF $\propto \lambda^{\alpha}$) with the power law index listed in the top-left corners. The dash-dotted lines show the functional dependence adopted by I04 ($\alpha = -0.3$). Both curves include a photometric noise component corresponding to $\sigma_{phot} = 0.02$ mag. The dotted lines with cross symbols show the DRW model prediction (for 100 realizations, see Section 3.4.2). The green dotted lines with circles show the prediction of the DRW model trained on S82 data with all τ and SF_{∞} values multiplied by 2 and $\sqrt{2}$, respectively.

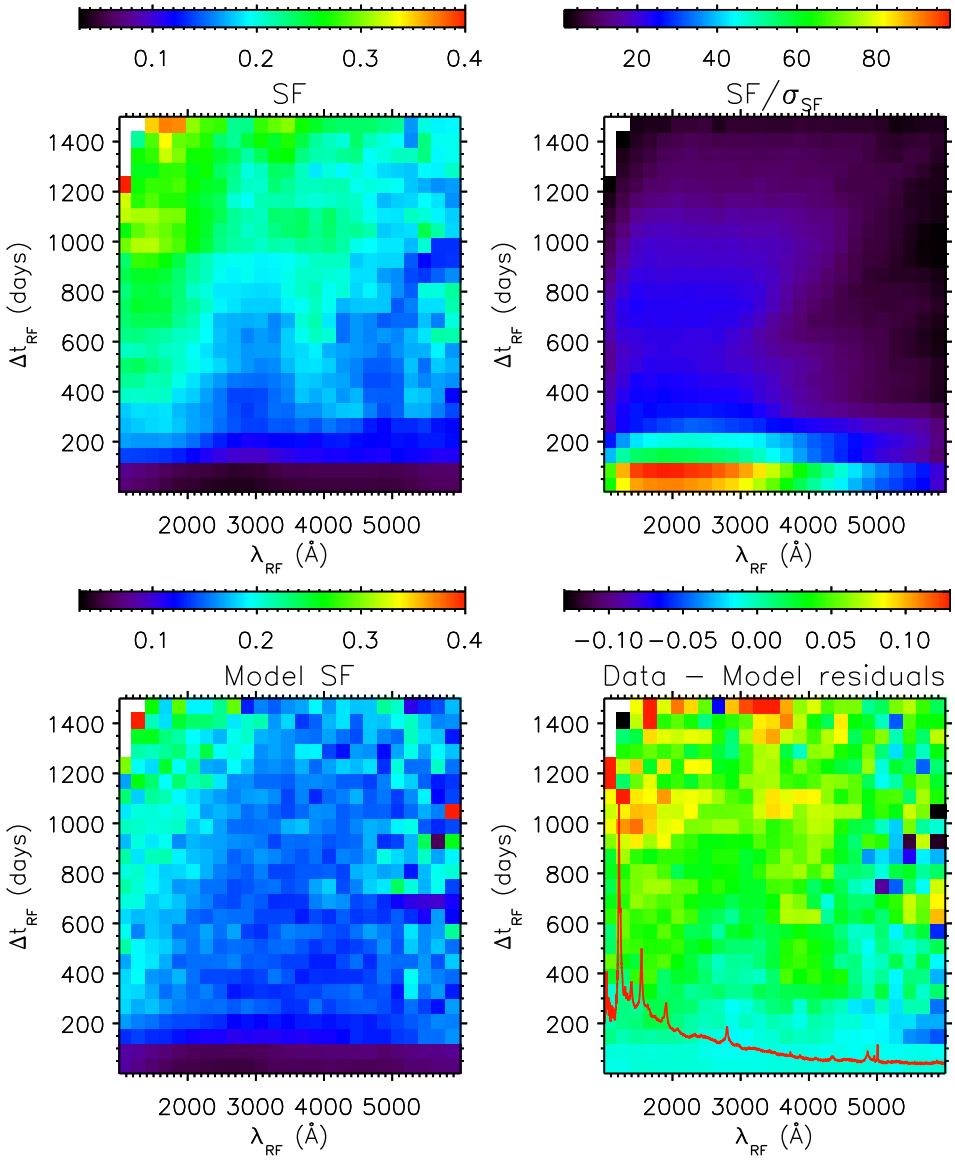

Figure 3.13 The top-left panel displays the structure function for quasar variability measured by the SDSS as a function of rest-frame wavelength and time lag. SF is computed as $0.74(\mathrm{IQR})/\sqrt{N-1}$, where IQR is the 25% – 75% interquartile range. The figure is based on 79,787 SDSS measurements, and shows the structure function on a linear scale from 0.03 to 0.4 mag, as indicated in the top bar. The top-right panel shows the SF signal-to-noise, where the error σ_{SF} is computed as $1.15SF/\sqrt{N-1}$. The bottom-left panel shows the model structure function (see text), and the bottom-right panel shows the (data–model) residuals. The red line shows the SDSS composite quasar spectrum in arbitrary flux (F_λ) units from Vanden Berk et al. (2001). The weak vertical feature in the residual map at ∼2800Å is coincident with the MgII line visible in the composite spectrum. The white pixels in each panel contain fewer than 5 data points even when including all adjacent bins (and thus are not used in our analysis).

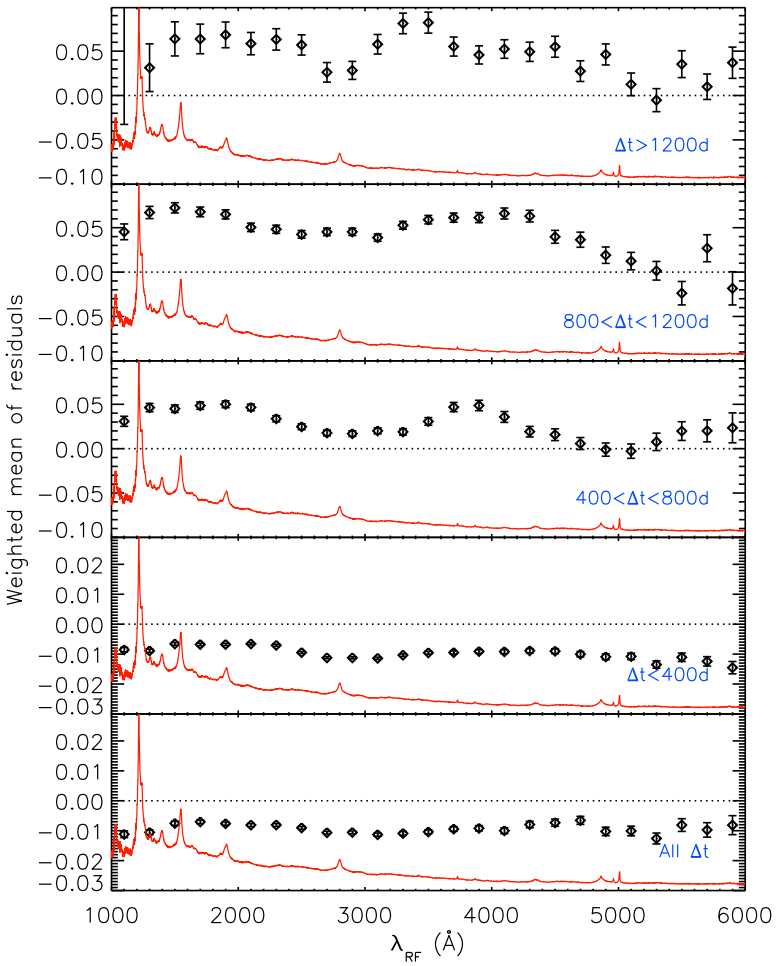

Figure 3.14 Weighted mean of the data–model residuals as a functions of rest wavelength (the bottom-right panel of Fig. 3.13 collapsed onto the wavelength axis). The weighted mean of all time lag bins i is defined as $\sum(SF_i - SF_{model,i})w_i / \sum w_i$, where $w_i = 1/\sigma^2_{SF}$, the uncertainty from Fig. 3.13. The error bars of the weighted means are computed as $(\sum w_i)^{-1/2}$. The composite quasar spectrum is over-plotted in arbitrary flux (F_λ) units. The dip at 2800Å results from the lagged and smeared variability of the MgII emission line with respect to the continuum, and is widened by the errors in estimating the rest wavelength from the SDSS bands.

3.5 Combining Short- and Long-term Quasar Variability Measurements (SDSS–POSS)

The exponential distribution analysis in Section 3.4.2 is sensitive to both SF_∞ and τ through the better determined combination defined by $\hat{\sigma} = SF_\infty/\sqrt{\tau}$. However, for the long-term SDSS–POSS data, the Δm distribution is mainly sensitive to SF_∞ or τ at fixed $\hat{\sigma}$. It is then of interest to see whether the Δm distribution remains exponential at large $|\Delta m|$ and if we can reproduce the ensemble $SF(\Delta t)$ on these long timescales. Therefore, we repeat the analysis in Section 3.4.2 using the 81,189 SDSS quasars that are also observed in DPOSS. We only use the DPOSS data for the Δm analysis because the errors for POSS-I are significantly larger. After applying the data quality cuts, there are 56,732 total Δm measurements in the G, R, and I bands. Figure 3.15 shows the distribution of time lags in the observed and rest frames.

The SDSS–DPOSS Δm distributions are shown in Figure 3.16. Due to the larger photometric errors, these distributions appear more Gaussian than those restricted to SDSS observations. In some panels, the peak is also offset from zero toward positive Δm values. This is likely due to the Malmquist-like bias discussed in dV05, where many objects that were fainter at the time of the DPOSS observation were then lost and not included. In general, the model curves (solid lines) are able to reproduce the shape of the distributions but underestimate the rms, even when accounting for the estimated photometric errors of $\sigma_{phot} = 0.1$ mag.

Finally, Figure 3.17 shows the SF over a large range of rest-frame time lag, including the SDSS–POSS data as measured in Ses06. We find an overall characteristic timescale of 835 ± 82 days with $SF_\infty = 0.26\pm0.01$ mag. Note that in dV05, the SF reaches 0.46 ± 0.02 mag at 40 years in the rest frame for the g and r bands combined.

3.5.1 Data–Model Comparison

In Figure 3.17, we show our model SF as described in Section 3.4.2. We again see that the model is biased to higher values of the SF between 1 and 200 days. We attribute this discrepancy to one or both of the following.

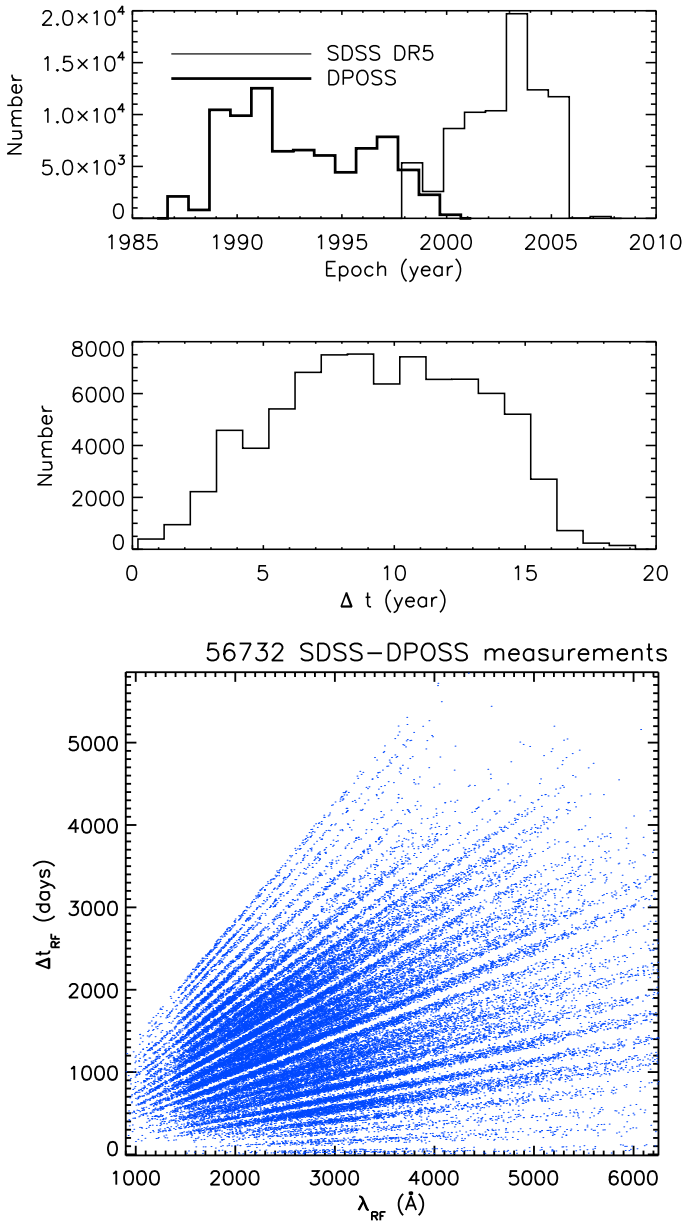

Figure 3.15 Top: Distribution of epochs for the SDSS and DPOSS observations. Middle: Distribution of the SDSS–DPOSS time lags in the observer's frame. Bottom: Distribution of the SDSS–DPOSS time lags and wavelengths in the rest frame.

1. At short Δt, the errors in our measured SF are underestimated, and some source of systematic error, such as an inaccurate correction for SDSS photometric errors, causes

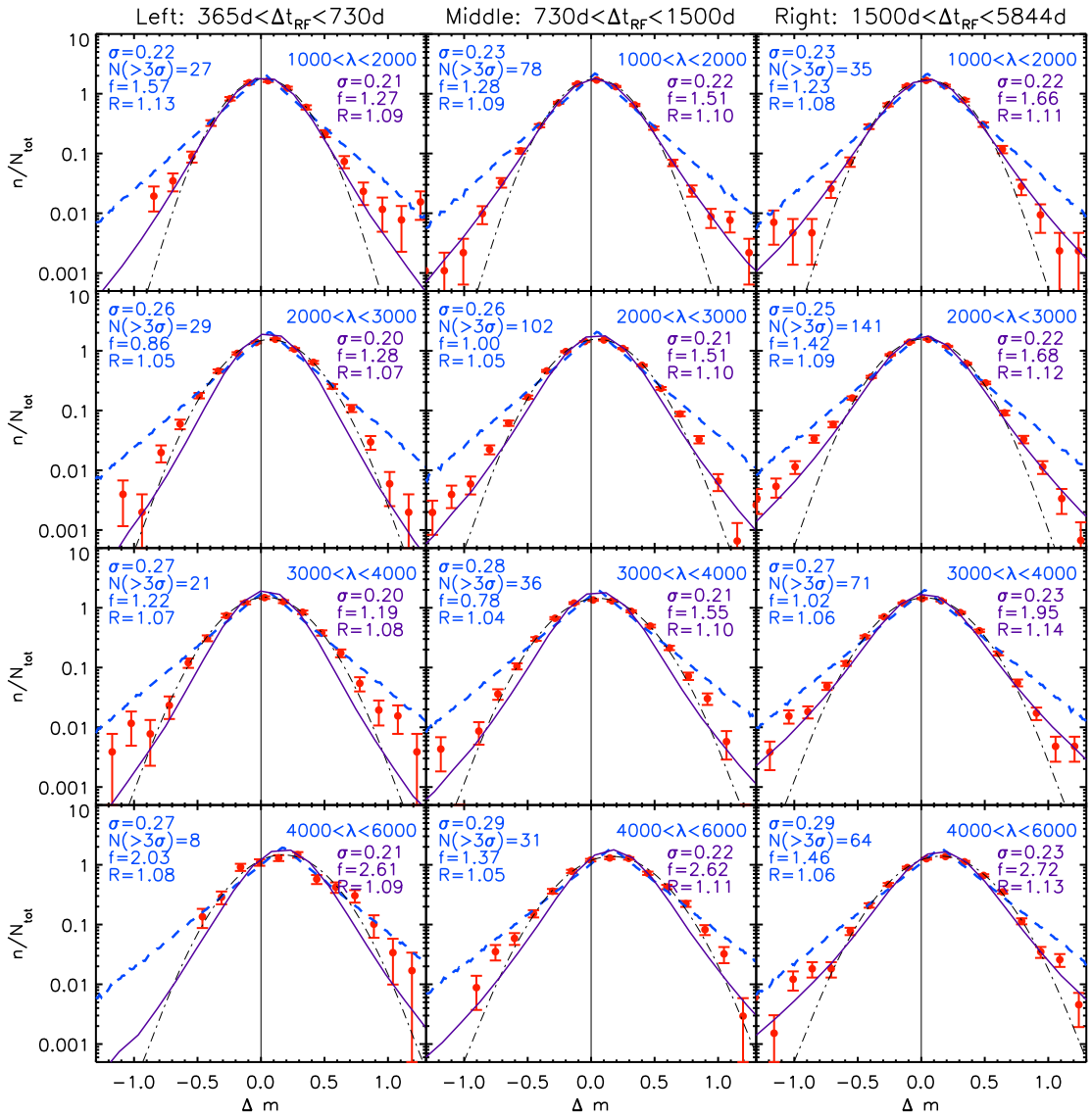

Figure 3.16 Similar to Fig. 3.10 but for SDSS–DPOSS observations over the northern DR5 footprint (with $i < 19.1$ in SDSS). The photometric accuracy was assumed to be $\sigma_{phot} = 0.1$ mag when computing the models. The Malmquist-like bias seen in the bottom panels is taken into account by shifting the model Δm distribution to the right by the observed mean.

the measured SF to appear smaller than what the model predicts.

2. The DRW model is inaccurate at short Δt and therefore overpredicts the ensemble

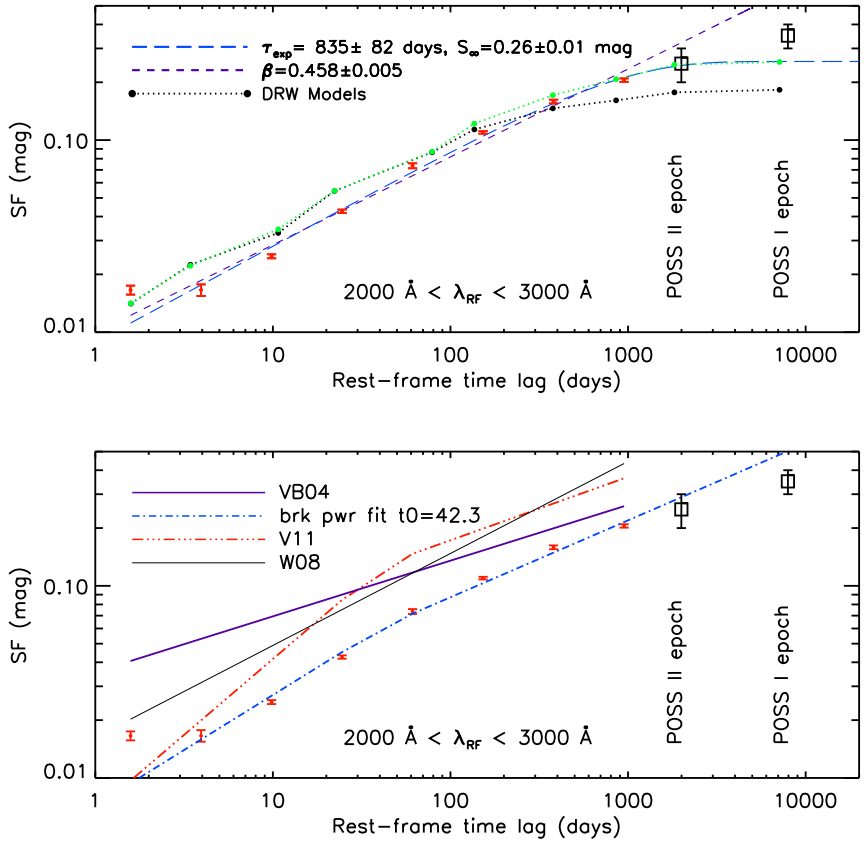

Figure 3.17 Top: Similar to Fig. 3.5, but plotted against rest-frame time lag, for data with rest-frame wavelength in the range 2000–3000Å. The data have been corrected for photometric errors by subtracting $0.018\sqrt{2}$ in quadrature. The green dotted line shows the prediction of the DRW model trained on S82 data with all τ and SF_∞ values multiplied by 2 and $\sqrt{2}$, respectively. Bottom: The data are shown along with comparison g-band fits from VB04 (purple line), Wilhite et al. (2008; thin line) and Voevodkin (2011; red dash-triple-dotted line). The normalization for these curves is arbitrary, since we are simply comparing the slopes. The blue dash-dotted line shows the best-fit broken power law to our data with the break set to 42.3 days, as in Voevodkin (2011).

SF. However, this is unlikely given that the results using S82 light curves as well as densely-sampled OGLE light curves (Zu et al. 2012) show the DRW model to be a good fit on timescales between ∼5 and 200 days.

Nevertheless, our measured slope of the SF seems well constrained. We again see that beyond 200 days, the model is biased to lower values of the SF. The difficulty to exactly

reproduce the long-term variability amplitudes suggests that the distributions of long-term amplitudes and/or timescales used in our model underestimate the true distributions. This bias may result from the following:

1. A bias in the measured τ (and thus SF_∞) values for S82 quasars due to insufficient light curve lengths. Since $SF(\Delta t << \tau) = \hat{\sigma}\sqrt{|\Delta t|}$, $\hat{\sigma} = \mathrm{SF}_\infty/\sqrt{\tau}$ is the most strongly constrained quantity on timescales shorter than τ. That is, on a grid of $\log(\tau)$ versus $\log(\mathrm{SF}_\infty)$, the best-fit values will be scattered due to fitting errors along lines of constant $\hat{\sigma}$, but much less perpendicular to it (i.e., along lines of constant $K = \tau\sqrt{\mathrm{SF}_\infty}$). Therefore, as the light curve length decreases, the mean best-fit $\hat{\sigma}$ will not vary significantly, while the best-fit parameters τ, SF_∞, and K will become biased (see Kozł10; Mac10; MacLeod et al. 2011b).

2. A bias in the deterministic model (Eq. 3.6). Due to the finite length of the S82 survey, 22% of the S82 sample was excluded from the analysis in Mac10 due to indeterminately long timescales. In this case, the measured parameters τ and SF_∞ are accurate for the remaining 78% of the S82 sample, but Eq. 3.6 is only accurate for the lower values of τ ($\tau \lesssim 10^3$) and SF_∞ from which the correlations are derived.

In principle, both effects will lead to lower overall τ and SF_∞ in the distributions for S82 quasars, which will in turn cause the model ensemble SF to flatten at $\Delta t_{RF} < 700$ days as seen here. Indeed, the excluded 22% of the sample will mostly contribute power to the long-term rather than the short-term SF given their indeterminately long timescales. However, when simply including an additional population of long-τ objects to fill in the missing 22%, we are unable to quantitatively reproduce the last SDSS data point at $\Delta t_{RF} = 1000$ days unless all the additional objects have $\tau \gtrsim 10,000$ days in the rest frame (while keeping the same distribution of $\hat{\sigma}$ as observed). On the other hand, if we double all the τ values in our model (instead of adding an extra population of quasars at long τ), while keeping $\hat{\sigma}$ fixed so that the SF_∞ values are multiplied by $\sqrt{2}$, we are able to reproduce the observed long-term SF, as shown by the agreement between the green lines in Figures 3.17 and 3.11 to the data points at long Δt_{RF}. In summary, the best-fit A coefficients from Eq. 3.6 need

to be altered upwards by 0.15 dex in the case of SF_∞, and by 0.30 dex in the case of τ, in order to explain the long timescale constraints provided by the SDSS–POSS data set.

At least half of these corrections can be understood as due to a fitting bias toward shorter τ (see Figure 7 in Mac10 and Figure 15 in MacLeod et al. 2011b). Another effect could be due to uncertain behavior for long timescales. The DRW process corresponds to a power spectral distribution (PSD) proportional to $1/f^2$ at frequencies $f > (2\pi\tau)^{-1}$, flattening to a constant at lower frequencies. Using the S82 data and computational technique described in Mac10, we were able to rule out an extrapolation of the $1/f^2$ power law. However, we were unable to distinguish between a $1/f^0$ or a $1/f$ PSD at frequencies $f < (2\pi\tau)^{-1}$, where the latter dependence is observed in the X-ray PSDs for Galactic black holes as well as AGN (McHardy et al. 2006; Kelly et al. 2011).

We compare the observed SF slopes from Voevodkin (2011), VB04, and Wilhite et al. (2008) to our data in the bottom panel of Figure 3.17. Voevodkin (2011) found that a broken power law provides a good fit to the S82 g-band ensemble SF with a slope of 0.33, steepening to 0.79 below 42 days. We also show the best-fit broken power law to our data with the break fixed to 42 days, and we find power law indices at short and long Δt of 0.53 and 0.40, respectively. While our two-epoch SDSS data are consistent with the shallower slope of 0.33, our results do not support the conclusion of a much steeper $SF(\Delta t)$ for small Δt found by Voevodkin (2011) for either the S82 or two-epoch data sets. We interpret the broken power law form to be a consequence of the turnover in the SF due to the mean characteristic timescale, as the observed SF is fully consistent with the form expected for a DRW (Eq. 3.5).

3.6 Predictions for Future Surveys

Using the observed Δm distributions, we can predict the number of quasars with Δm exceeding an arbitrary limit that might be seen in a survey with a given number of quasars. This information is useful for transient identification, in particular to identify quasars that contaminate candidate lists of other objects. For example, Vanden Berk et al. (2002) reported an orphan gamma-ray burst afterglow based on a 2.5 magnitude decrease in the optical flux of an unidentified point source. Instead, as pointed out by Gal-Yam et al.

(2002), the observations are best explained by the presence of a spectroscopically identified, highly variable quasar.

In order to quantify the importance of quasar contamination in future transient surveys, such as the Palomar Transient Factory (PTF; Law et al. 2009) and the Large Synoptic Survey Telescope (LSST; Ivezić et al. 2008), we need to know the probability that a quasar can increase brightness by Δm (over the faint survey limit) within a time Δt. For example, assume that no source is detected above a faint limit of m_{faint} on a given night, but when repeating the observation some time later, a source is detected with magnitude $m = m_{faint} - \Delta m$. In this case we would like to know how many quasars with a Δm at least as large as observed could be present in a particular scanned area and to a faint limit of m_{faint}. Given that our stochastic model performs well at reproducing the observed ensemble variability of quasars, we can make robust and useful predictions for future surveys.

We make use of a mock LSST quasar sample which includes absolute B magnitudes (M_B) and redshifts generated over 100 deg^2 of sky using the luminosity function from Bongiorno et al. (2007) for the purposes of LSST image simulations. We limit the sample to $M_B < -20$ following Table 10.2 in the LSST Science Book (LSST Science Collaborations and LSST Project 2009). The distance modulus is computed assuming a standard cosmology: $\Omega_m = 0.27$, $\Omega_\Lambda = 0.73$, and $h = 0.71$. The LSST magnitudes (m) and rest-frame M_i (based on M_B) are computed using the composite quasar spectrum from Vanden Berk et al. (2001). The black hole masses are estimated from the M_i values using the prescription in Mac10.

We generate magnitude differences for each quasar in the mock LSST sample based on the DRW model most appropriate for the quasar's physical parameters, as described in Section 3.4.2. When computing the DRW model, the τ (SF$_\infty$) values are multiplied by 2 ($\sqrt{2}$) with respect to the expected values based on Mac10 in order to correct for the bias due to limited time sampling in S82 (see the previous section). We note that correcting for this bias should also account for at least some of the residual scatter in K estimated in Section 3.4.2; however, for simplicity we still include this scatter along with the bias correction in our simulations. First, we consider three different survey faint limits of $m_{faint} = 19.1$ (to establish similarity with the SDSS results), $m_{faint} = 22$, and $m_{faint} = 24.5$, excluding all mock quasars with $m \geq m_{faint}$ in each case. For the $m < 19.1$,

22, and 24.5 simulations respectively, 1000, 100 and 11 model Δm values are generated per quasar to increase the sample size. A Gaussian noise component of width $\sqrt{2}\sigma_{phot}$ is added to all curves to simulate a photometric accuracy similar to the SDSS (the photometric accuracy for future surveys such as LSST will likely be better than that for the SDSS, but this is a higher order question than investigated here). Note that we retain all Δm values in the simulations, including $|\Delta m| > 3$.

Figure 3.18 shows the simulated cumulative distribution of Δm in the urz bands for three faint magnitude limits. Also shown are Gaussian analytic functions as thin curves with the same rms as the data. Using these cumulative distributions, which are based on $\sim 10^6$ mock quasars, we predict probabilities down to our resolution limit ($P \gtrsim 10^{-6}$). Table 3.1 lists the predicted probabilities of observing a quasar with $m < 24.5$ and a magnitude difference of $\Delta m > 1$ mag and $\Delta m > 2$ mag over 3, 30, and 300 days, using the observed frame, in the urz bands. Due to the exponential nature of the Δm distributions, the probability of observing $\Delta m > 1$ mag reaches 0.02 in the u band (where variability is strongest) for time lags of 300 days, and 6×10^{-4} for $\Delta m > 2$ mag. Assuming Gaussian Δm distributions will result in erroneous probability estimates of 9×10^{-4} and 10^{-6}, respectively.

Next, we adopt realistic PTF photometric errors which vary as a function of magnitude, and repeat the simulation in g and r using the PTF magnitude limits ($r < 20.6$ and $g < 21.0$). The errors range from 0.008 mag at the brightest limits to 0.2 mag at the faint end. As shown in Figure 3.19, the resulting distributions are more Gaussian than those in the previous simulations due to the different dependence of errors on magnitude. However, at extreme values of Δm, the Gaussian curves significantly underestimate the detection probabilities in both g and r.

These results would have been useful to Vanden Berk et al. (2002), who reported a single transient with $\Delta m = 2.5$ mag over 410 days in the gri bands in early SDSS data covering 1500 deg^2. Since the object had the spectrum of a normal galaxy in its faint phase (rather than a nonstellar spectrum with broad emission lines indicative of quasars), the authors concluded that it was more likely to be a gamma-ray burst afterglow than a highly variable quasar, given the large drop in flux. We find that the probability of a quasar having $\Delta m = 2.5$ mag over 410 days with $m_{faint} = 22$ is 10^{-5} in the g and r bands. To find the

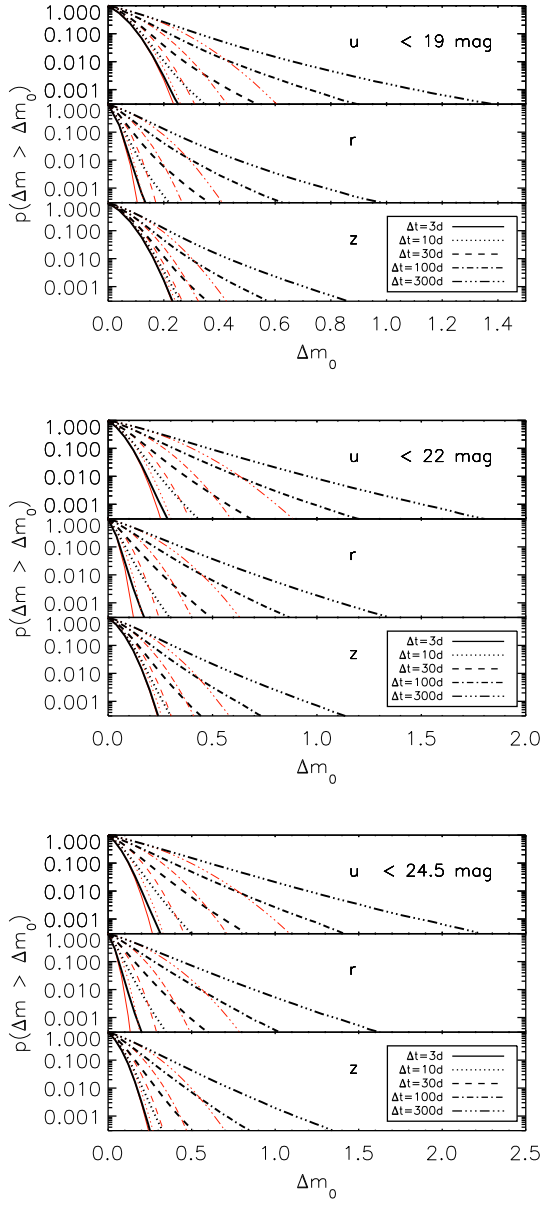

Figure 3.18 The predicted cumulative distribution of magnitude differences in u, r, and z bands as a function of observed time lag (Δt) for $\sim$1 million quasars with magnitudes less than 19.1 (top three panels), 22 (middle three panels), and 24.5 (bottom three panels). The thin curves show the Gaussian distributions with the same rms widths.

number of quasars expected to exhibit this variability in 1500 deg^2 of sky, one needs to know the quasar density. We extrapolate Figure 13 in Richards et al. (2006) to a density

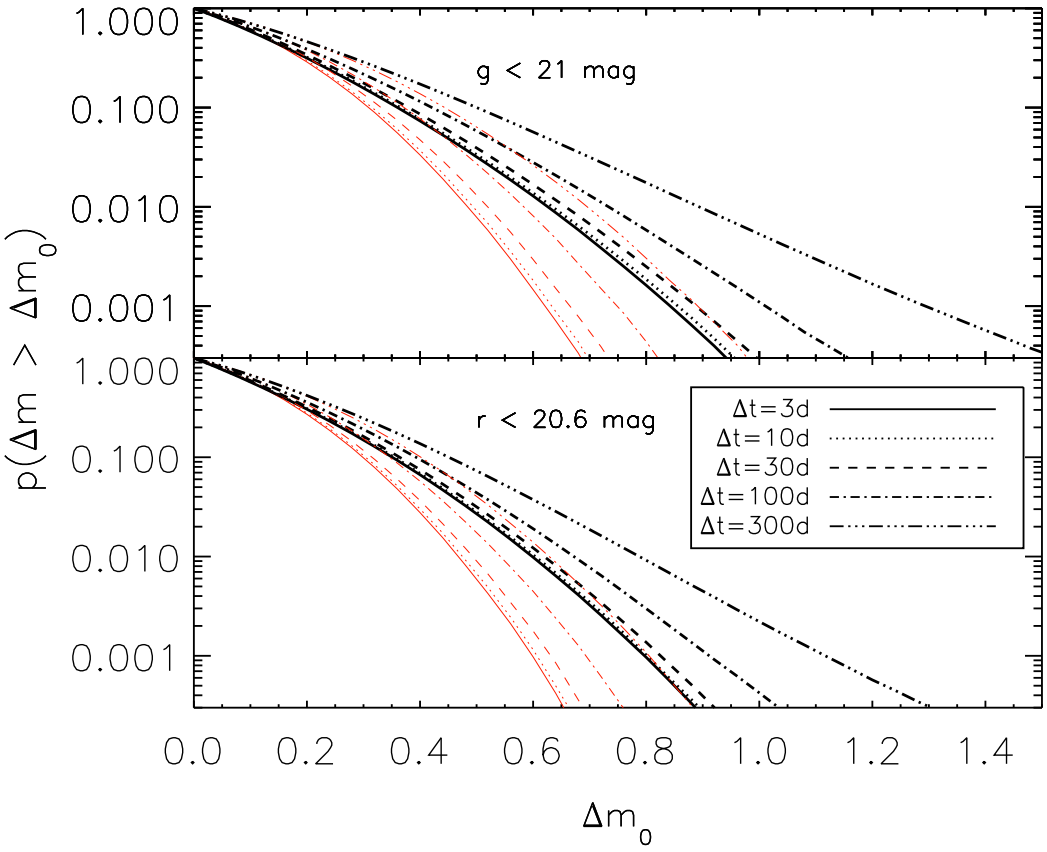

Figure 3.19 The predicted cumulative distribution of g-band (top) and r-band (bottom) magnitude differences, including realistic PTF photometric errors, as a function of observed time lag (Δt) for quasars brighter than the PTF magnitude limit ($g < 21.0$, $r < 20.6$). The thin curves show the Gaussian distributions with the same rms widths.

of $\sim$120 deg^{-2} at $i < 22$. Therefore, we expect roughly 10^5 quasars in 1500 deg^2 to $i < 22$. Given a probability of 10^{-5}, we expect that roughly one quasar will show $\Delta r = 2.5$ mag over $\Delta t = 410$ days in the SDSS sample studied by Vanden Berk et al. (2002). This is consistent with their finding, and with the follow-up study of Gal-Yam et al. (2002) who confirmed that it is indeed a highly variable quasar. If one assumes a Gaussian distribution of Δm, the probability is less than 10^{-6}, and one would expect $\lesssim 0.1$ quasar to be found with these parameters. Given this expectation value, the probability to detect one quasar is at most 0.09, and the quasar hypothesis would be (erroneously) rejected at a $\gtrsim 2\sigma$ level.

Table 3.1. Predicted Δm Probabilities for $m < 24.5$

Δt (days)	Band	$P(\Delta m > 1$ mag)	$P(\Delta m > 2$ mag)
3	u	$< 2 \times 10^{-6}$	$< 2 \times 10^{-6}$
3	r	$< 2 \times 10^{-6}$	$< 2 \times 10^{-6}$
3	z	$< 2 \times 10^{-6}$	$< 2 \times 10^{-6}$
30	u	7×10^{-5}	$< 2 \times 10^{-6}$
30	r	4×10^{-6}	$< 2 \times 10^{-6}$
30	z	$< 2 \times 10^{-6}$	$< 2 \times 10^{-6}$
300	u	0.02	6×10^{-4}
300	r	0.005	6×10^{-5}
300	z	0.002	1×10^{-5}

3.7 Discussion

We have assembled, organized, and publicly released a data set including ∼3.5 million photometric measurements for 80,000 spectroscopically confirmed quasars. The available time lags span 0.8 days to almost 20 years in the observer's frame. We have analyzed and quantified the observed variability in the observer's and rest frames. By assuming a DRW model for each quasar in our sample, we reconcile the observed variability of individual quasars in S82 with their ensemble statistics. Our principal results are as follows.

1. Long-term quasar variability measurements, constrained using SDSS and POSS data for time lags up to 50 years (in the observer's frame), conclusively show that a simple power law dependence for the SF cannot be extrapolated beyond a decade, and suggests an average characteristic timescale for quasar variability in the rest frame of ∼2 years and an average long-term dispersion of ∼0.26 mag (for rest wavelengths 2000–3000Å). This behavior extrapolates well to the UV results of Welsh, Wheatley

& Neil (2011), who find that the SF for *GALEX* NUV data reaches about 0.4 mag and flattens at $\Delta t_{RF} > 300$ days. This SF limit corresponds to a limiting SF value of 0.33 mag when using our definition of the SF (see Section 3.4.1), and 0.27 mag when also scaling to the u band using the wavelength dependence from Eq. 3.6. This result is in close agreement with the SF at the shortest wavelengths in our data set. Voevodkin (2011) found that a broken power law provides a good fit to the S82 ensemble SF with a slope of 0.33 at long timescales steepening to 0.79 below 42 days. Our two-epoch SDSS data are consistent with the shallower slope of 0.33, but our data do not support the conclusion of a much steeper SF(Δt) for small Δt found by Voevodkin (2011). While we cannot rule out a broken power law dependence with the available data, the observed SF is fully consistent with the form expected for a DRW (Eq. 3.5).

2. We tested the DRW model results based on SDSS Stripe 82 data on an independent data set, and confirm that the variability parameters τ and SF_∞ correlate with physical parameters as found for individual quasars (e.g., Mac10 and references therein). This is evident from the agreement of our model with the observed ensemble variability of SDSS quasars. However, the results indicate that the measured τ and SF_∞ distributions are biased low for the S82 sample by a factor of about 2 and $\sqrt{2}$, respectively. This bias most likely results from the 10-year length limit of the S82 light curves, although it could also be due to uncertain behavior for long timescales. The best-fit A coefficients from Eq. 3.6 (Mac10) need to be shifted upwards by 0.15 dex in the case of SF_∞, and by 0.30 dex in the case of τ, in order to explain the long timescale constraints provided by the SDSS–POSS data set. These shifts leave the shorter timescale variability statistics unchanged.

3. For a given time lag and wavelength, the magnitude difference (Δm) distribution is exponential rather than Gaussian for large magnitude changes. This is well explained as a cumulative effect of averaging over quasars with a range of different τ and SF_∞. This is a remarkable result given that the Δm distribution of every individual quasar

is Gaussian.

4. We made predictions for the incidence of quasar contamination in transient surveys using detailed simulations of quasar light curves from a mock LSST catalog. Due to the exponential nature of the Δm distributions for quasars, the probability of observing $\Delta m > 1$ mag reaches 0.02 in the u band (where variability is strongest) for time lags of 300 days, and 6×10^{-4} for $\Delta m > 2$ mag. Assuming Gaussian Δm distributions will result in erroneous likelihood estimates that are about 10 and 1000 times smaller, respectively.

It is clear that a major limitation for the S82 quasars is the quality of light curves in both sampling density and time span. It is also clear that our variability model needs to be better tested given the evidence for a likely bias in the S82 timescale estimates. The best current sample for these improvements is that from the OGLE microlensing survey, since the light curves are more densely-sampled and longer than for S82. Here, the problem is the lack of spectroscopic identification of quasar candidates, although the follow-up confirmation of quasars is rapidly improving (Kozłowski et al. 2012). The next-generation surveys will also greatly improve the constraints on the long-term SF both individually and for ensembles of quasars. The best short-term prospects are PTF (Law et al. 2009), Pan-STARRS (Kaiser et al. 2002), and the Dark Energy Survey (DES; Honscheid et al. 2008). In particular, the DES supernova program will greatly expand many of the S82 quasar light curves with *griz* sampling once per week for $\sim$3 months per year over 5 years. The combination of SDSS, Pan-STARRS, DES, and LSST will yield well-sampled light curves covering over 25 years for 10,000 quasars in Stripe 82.

The success of the model presented here suggests that a range of characteristic timescales exists among an ensemble of quasars, which can be related to physical timescales in the accretion disk. While we assumed a single τ per quasar, there is evidence that multiple timescales can exist for a given quasar (Collier & Peterson 2001; Kelly et al. 2011). Therefore, the study presented here can be extended to adopt the model in Kelly et al. (2011), which fits more than one τ for a given object. This model, also called a mixed Ornstein-Uhlenbeck (OU) process, reproduces a PSD of the form exhibited by the X-ray light curves of galactic

black holes and AGN, which is flat below a low-frequency break, decays as $1/f$ above the low-frequency break, and steepens to $1/f^2$ above a high-frequency break. In this case, with two characteristic timescales for each quasar, the long-term ensemble SF can be revisited and possibly explained in the context of a mixed OU process. Note that recent optical data from the Kepler mission (Mushotzky et al. 2011), which have a sampling of 1 data point roughly every 30 minutes and 0.1% errors, suggest an additional break to a steeper slope ($\sim 1/f^3$), but this dependence is seen on timescales shorter than can be resolved in SDSS data.

Chapter 4

QUASAR SELECTION BASED ON PHOTOMETRIC VARIABILITY

4.1 Abstract

Up until Section 4.8, this Chapter contains the text of the published paper, MacLeod et al. (2011b), where we develop a method for separating quasars from other variable point sources using SDSS Stripe 82 light curve data for ∼10,000 variable objects. To statistically describe quasar variability, we use a damped random walk model parametrized by a damping timescale, τ, and an asymptotic amplitude (structure function), SF_∞. With the aid of an SDSS spectroscopically confirmed quasar sample, we demonstrate that variability selection in typical extragalactic fields with low stellar density can deliver complete samples with reasonable purity (or efficiency, E). Compared to a selection method based solely on the slope of the structure function, the inclusion of the τ information boosts E from 60% to 75% while maintaining a highly complete sample (98%) even in the absence of color information. For a completeness of $C = 90\%$, E is boosted from 80% to 85%. Conversely, C improves from 90% to 97% while maintaining $E = 80\%$ when imposing a lower limit on τ. With the aid of color selection, the purity can be further boosted to 96%, with $C = 93\%$. Hence, selection methods based on variability will play an important role in the selection of quasars with data provided by upcoming large sky surveys, such as Pan-STARRS and the Large Synoptic Survey Telescope (LSST). For a typical (simulated) LSST cadence over 10 years and a photometric accuracy of 0.03 mag (achieved at $i \approx 22$), C is expected to be 88% for a simple sample selection criterion of $\tau > 100$ days. In summary, given an adequate survey cadence, photometric variability provides an even better method than color selection for separating quasars from stars.

Section 4.8 describes our spectroscopic follow-up program using the Dual Imaging Spectrograph on the ARC 3.5m at Apache Point Observatory to confirm our top quasar candidates based on variability and detections in other wavelength regimes (IR, UV, X-ray, and

radio). New results on their spectroscopic confirmations, and newly-discovered interesting objects, are presented.

4.2 Introduction

The selection of large samples of quasars, or active galactic nuclei (AGN), is a valuable tool in many areas of study, including galaxy evolution and black hole growth (Kollmeier et al. 2006), absorbing systems in the intergalactic medium (Hennawi & Prochaska 2007; McDonald & Eisenstein 2007), and determining the large scale structure of the universe (Ross et al. 2009). Extending these studies to significantly fainter quasars at any given redshift is a major goal for the next-generation Large Synoptic Survey Telescope (LSST; Ivezić et al. 2008). LSST will probe the luminosity function to $M_B = -23$ at $z \sim 2$, breaking the luminosity-redshift degeneracy that arises in current flux-limited samples such as the Sloan Digital Sky Survey (SDSS), which can only detect $M_B > -23$ quasars at $z < 0.5$. Also, our understanding of AGN fueling mechanisms and the detailed physics of accretion disks is far from complete (for a review, see McHardy 2010). Therefore, larger and more representative samples will only provide more potential for discovery.

Unfortunately, the resemblance of unresolved quasars (i.e., Type I broad line AGN) to stars in photometric surveys poses a difficulty in their identification. Quasars can be reliably identified by their spectra, and the SDSS provides a spectroscopically complete survey to $i < 19$ for the quasar region in SDSS color space (Richards et al. 2002) which includes ∼105,000 quasars (Schneider et al. 2010). However, it is becoming increasingly important to identify quasars in the absence of spectra, particularly since complete spectroscopy will be too expensive in upcoming large-area sky surveys. Previously, quasar selection in the absence of spectroscopy has best been achieved using $u - g$ and $g - r$ colors (e.g., Richards et al. 2002; Jiang et al. 2006; Richards et al. 2009). This technique requires multiple filters and results in samples with efficiencies[1] as low as 73%, and there is a redshift regime near $z \simeq 3$ where optical color selection largely fails. The upcoming large-area sky surveys, such as LSST and Pan-STARRS (Kaiser et al. 2002), plan to use their time-domain information

[1]Throughout this paper, "efficiency" (E) is equivalent to "purity" (p) in Schmidt et al. (2010).

to separate quasars from stars. In addition to color selection, it is expected that the lack of proper motion and selection by photometric variability properties will result in fairly clean samples with an efficiency much higher than the color-based 73%. Also, having an alternative method to color selection is needed to avoid a bias against extincted systems with non-standard colors. However, until very recently, it was not clear how variability could be used in a quantitative fashion to achieve these goals.

A simple way of using variability to separate quasars from stars is to take advantage of the fact that quasars are generally more variable than stars. Sesar et al. (2007) demonstrated that at least 90% of quasars have a root-mean-square (rms) variability of at least 0.03 mag. However, the rms by itself does not enable the separation of quasars from other variable objects such as RR Lyrae. A recent study by Schmidt et al. (2010, hereafter Schm10) showed that by measuring the rms as a function of time lag (Δt) between observations, i.e., the structure function $SF(\Delta t) = A(\Delta t/1\mathrm{yr})^{\gamma}$, the parameters A and γ enable the separation of quasars and stars with great efficiency ($\approx$95% for ultraviolet-excess objects). This separation is efficient because the rms of quasar variability is a few times less on monthly than on yearly time scales. Boutsia et al. (2009) showed that the variability of AGN enables their selection even for extended, host-dominated AGN (and X-ray quiet AGN).

An important piece of information missing from these analyses is the characteristic timescale for quasar variability, τ. Studies by de Vries et al. (2005) and Sesar et al. (2006) using SDSS combined with earlier Palomar Observatory Sky Survey measurements for 40,000 SDSS quasars constrained quasar continuum variability on timescales of 10 to 50 years in the observer's frame. They report that the characteristic timescale, which in this context is the time lag above which $SF(\Delta t)$ flattens to a constant value, is of order 1 year in the quasar rest frame. Recently, Kelly et al. (2009, hereafter KBS09) proposed a model where the optical variability of a given quasar is described by a damped random walk (DRW). The difference with respect to the well-known random walk is that an additional self-correcting term pushes any deviations back towards the mean flux on a timescale τ. It has been established by KBS09, Kozłowski et al. (2010b, hereafter Kozł10), and MacLeod et al. (2010) that a DRW can statistically explain the observed light curves of quasars at an impressive fidelity level (0.01-0.02 mag). The model has only three free parameters: the mean value of

the light curve, the driving amplitude of the stochastic process, and the damping (or characteristic) timescale τ. The predictions are only statistical, and the random nature reflects our uncertainty about the details of the physical processes. Kozł10 showed that the DRW model is an efficient quantitative method to separate quasars from most variable stars even in the dense stellar environments of the Magellanic Clouds.

Here, we test quasar selection in extragalactic fields by applying the DRW model to the individual light curves of ∼10,000 variable point sources with $i < 19$ from SDSS Stripe 82 (S82). This was also recently done by Butler & Bloom (2010, hereafter B&B) using a different approach based on the DRW model. About 15% of these sources are spectroscopically confirmed quasars whose variability properties have been analyzed in detail by MacLeod et al. (2010). We extend the selection technique in Schm10 by including information on the characteristic (damping) timescale, extend the study of Kozł10 to an extragalactic survey with extensive spectroscopic data, and extend the study in B&B by applying the DRW model to individual light curves rather than to the ensemble variability of quasars. We describe our variability model in Section 4.3, describe the S82 data set in Section 4.4, and present our results in terms of efficiency and completeness in Section 4.5. In Section 4.6, we explore the effects that the light curve length, cadence, and photometric errors have on the best-fit DRW parameters. We investigate the performance of the quasar selection technique presented here, in Kozł10, and in B&B, for typical (simulated) LSST and Pan-STARRS cadences. We summarize our findings in Section 4.7.

4.3 Quasar Variability as a Damped Random Walk

We model the time variability of quasars as a stochastic process described by the exponential covariance matrix

$$S_{ij} = \sigma^2 \exp(-|t_i - t_j|/\tau) \tag{4.1}$$

between times t_i and t_j. As detailed by KBS09 and Kozł10, this corresponds to a DRW (more specifically, an Ornstein-Uhlenbeck process) with a damping, or characteristic, time scale τ, and a long-term standard deviation of variability σ. The driving amplitude of short-term variations is defined as $\hat{\sigma} = \sigma\sqrt{2/\tau}$. Following Kozł10, we model the light

curves and estimate the parameters and their uncertainties using the method of Press et al. (1992), its generalization in Rybicki & Press (1992), and the fast computational implementation described in Rybicki & Press (1994). Kozł10 demonstrate that this approach is more statistically powerful than the forecasting methods used by KBS09, while still having computation times scaling linearly with the number of data points.

As in MacLeod et al. (2010), we express the long-term variability in terms of the structure function (SF), where the SF is the rms magnitude difference as a function of the time lag (Δt) between measurements. The characteristic timescale for the SF to reach an asymptotic value SF_∞ is the damping timescale, τ. The SF for a DRW is

$$SF(\Delta t) = \mathrm{SF}_\infty(1 - \mathrm{e}^{-|\Delta \mathrm{t}|/\tau})^{1/2}, \tag{4.2}$$

and the asymptotic value at large Δt is

$$SF(\Delta t >> \tau) \equiv \mathrm{SF}_\infty = \hat{\sigma}\sqrt{\tau}. \tag{4.3}$$

For short time lags,

$$SF(\Delta t << \tau) = \mathrm{SF}_\infty\sqrt{\frac{|\Delta \mathrm{t}|}{\tau}} = \hat{\sigma}\sqrt{|\Delta \mathrm{t}|}. \tag{4.4}$$

Therefore, $\hat{\sigma}$ regulates the rise of $SF(\Delta t << \tau)$. The $\hat{\sigma}$ parameter is related to γ in the equation $SF(\Delta t) = A(\Delta t/1\mathrm{yr})^\gamma$, where A and γ were used to select quasars in Schm10. However, A and γ do not provide unique information on the characteristic timescale τ (although since $\hat{\sigma} \propto \tau^{-1/2}$, there is some information on τ). While we obtain individual estimates of τ and $\hat{\sigma}$ for every light curve, in B&B τ and $\hat{\sigma}$ are estimated based on the apparent magnitude, which follows from modeling the ensemble SF as a DRW as a function of magnitude. We discuss this difference further in Section 4.5.2.

In Section 4.5, we demonstrate that the inclusion of τ, in addition to $\hat{\sigma}$, in the selection of quasars enables the selection of highly pure samples with high completeness.

4.4 The SDSS Stripe 82 Data Set

The Sloan Digital Sky Survey (SDSS, York et al. 2000) provides homogeneous and deep ($r <$ 22.5) photometry in five passbands (*ugriz*, Fukugita et al. 1996; Gunn et al. 1998; Smith et

al. 2002). To test our quasar selection method, we utilize the g-band light curves from the SDSS Stripe 82 (S82), which covers the sky region defined by 22hr 24m < RA < 4hr 8m and $-1.27^\circ <$ Dec $< 1.27^\circ$ (an area of $\approx$290 deg^2). The light curves span about ten years, and the observations are clustered into yearly seasons about 2–3 months long. There are on average more than 60 available epochs. Because some observations were obtained in non-photometric conditions, improved calibration techniques have been applied to the SDSS S82 data by Ivezić et al. (2007) and Sesar et al. (2007), and we use their results. For these data, photometric zero-point errors are 0.01–0.02 mag.

We apply the DRW model to all objects in the S82 variable point source catalog (Ivezić et al. 2007). These are objects with at least ten observations, an rms of variability in the g and r bands exceeding 0.05 mag, and a χ^2 per degree of freedom exceeding 3 for the light curves in these bands being fit with a constant flux. We only include observations spaced by at least half a day apart when applying the model, since we are interested in the selection of quasars which generally vary on longer timescales than variable stars. The S82 data set is spectroscopically complete for $i < 19$ in the quasar region of $(u-g)$, $(g-r)$, and $(r-i)$ color space (Richards et al. 2002; see the top panel of Figure 4.1 for the total distribution of quasars in i). We include newly confirmed DR7 (Abazajian et al. 2009) quasars, and limit our data set to the 52,547 variable objects with $i < 19$ when evaluating quasar selection based on variability (Section 4.5.2). In this sample, there are 1,912 (4%) spectroscopically identified quasars.

The distribution of the known quasars in RA is compared to the distribution of all variable sources in the bottom panel of Figure 4.1. The quasar counts fall off at RA < -35 deg, while the full distribution rises due to the rising density of Galactic stars. Therefore, we further restrict our sample to the 10,024 objects with $-35^\circ <$ RA $< 50^\circ$ in order to minimize the stellar contamination from the Galaxy. This sample of 10,024 variable sources, including 1,490 (15%) spectroscopically confirmed quasars, defines our final sample when evaluating quasar selection based on variability in Section 4.5.2. However, to demonstrate the performance in regions of increased stellar density, we also compare to the results for the region $-35^\circ <$ RA $< -25^\circ$.

It is important to note that for variable objects, the single-epoch i magnitudes can be

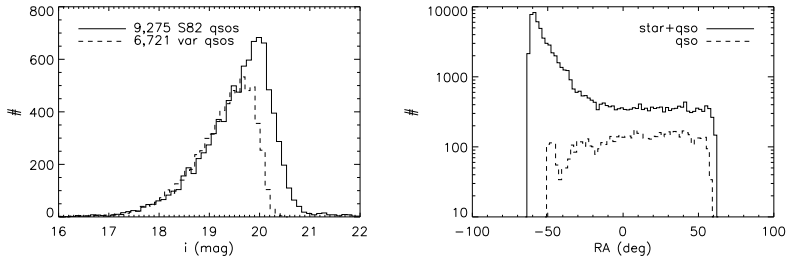

Figure 4.1 *Left panel:* quasar counts as a function of i magnitude for all spectroscopically confirmed S82 quasars (MacLeod et al. 2010; solid line) and those in the S82 variable point source catalog (Ivezić et al. 2007; dashed line). *Right panel:* counts for all variable point sources (solid) and for spectroscopically confirmed quasars (dashed), as a function of Right Ascension. In order to minimize stellar contamination from the Galactic disk, we only include sources with $-35^\circ < \mathrm{RA} < 50^\circ$ in the subsequent analysis (with Galactic latitudes $-45^\circ < b < -37^\circ$).

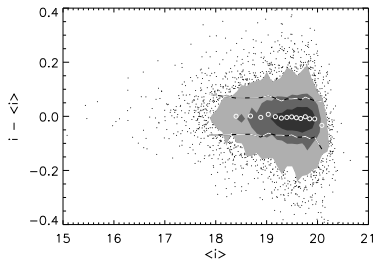

Figure 4.2 Differences in the SDSS BEST i magnitude and best-fit mean i magnitude, $\langle i \rangle$, as a function of $\langle i \rangle$ for S82 quasar light curves modeled as a DRW. The differences arise because the SDSS best i magnitudes are for a single epoch. The $\pm 1\sigma$ range is shown with dashed lines. The contours show regions containing 25%, 50%, and 90% of the total number of data points.

very different from the best-fit mean magnitudes from the DRW model, $\langle i \rangle$, since the latter are based on the entire time series. In Figure 4.2, the distribution of $i - \langle i \rangle$ is shown as a function of $\langle i \rangle$ for S82 quasars. It is centered on zero with the $\pm 1\sigma$ range spanning almost 0.1 mag. Therefore, these differences are significant. For subsequent analysis, we use the median PSF i-band magnitudes as listed in the variable point source catalog, which are similar to the $\langle i \rangle$ values. All magnitudes are corrected for interstellar dust extinction using Schlegel et al. (1998).

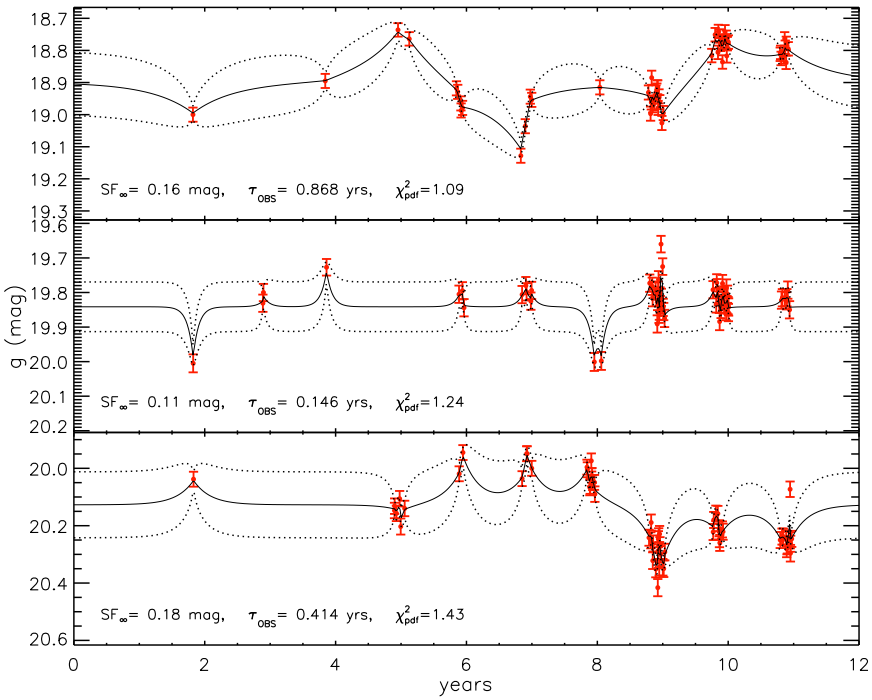

Figure 4.3 The top panel shows a sample light curve for a confirmed quasar. Data points with error bars show the observed data, and the solid line shows the weighted average of all consistent DRW models. Dotted lines show the $\pm 1\sigma$ range of these possible stochastic models about the average. The middle panel is a similar plot for an unknown object whose light curve can only marginally be distinguished from noise ($\Delta L_{\rm noise} = 0.4$). The bottom panel is a similar plot for a confirmed quasar with an outlying data point around 11 years, leading to a large $\chi^2_{\rm pdf}$. The best-fit variability parameters along with $\chi^2_{\rm pdf} = \chi^2/N_{\rm dof}$ are listed at the bottom of the panels.

4.5 Results

4.5.1 Variability of Quasars and Other Objects

In this Section, we present the best-fit variability parameters after applying the DRW model to the light curves of all variable point sources in S82. These parameters are used to develop a method for separating quasars from other variable point sources. We compare the results for two subsamples: one containing the spectroscopically confirmed quasars, and one containing all other objects (which is dominated by stars). We also investigate in detail the objects in the latter subsample that have best-fit timescales greater than 100 days, which is indicative of quasar variability.

Figure 4.3 shows examples of fitted light curves. The top panel shows a spectroscopically confirmed quasar whose variability is well-fit by the model, and the middle panel shows an unknown object. The weighted average of all model light curves consistent with the data is shown, and the "error snake" is the $\pm 1\sigma$ range of those light curves about this mean. The best-fit variability parameters along with $\chi^2_{\rm pdf} = \chi^2/N_{\rm dof}$ are listed at the bottom of the panels, where $N_{\rm dof}$ is the number of degrees of freedom. Additional information on the goodness of fit is provided by the parameter $\Delta L_{\rm noise} \equiv \ln(L_{\rm best}/L_{\rm noise})$ which was used in Kozł10 and MacLeod et al. (2010) to select light curves that are better described by a DRW than by pure white noise. Here, $L_{\rm best}$ is the likelihood of the best-fit stochastic model and $L_{\rm noise}$ is that for a white noise solution ($\tau \equiv 0$). In addition to having a larger $\chi^2_{\rm pdf}$, the light curve for the unclassified object in the middle panel of Figure 4.3 has $\Delta L_{\rm noise} = 0.4$ and is therefore only marginally more consistent with the DRW model than with white noise. In some cases (see the bottom panel), outliers in the light curve cause a relatively large $\chi^2_{\rm pdf}$ while the likelihood is high for a DRW ($\Delta L_{\rm noise} = 34$). Outliers such as these could either represent a model failure or a bad data point with a true error that is larger than the reported value. A detailed analysis of these cases will be discussed in a future publication.

In Figure 4.4, $\chi^2_{\rm pdf}$ is shown as a function of $\Delta L_{\rm noise}$ for the known quasars and all other objects in the variable S82 sample. The quasars typically have $\chi^2_{\rm pdf} \sim 0.85$ while all other variable S82 objects typically have $\chi^2_{\rm pdf} \sim 1.0$, with a large overlap between the two distributions. The overlap is greatly reduced when considering the marginal distributions of $\Delta L_{\rm noise}$. While $\chi^2_{\rm pdf}$ barely resolves two populations (quasars and stars), $\Delta L_{\rm noise}$ efficiently separates quasars (which are well-fit by the stochastic model) from variable stars (which are better fit by pure noise). Objects with high $\Delta L_{\rm noise}$ also tend to vary on longer timescales (see Figure 4.5). Therefore, the characteristic timescale τ also provides an efficient way to distinguish quasars from stars.

The distributions of τ and SF_∞ for the quasar and stellar subsamples are shown in Figure 4.6. The τ distribution (in the observer's frame) peaks near 500 days for the known quasars and around 1 day for the other objects, providing an efficient and simple way to distinguish between the two subsamples. We note that because quasars are known to have aperiodic variability, eliminating periodic sources should lead to a cleaner sample of quasars.

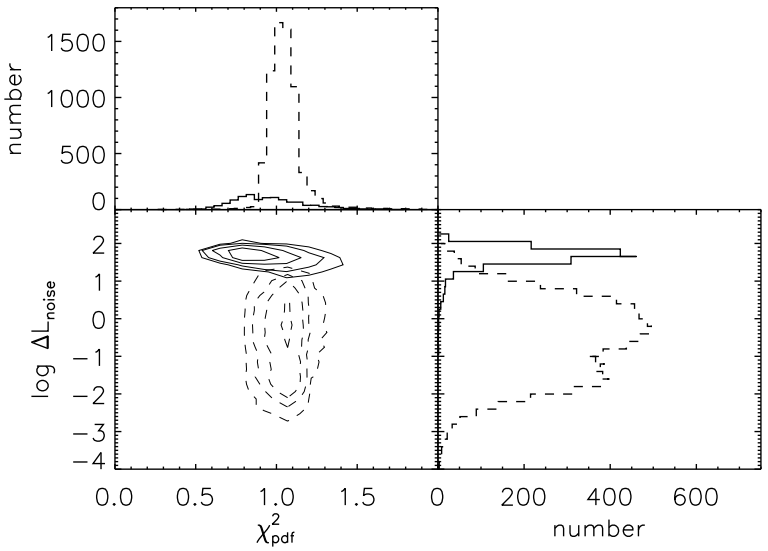

Figure 4.4 Goodness of fit parameters $\chi^2_{\rm pdf} = \chi^2/N_{\rm dof}$ versus $\log \Delta L_{\rm noise}$ for spectroscopically confirmed quasars (solid) and all other objects (dashed) in our S82 data set restricted to $i < 19$. The contours show regions containing 30, 60, 80, and 90% of data points in each subsample. The marginal distributions are shown in the upper and right-hand panels.

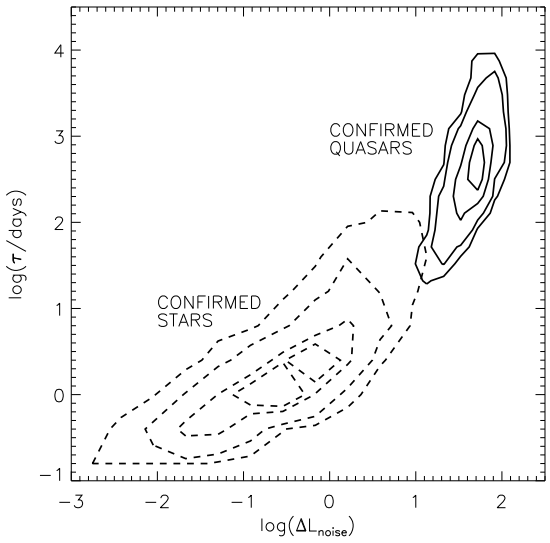

Figure 4.5 Dependence of the best-fit characteristic time scale on the relative DRW likelihood $\Delta L_{\rm noise}$. Solid (dashed) contours show the 25%, 50%, 75%, and 90% levels for spectroscopically confirmed quasars (stars).

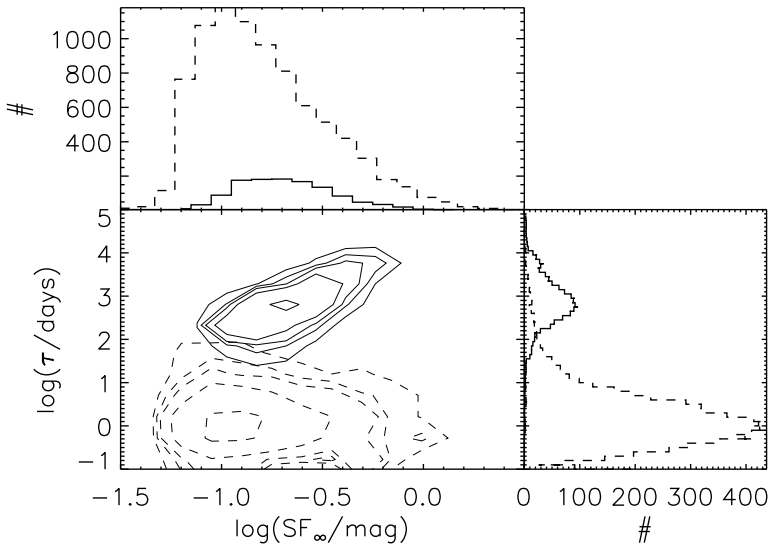

Figure 4.6 Characteristic timescale τ versus asymptotic rms variability SF_∞ for spectroscopically confirmed quasars (solid) and all other objects (dashed) in our S82 data set restricted to $i < 19$. The contours show regions containing 10, 60, 75, 85, and 90% of data points in each subsample. The marginal distributions are shown in the upper and right-hand panels.

We could classify variable sources as periodic if the likelihood for the peak in the Lomb–Scargle periodogram (as implemented by Press et al. 1992b) being observed at random is $\log(p_{\rm periodic}) < -3$ and the estimated period is less than ∼200 days. For longer periods, many quasars are falsely flagged as periodic because the periodogram only tests for whether a sine wave is a better fit than a constant, and this is frequently true for quasars with only a few "oscillations" in the data (see Kozł10). Because we have few longer period variable stars in S82, adding such a cut for periodic stars adds little and so we have not included it. This differs from the Magellanic Cloud fields examined by Kozł10 where there are many longer-period periodic stars (e.g., Cepheids).

The use of τ as an efficient classifier is further illustrated in Figure 4.7. The top-left panel shows the $(u-g)$ and $(g-r)$ colors for spectroscopically confirmed quasars and stars. The top-right panel shows the median value of τ as a function of $(u-g)$ and $(g-r)$ color. This panel shows that the quasar locus (region II) as well as the high redshift quasars ($z \gtrsim 3.2$; region IV) are dominated by long timescales, while the stellar locus (region I) and the RR Lyrae (region III) have short timescales. The timescale information is able to clearly

distinguish between quasars and RR Lyrae, as opposed to when using the rms of variability alone (see Figure 3 in Sesar et al. 2007). The quasar regions of color space also tend to have larger median values of SF_∞, as seen in the bottom-left panel of Figure 4.7. The remaining panel shows the median driving amplitude of short-term variations, $\hat{\sigma} = \mathrm{SF}_\infty/\sqrt{\tau}$, against $(u-g)$ and $(g-r)$ color. Because it is anti-correlated with τ (see Kozł10), the quasar regions show systematically smaller $\hat{\sigma}$ than the stellar regions.

Figure 4.8 compares the distribution of objects with $\tau \geq 100$ days and $\tau < 100$ days in $(u-g, g-r)$ space, as well as the subset of 233 objects with $100 \leq \tau < 10^5$ days and $\Delta L_{\mathrm{noise}} > 2$ that are not spectroscopically confirmed quasars. The latter contains good quasar candidates, especially those in regions II or IV (51 objects). Thirty-five of these are spectroscopically confirmed stars, represented by the star symbols. Figure 4.9 shows the distribution of these 233 objects in $(\tau, \mathrm{SF}_\infty)$, $(i, i-z)$, and $(r-i, i-z)$ space. The subset in regions II or IV, represented by triangles, are clustered near the quasar region at faint i and blue $(r-i, i-z)$ colors. We inspected their observed and phased light curves (using the Supersmoother algorithm, see Reimann 1994), and they appear consistent with quasar light curves with no signs of periodicity. About half of these 51 objects are detected in GALEX[2] GR6 with GALEX-SDSS UV-optical colors resembling those of quasars ($n-u < 1.5$ and $u-g < 0.8$, or $f-n > 0$ and $g-r < 0.5$; see Agüeros et al. 2005). About 15% have 2MASS[3] colors $J-K > 1$ that are also indicative quasar nature. We also investigated the 15 unknown objects with $\tau \geq 100$ days that lie in region I but off the stellar locus (above the solid line in Figure 4.8). These objects (represented by asterisks in Figure 4.9) seem consistent with reddened quasars, as their light curves are consistent with quasars, and they are rather faint and clustered at redder $(r-i, i-z)$ colors. However, their median SF_∞ (0.15 mag) is lower than that for the quasar candidates represented by triangles (0.22 mag). None of these objects have GALEX-SDSS UV-optical colors indicative of quasars, but about 30% have $J-K > 1$. Follow-up spectroscopy will be useful for classifying these two subsamples.

The remaining objects with $\tau \geq 100$ days lying on the stellar locus or in region III seem

[2]Galaxy Evolution Explorer; see http://galex.stsci.edu/GR6/. A matching radius of 3″ was used.

[3]Two Micron All Sky Survey; see http://www.ipac.caltech.edu/2mass/. A matching radius of 1.5″ was used.

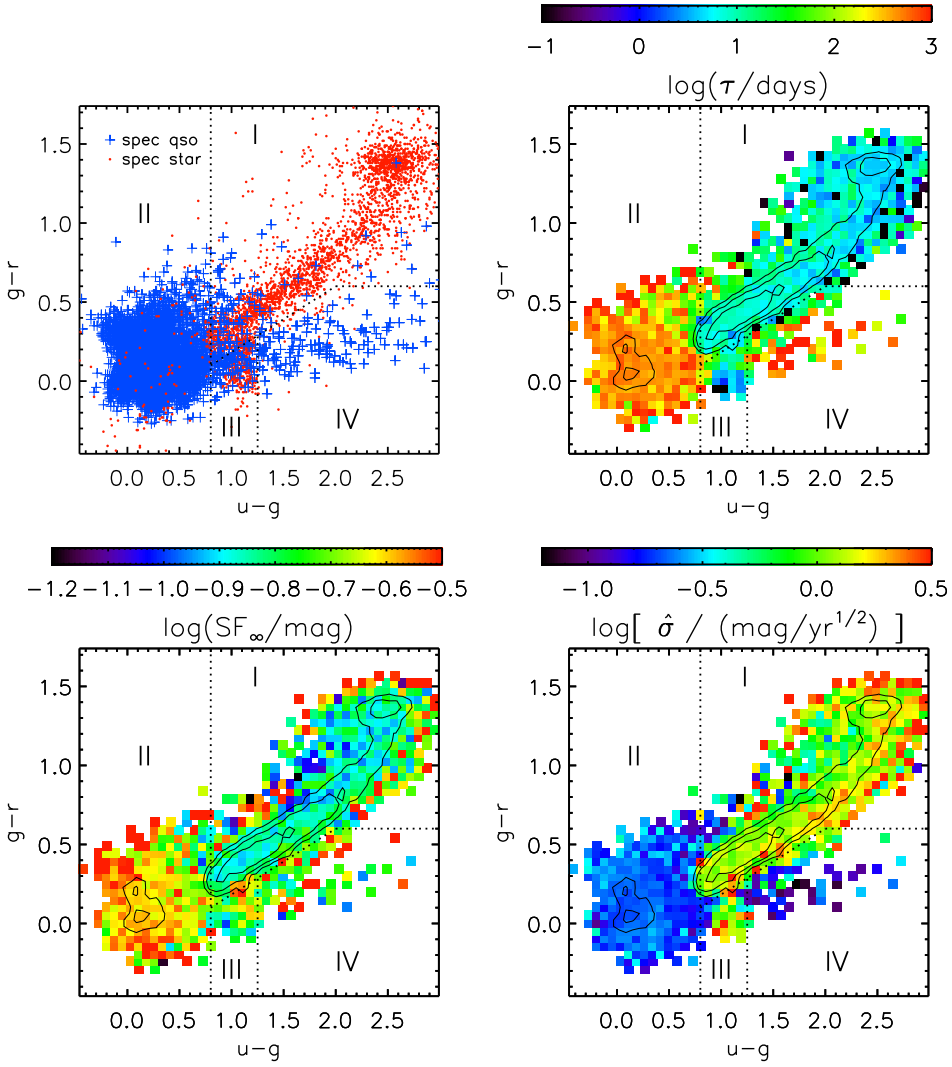

Figure 4.7 *Top-left:* $u-g$ versus $g-r$ colors of all spectroscopically confirmed S82 quasars (crosses) and spectroscopically confirmed stars (dots). *Top-right:* The median characteristic time scale τ for all variable point sources in S82 that are well-described by the DRW model ($\Delta L_{\rm noise} > 2$), as a function of $u-g$ and $g-r$, following the legend at the top. The contours show regions containing 25%, 50%, and 75% of the total number of data points. The red regions with large τ coincide with the quasar color regions, providing a convenient method for quasar selection based on variability. *Bottom-left:* As in top-left panel but for the asymptotic rms of variability, SF_∞. *Bottom-right:* As in top-left panel but for the driving amplitude of short-term variations, $\hat{\sigma} = \mathrm{SF}_\infty/\sqrt{\tau}$. The dotted lines divide the color space into four characteristic regions: the stellar locus (I), the low-redshift ($z \lesssim 2$) quasar region (II), the RR Lyrae region (III), and the high-redshift ($z \gtrsim 3.2$) quasar region (IV).

consistent with stars based on their i magnitude and colors, and many are spectroscopically confirmed as stars. The majority have small amplitudes ($\mathrm{SF}_\infty \approx 0.1$ mag). However,

there is a subset with large amplitudes ($SF_\infty > 0.8$ mag). The majority of these sources have a single outlier at a significantly fainter magnitude in their light curves (e.g., SDSS J001014.28+005938.6, see Figure 4.10) causing the model to fit a long timescale. These could be wide eclipsing binaries or simply a bad data point. If we reject the data point with the largest magnitude difference from the median in each light curve with $SF_\infty > 0.8$ mag, and then refit the models, half now have $SF_\infty < 0.2$ mag and 60% now have $\Delta L_{\rm noise} < 2$ and $\tau < 100$ days. There is a subset of these sources with large amplitude semi-periodic outbursts, suggestive of AGB stars (e.g., SDSS J220514.58+000845.7, see Figure 4.10). In general, however, inspection of both the phased and observed light curves revealed no examples of truly periodic behavior. A few (2%) of these objects have GALEX-SDSS UV-optical colors indicative of quasars, and about 5% have $J - K > 1$.

We compile a list of quasar candidates by removing those that already have SDSS spectra from the list of 233 and including cases where τ has saturated to 10^5 days. We retain 255 objects lacking SDSS spectra that have $i < 19$, $\Delta L_{\rm noise} > 2$, and $\tau \geq 100$ days, leading to a sample of 255. The coordinates, SDSS photometry, and DRW parameters for these candidates are provided in Table 1 in the online version of MacLeod et al. (2011b). Note that 30% of the target sample fails the $\Delta L_{\rm noise} > 2$ and $\tau \geq 100$ day criteria when omitting the most outlying data point in the light curve. A subsample of these 255 were targeted in a spectroscopic follow-up study and are presented along with their spectroscopic classifications in Table 4.4.

4.5.2 Completeness and Efficiency

To measure the success of our quasar selection, we report the completeness (C) and efficiency (E), defined by

$$C = \frac{\#\ of\ selected\ confirmed\ quasars}{total\ \#\ of\ confirmed\ quasars} \times 100$$
$$E = \frac{\#\ of\ selected\ confirmed\ quasars}{total\ \#\ of\ selected\ objects} \times 100, \tag{4.5}$$

for a variety of selection criteria, where "confirmed" means that the source is spectroscopically identified as a quasar. Here, we only consider sources with $i < 19$, and E is a lower

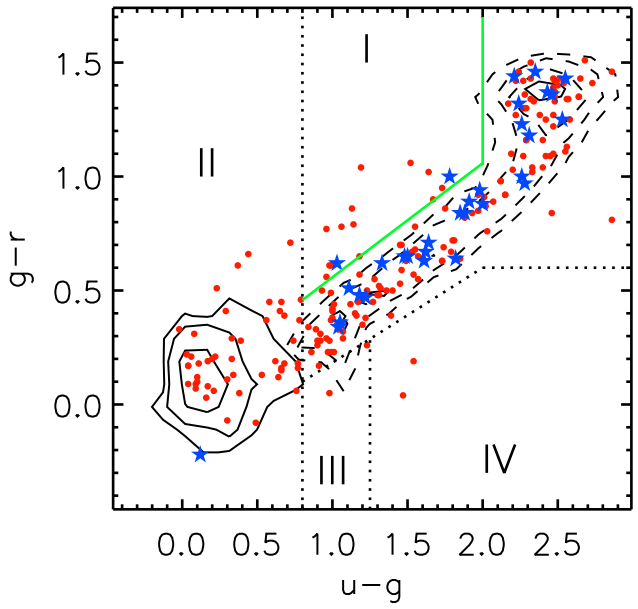

Figure 4.8 $u-g$ versus $g-r$ colors of bright ($i < 19$) variable point sources with $\tau \geq 100$ days (solid contours) and $\tau < 100$ days (dashed; 25%, 60%, and 75% levels). Also shown are the 233 objects with $100 \leq \tau < 10^5$ days and $\Delta L_{\rm noise} > 2$ that are not spectroscopically confirmed quasars (red dots). Those with confirmed stellar origin are represented by blue stars. We use the solid line in region I to define a subsample of objects that lie off the stellar locus.

limit for the true efficiency because the sample of confirmed quasars may not be 100% spectroscopically complete. These quantities for a quasar selection based on τ alone are shown in Figure 4.11 as a function of the minimum threshold τ. The two quantities are anti-correlated due to the increased overall fraction of selected quasars as well as the increased number of missed quasars when restricting τ to longer timescales. The completeness drops to 94% (where $E = 81\%$) at a threshold of $\tau > 100$ days, beyond which C dramatically decreases. At the same time, E for $\tau > 100$ days is almost as high as its asymptotic value (85%). Therefore, we adopt $\tau \geq 100$ days as the optimal cut for selecting quasars based on timescale alone. An additional constraint of $\Delta L_{\rm noise} > 2$ (thick line in Figure 4.11) does not affect the resulting C or E by more than 1% after requiring $\tau \geq 100$ days, while the

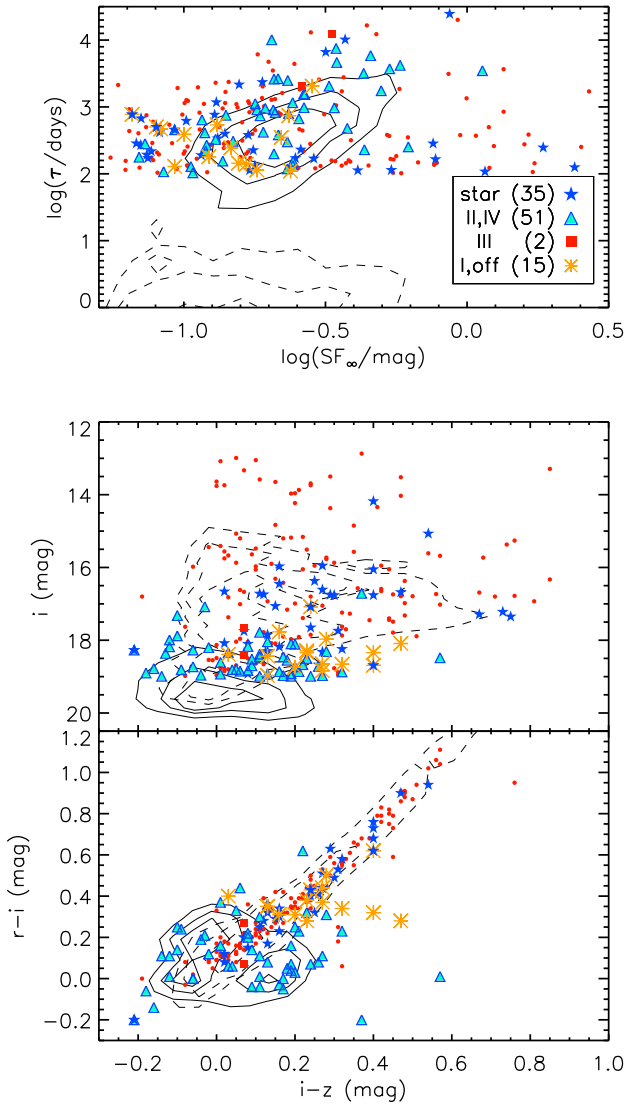

Figure 4.9 *Top*: Variability timescale τ versus amplitude SF_∞ for all spectroscopically confirmed quasars (solid contours), all confirmed stars (dashed), and the 233 objects with $100 \leq \tau < 10^5$ days, $\Delta L_{\rm noise} > 2$, and $i < 19$ that are not already confirmed quasars (symbols). Blue stars represent those with confirmed stellar origin, and cyan triangles represent those with quasar-like colors (regions II or IV of Figure 4.7). Red squares represent those in region III, and orange asterisks represent those that lie in region I but above the solid line in Figure 4.8. The number of objects in each subset is listed in the legend, and the remaining 130 objects are represented by red dots. The contour levels are 25%, 50%, and 75%. *Middle and bottom*: As in top panel but for i magnitude and $r - i$ color, respectively, as a function of $i - z$.

constraint $\Delta L_{\rm noise} > 10$ (gray lines) increases E with a minor drop in C.

Table 4.1 lists C and E for various selection criteria, beginning with the i magnitude and

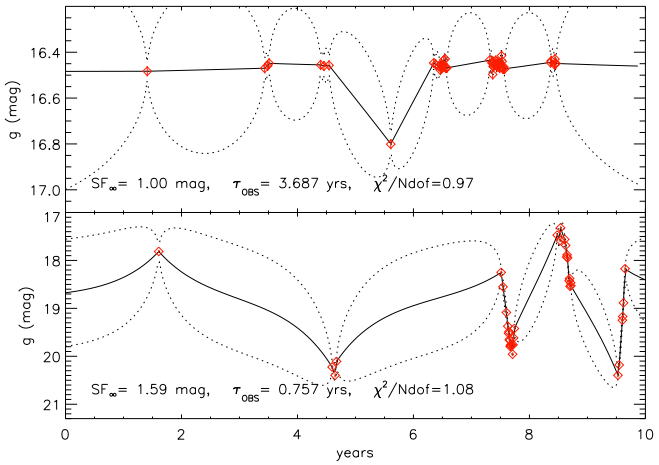

Figure 4.10 Light curves for SDSS J001014.28+005938.6, a probable example of a false variable created by a bad data point (top), and SDSS J220514.58+000845.7, a probable example of a non-periodic variable star (bottom). Data points with error bars show the observed data, and the solid line shows the weighted average of all consistent DRW models. Dotted lines show the $\pm 1\sigma$ range of these possible stochastic models about the average. The best-fit variability parameters are listed at the bottom of the panels.

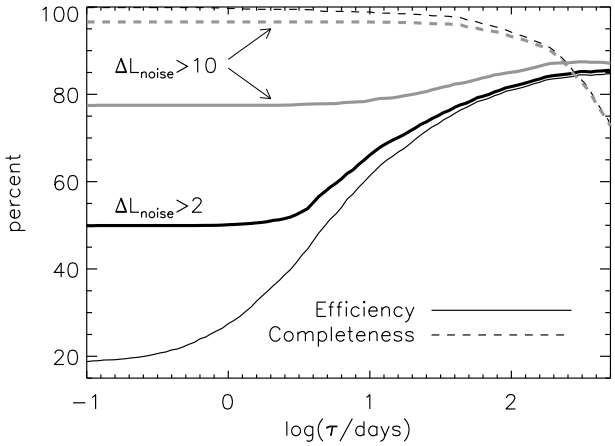

Figure 4.11 E (solid curves) and C (dashed curves) measured as a function of the minimum threshold τ for variable S82 point sources with $i < 19$ and $-35^\circ < \mathrm{RA} < 50^\circ$. The thin, solid and dashed black curves show the results when using a τ criterion alone. The thick, solid black curve shows the effect of adding a criterion of $\Delta L_{\mathrm{noise}} > 2$ (C is unaffected), and the gray curves show the effect of adding $\Delta L_{\mathrm{noise}} > 10$.

RA limits that define our starting sample (first row, hence the 100% completeness), and the restrictions $\tau \geq 1$ and 100 days (second and third rows). When excluding the objects that fail the $\Delta L_{\mathrm{noise}} > 2$ and $\tau \geq 100$ day criteria after omitting the most outlying data point in

the light curve, E rises from 81% to 87%, while C only drops by 1% (fourth row). This cut is useful for eliminating potentially bad data in order to boost the efficiency of the selected sample. C can be increased to 96% while maintaining $E = 87\%$ by reducing the limit to $\tau \geq 10^{1.5}$ days and requiring $\Delta L_{\rm noise} > 10$ (fifth row). Here, we have excluded objects that drop out of the selection after removing the most outlying data point. When further restricting the variability-selected sample (with $\tau \geq 100$ days) to those objects which have quasar colors (regions II and IV in Figure 4.7), C decreases to 91% while E rises to 96% (sixth row; these results are unaffected when rejecting light curve outliers). Table 4.1 also lists the percentage of each selected sample that have SDSS spectra, and the breakdown into quasars, stars, and galaxies based on their SDSS spectroscopic class.

We also compute E and C for the ultraviolet-excess ("UVX") and non-ultraviolet-excess ("nUVX") color boxes as defined in Schm10. The former should be spectroscopically complete to $i < 19$ (see Richards et al. 2002). The latter corresponds to the quasar redshift regime of $\sim$3, where color selection largely fails. Here, C is the percentage of all confirmed variable (n)UVX quasars that have $\tau \geq 100$ days. For UVX objects, $C = 95\%$ and $E = 97\%$ for the $\tau \geq 100$ days selection criteria (eighth row of Table 4.1). For comparison, the Schm10 selection method, which does not use the τ information, gives $C = 90\%$ and $E \simeq 95\%$. For nUVX objects (only 52 objects; tenth row), $C = 100\%$ for the $\tau \geq 100$ days selection criteria, and E is only known to be $> 69\%$ due to the fact that only 69% have spectra.

The driving amplitude $\hat{\sigma}$ also shows a large contrast between the quasar color regions (II and IV) and the stellar regions (I and III) in Figure 4.7. This suggests that including information on the driving amplitude may improve the quasar selection based on variability. Indeed, Schm10 found high-efficiency quasar samples by using information closely corresponding to $\hat{\sigma}$. In Figure 4.12, we show how the selection can be improved by including information on both τ and $\hat{\sigma}$. In the top two panels, E and C are mapped out as a function of the minimum τ and the maximum $\hat{\sigma}$ allowed for quasar classification. For every value of $\hat{\sigma}$, there is a lower limit on τ that boosts efficiency with a minor drop in C. This essentially follows one of the quasar selection criteria for τ and $\hat{\sigma}$ in Kozł10. A similar improvement is seen when instead using lower limits in both τ and SF_∞, or a combination of τ, SF_∞,

and $\hat{\sigma}$ information (the differences are all $\lesssim 1\%$). In the bottom panels of Figure 4.12, it can be seen more quantitatively how the inclusion of τ improves the selection of quasars as compared to using $\hat{\sigma}$ alone. The panels show the maximum C achieved for an E exceeding that plotted on the x-axis, for three cases. In the first case (solid), only the $\hat{\sigma}$ parameter is used to separate quasars and stars, i.e., all values of τ are allowed, as in Schm10. The second case (dashed) shows the effect of including information on τ. The third case (dotted) shows results for using τ information alone. When including all SDSS colors (bottom-left panel), the τ information alone leads to significantly higher E than a selection based on $\hat{\sigma}$ alone. However, for UVX objects (bottom-right panel), a combination of τ and $\hat{\sigma}$ information is required to significantly boost E from a selection based on $\hat{\sigma}$ alone.

Using both τ and $\hat{\sigma}$ as criteria for quasar selection in the absence of color information, one can achieve an efficiency of $E = 82\%$ with a completeness of $C = 96\%$ (seventh row of Table 4.1), or alternatively, $E = 75\%$ for $C = 98\%$, and $E = 85\%$ for $C = 90\%$. Here, objects that drop out of the selection after omitting the most outlying data point in the light curve are excluded. When restricting the selection to color regions II and IV in Figure 4.7, one can achieve $E = 96\%$ with $C = 93\%$. At the same level of completeness, $E = 84\%$ without color cuts, and $E = 78\%$ when using $\hat{\sigma}$ alone. For UVX objects, $E = 97\%$ and $C = 98\%$ when selecting $\tau \geq 10^{1.56}$ days and $\hat{\sigma} \leq 10^{-0.18}$ mag yr$^{-1/2}$ (ninth row of Table 4.1). For illustration, Figure 4.13 shows τ versus SF_∞ for the variable S82 sample, with lines of constant $\hat{\sigma}$ overlaid. By choosing $\hat{\sigma} < 10^{-0.2}$ mag yr$^{-1/2}$, the entire quasar locus is selected as well as the gray shaded region containing stars. However, when imposing a lower limit at $\tau = 10^{1.56}$ days, these contaminants are excluded from the sample, leading to a higher efficiency of quasar selection.

Since the selection of quasars based on variability is a useful tool in regions of high stellar density (Kozł10), we repeat the analysis in Table 4.1 for the region $-35\,^\circ < \mathrm{RA} < -25\,^\circ$, where the density of Galactic stars rises. Since the rate of contamination is higher in this region, it is useful to reject outliers in light curves in order to boost efficiency as in the fourth row of Table 4.1. These results can be found in Table 4.2. Here, E for the $\tau \geq 100$ days selection is 79%, after omitting outliers in light curves. Restricting this selection to $\Delta L_{\mathrm{noise}} > 10$ boosts E to 82% without a drop in completeness. Therefore, a

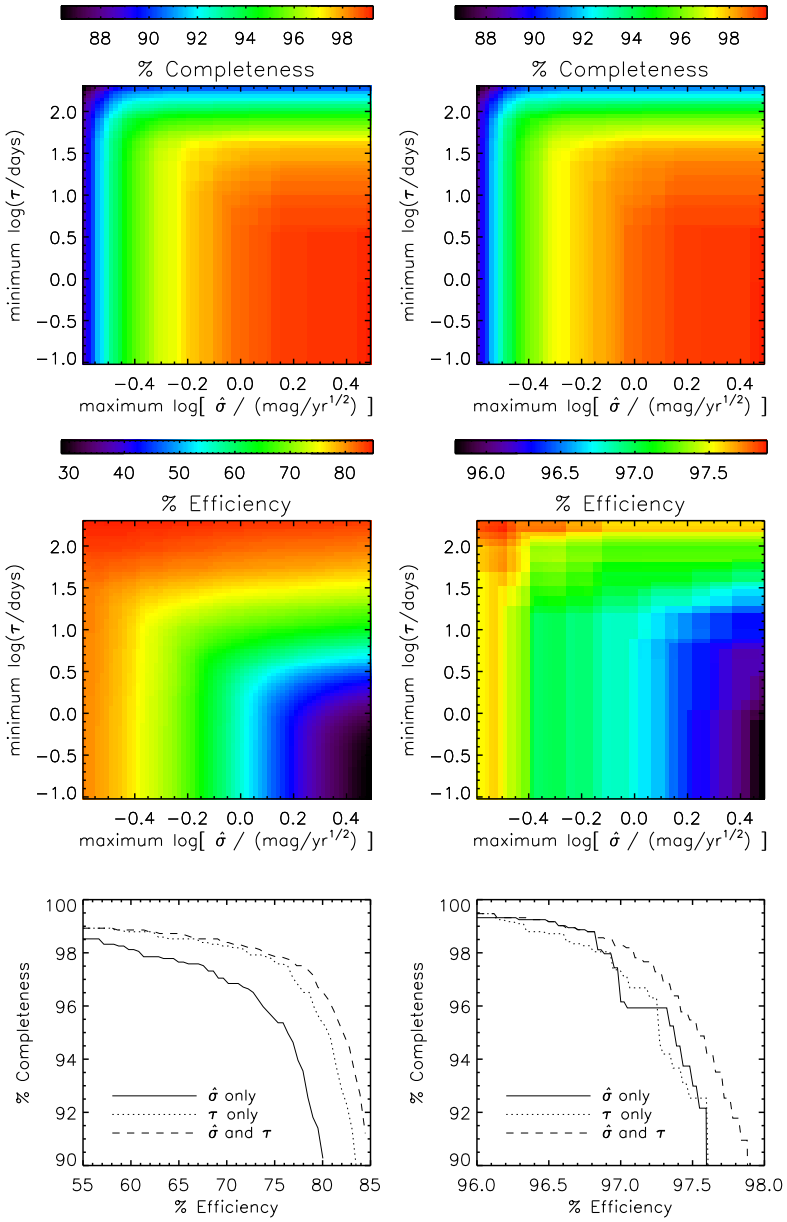

Figure 4.12 *Left panels:* Improvements in variability-based selection of quasars due to added timescale information. The top two panels show the the completeness (C) and efficiency (E) for quasar classification as a function of the minimum τ and maximum $\hat{\sigma}$ allowed for the variable sample with $i < 19$ and $-35^\circ <$ RA $< 50^\circ$. The solid line in the bottom panel shows the maximum C as a function of E achieved when using $\hat{\sigma}$ information alone (i.e., allowing all timescales τ). The dashed line shows how E improves when including a cut in τ to select quasars in addition to a cut in $\hat{\sigma}$. The dotted line shows results when using the τ information alone. *Right panels:* As in left panels but for UVX objects only.

strict cut in $\Delta L_{\rm noise}$ is especially useful in regions of high stellar density.

Table 4.1. Selection Criteria for $-35\,^\circ < \mathrm{RA} < 50\,^\circ$ and $i < 19$

	Selection	N	C	δC	E	δE	Spec.	QSO	Star	Gal.	Unk.
1.	All Sources	10024	100	4	15	0.4	33	45	53	2.5	0.15
2.	τ >1 d	5385	99.7	4	28	0.8	43	64	33	2.0	0.22
3.	$\tau \geq 100$ d	1734	94	3	81	3	84	96	3.5	0.55	0.14
4.	$\tau \geq 100$ d*	1594	93	3	87	3	89	98	1.76	0.49	0.14
5.	$\tau \geq 10^{1.5}$ d, $\Delta L_{\mathrm{noise}} > 10$*	1647	96	4	87	3	89	97	2.0	0.61	0.20
6.	$\tau \geq 100$ d, qso-like colors	1420	91	3	96	4	97	99	0.15	0.44	0.15
7.	$\tau \geq 10^{1.56}$ d, $\hat{\sigma} \leq 10^{-0.18}$ mag yr$^{-1/2}$*	1751	96	4	82	3	86	96	3.3	0.87	0.07
8.	**UVX:** $\tau \geq 100$ d	1297	95	3	97	4	98	99	0.16	0.32	0.16
9.	$\tau \geq 10^{1.56}$ d, $\hat{\sigma} \leq 10^{-0.18}$ mag yr$^{-1/2}$	1332	98	3	97	4	98	99	0.15	0.31	0.08
10.	**nUVX:** $\tau \geq 100$ d	52	100	0.4	69	15	69	100	0.00	0.00	0.00
11.	B&B	1702	96	4	84	3	87	97	2.7	0.40	0.13

*The most outlying data point in each light curve is omitted.

Note. — The first column lists various selection criteria ("qso-like colors" signifies having $u-g$ and $g-r$ colors in regions II or IV of Figure 4.7). N is the total number of objects that satisfy each criterion (the sample is defined by $i < 19$ and covers 213 deg^2). C is the quasar completeness, or the percentage of all confirmed quasars in the sample of 10,024 that are selected. E is the efficiency, or the percentage of the selected objects that are confirmed quasars. δC and δE are the Poisson errors on C and E. "Spec." indicates the percentage that have an SDSS spectrum, and the last four columns show the percentage breakdown of the subset with spectra into four spectroscopic classes: QSO (SpecClass is 3 or 4), Star (SpecClass is 1 or 6), Galaxy (SpecClass is 2), and Unknown (SpecClass is 0). In the eighth and ninth rows (tenth row), we list results for the (n)UVX color box described in Schm10, where C is computed as the percentage of all confirmed quasars contained in the (n)UVX color box that satisfy the variability criteria listed in the first column. The eleventh row lists the results for the selection criteria in B&B (see Section 4.5.2).

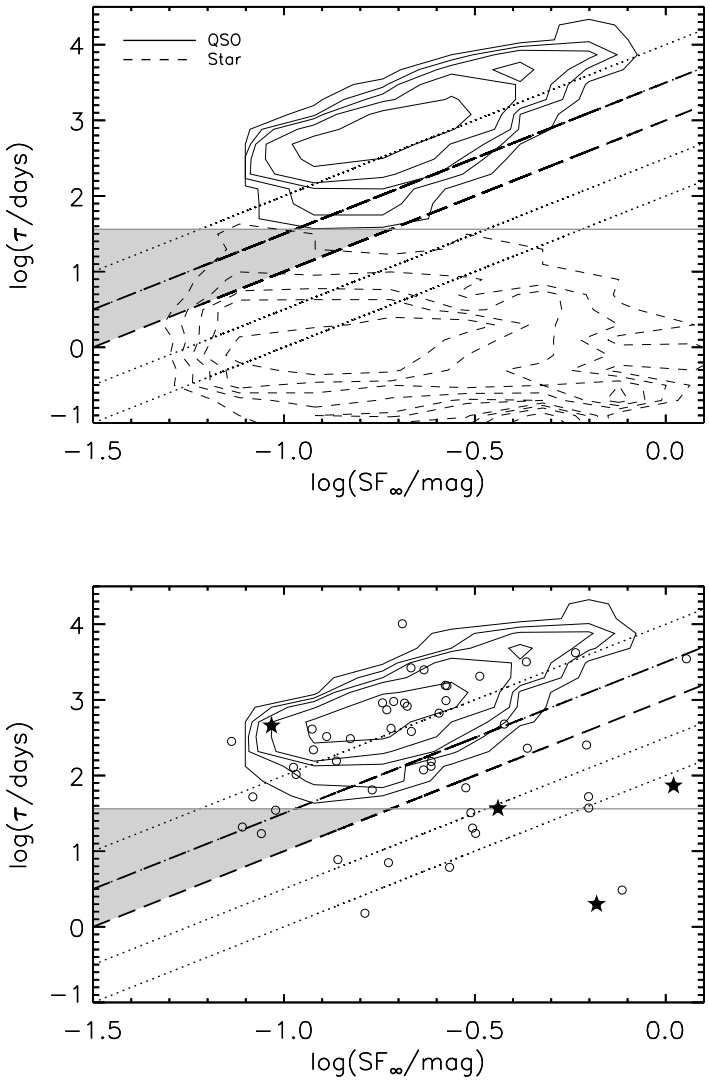

Figure 4.13 *Top:* Characteristic timescale τ versus asymptotic rms variability SF_∞ for confirmed quasars (solid) and stars (dashed) in our S82 data set restricted to $i < 19$. The contours show regions containing 40, 60, 75, 85, and 90% of data points in each subsample. The dotted lines represent lines of constant $\hat{\sigma} = SF_\infty/\sqrt{\tau}$. The two thick dashed lines correspond to $\hat{\sigma} = 10^{-0.22}$ and $10^{-0.47}$ mag yr$^{-1/2}$. The gray region represents the contaminating (stellar) region when selecting sources with $\hat{\sigma} < 10^{-0.22}$ mag yr$^{-1/2}$. When imposing a lower limit at $\tau = 10^{1.56}$ days (gray horizontal line), these contaminants are excluded from the sample, leading to a higher efficiency of quasar selection. *Bottom:* As in top panel but for UVX objects only. Here, the confirmed stars are shown with filled stars, and UVX objects without SDSS spectra are shown with open circles.

Table 4.2. Selection Criteria for $-35^\circ < \mathrm{RA} < -25^\circ$ and $i < 19$

	Selection	N	C	δC	E	δE	Spec.	QSO	Star	Gal.	Unk.
1.	All Sources	2298	100	12	6.5	0.5	20	32	65	2.17	0.22
2.	$\tau >1$ d	991	99	12	15	1	30	50	47	2.72	0.34
3.	$\tau \geq 100$ d*	167	89	11	79	9	83	95	2.2	2.16	0.72
4.	$\tau \geq 100$ d, $\Delta L_{\mathrm{noise}} > 10$*	160	89	11	82	10	85	97	0.74	1.47	0.74
5.	$\tau \geq 100$ d, qso-like colors*	137	86	10	93	11	96	98	0.76	1.53	0.76
6.	$\tau \geq 10^{1.56}$ d, $\hat{\sigma} \leq 10^{-0.18}$ mag yr$^{-1/2}$, $\Delta L_{\mathrm{noise}} > 10$*	179	93	11	78	9	80	97	1.4	1.4	0.70
7.	**UVX:** $\tau \geq 100$ d	127	89	10	94	12	97	98	0.00	1.63	0.81
8.	$\tau \geq 10^{1.56}$ d, $\hat{\sigma} \leq 10^{-0.18}$ mag yr$^{-1/2}$	134	95	10	96	12	97	98	0.00	1.54	0.00
9.	**nUVX:** $\tau \geq 100$ d	10	100	2	50	27	50	100	0.00	0.00	0.00

*The most outlying data point in each light curve has been omitted.

Note. — As in Table 4.1 but for regions of increased stellar density.

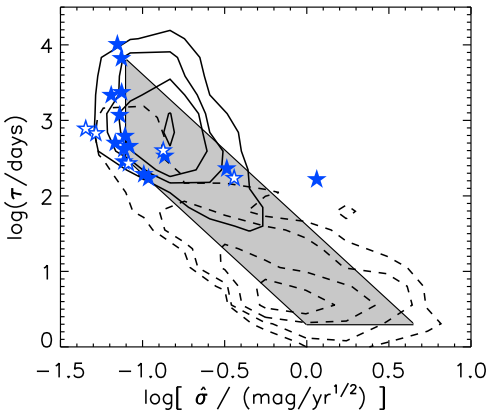

Figure 4.14 Variability timescale τ as a function of short-term driving amplitude $\hat{\sigma}$. The contours show the regions containing 25, 50, 75, and 90% of data points for quasars (solid, $N = 1170$) and all other objects (dashed, $N = 1783$). The shaded region shows the quasar selection cut from Kozł10. The star symbols represent the confirmed stars passing our $\tau > 100$ days and $\Delta L_{\rm noise} > 2$ selection criteria (except those flagged as having outliers in their light curves). The star symbols filled with white indicate late-type stars (SDSS SpecClass= 6).

Comparison to Kozłowski et al. (2010b)

Kozł10 proposed a quasar variability selection regions based on the light curves of OGLE-II and OGLE-III sources in the Magellanic Clouds with the mid-IR colors of AGN from Kozłowski & Kochanek (2009) also with the requirement that $\Delta L_{\rm noise} > 2$. Figure 4.14 shows the distribution of spectroscopically identified quasars and other sources in this selection space as well as the confirmed stars passing the $\tau > 100$ days and $\Delta L_{\rm noise} > 2$ selection cuts. The completeness of this selection region for the SDSS quasar sample is relatively low, $C = 55\%$, as is the efficiency, $E = 46\%$ (62% when excluding outliers in light curves), which demands some explanation. The lower completeness is likely a consequence of the differences in the cadence and duration of the OGLE light curves as compared to the SDSS light curves. The OGLE light curves are densely sampled over the full durations of the surveys, and the typical OGLE-III light curve has 370 epochs over 7 years. As a result, fewer light curves have run-away estimates of τ, making it relatively safe to impose an upper limit on τ in order to reduce contamination by long period variables (LPVs). In the very dense stellar environment of the Magellanic Clouds, the contamination by LPVs

is significant. For the SDSS regions, if we simply eliminate the upper boundary of the Kozł10 selection region, the completeness rises to $C = 91\%$ with a corresponding rise in the efficiency to $E = 52\%$ (69%, excluding outliers) because many more quasars are added than contaminating stars. Figure 4.14 does show significant contamination coming from sources with short timescales, and this was not observed in the Magellanic Cloud sample after removing the periodic variables (predominantly RR Lyrae and Cepheids). Part of the problem may again be the differences in the light curve cadences, where the better cadenced OGLE light curves can better distinguish $\tau \equiv 0$ from a small τ, leading to less "leakage" of false sources past the $\Delta L_{\rm noise} > 2$ boundary. There are probably also some basic differences in the sources of contamination. Kozł10 were designing selection criterion for a small area (1.3 deg^2) containing a large number of luminous stars whose natural timescales are relatively long. Here, we are examining a much wider area (213 deg^2) where the variable stars are generally not as luminous and/or are white dwarfs, leading to relatively shorter natural timescales. Thus, the generous boundary for short τ in Kozł10 may not be appropriate to a wide area survey. If we raise the $\tau > 2$ day lower limit in Kozł10 to $\tau > 10^{1.5}$ days, then for the SDSS region the completeness changes little ($C = 90\%$), but the efficiency rises dramatically to $E = 78\%$ (86%, excluding outliers). Finally, if we consider the variability properties of the variable spectroscopically identified stars in this diagram, they tend to lie in the region of low amplitudes $\hat{\sigma}$ and long timescales. In the Magellanic Cloud regions, where Kozł10 could take advantage of color magnitude diagrams to better characterize the variable stars, these tend to be various classes of evolved, red stars – long period variables (LPV), long secondary period (LSP) variables and OGLE Small Amplitude Variable Red Giants (OSARGS).

Comparison to Butler & Bloom (2010)

Butler & Bloom (2010, hereafter B&B) present a different approach to using the DRW model to select quasars based on variability. They model the ensemble SF for quasars (see, e.g., Vanden Berk et al. 2004) as a function of apparent magnitude with the DRW model, and use this to predict τ and $\hat{\sigma}$ for individual quasar candidates. The process model χ^2

is then computed for each light curve using these *fixed* DRW parameters to obtain $\chi^2_{\rm QSO}$. Essentially, they fit the DRW model with a (infinitely) strong prior on the values of τ and $\hat{\sigma}$ derived from the apparent magnitude. They also compute $\chi^2_{\rm FALSE}$, the goodness of fit expected if the same DRW model was fit to a light curve that was really white noise with the overall variance of the observed light curve. The main difference with respect to our method is that we obtain individual estimates of τ and $\hat{\sigma}$ for every light curve independent of the mean magnitude. This leads to a much wider distribution of τ and $\hat{\sigma}$ for our method with respect to B&B, allowing us to use τ as an efficient classifier. The selection technique used in B&B is very efficient, however, because only a single model rather than a parameter fit is required.

Here we compare the resulting C and E from each method using identical samples of light curves. In Table 4.1 we compare the two methods for the light curve sample with $-35^\circ < {\rm RA} < 50^\circ$ and $i < 19$, where the B&B method selects an object as a quasar candidate if

- $\chi^2_{\rm FALSE}/\nu > \chi^2_{\rm QSO}/\nu$
- $\chi^2_{\rm FALSE}/\nu > 3.1$ if RA > -20 deg (5σ significance)
- $\chi^2_{\rm FALSE}/\nu > 4.4$ if RA < -20 deg (7σ significance),

where ν is the number of degrees of freedom. Comparing the fourth and eleventh rows, one can see that C is higher by a few percent for the B&B method whereas E is slightly lower. When instead choosing $\tau > 10^{1.5}$ days and $\Delta L_{\rm noise} > 10$, a similar completeness is achieved (96%) with $E = 87\%$. Therefore, with some fine-tuning of parameters, the two methods can deliver samples with comparable C and E, which is not too surprising since they rely on the same mathematical model to describe the light curve behavior. The B&B method is faster if the only goal is to select quasars, but fitting for τ and $\hat{\sigma}$ then allows for further classification of the type of variability, as illustrated by the locations of different classes of variable stars in the parameter space (see Figure 9 in Kozł10).

In Table 4.3, we present the cross–correlation between our method and that in B&B in terms of how many $i < 19$ sources are selected using either technique. For this table,

Table 4.3. Cross–Correlation with B&B for $-35^\circ < \mathrm{RA} < 50^\circ$ and $i < 19$

	Selected By	N	C	δC	E	δE	Spec.	QSO	Star	Gal.	Unk.
1.	All Sources	10024	100	4	15	0.4	33	45	53	2.5	0.15
2.	Both Methods	1564	95	3	90	3	92	98	1.3	0.35	0.14
3.	Either Method	1763	97	4	82	3	85	96	3.1	0.60	0.13
4.	Neither Method	8237	2.6	0.4	0.5	0.1	22	2.1	94	4.0	0.11
5.	Only our Method	83	0.9	0.3	17	5	37	45	39	13	3.2
6.	Only B&B	138	1.5	0.3	16	4	33	49	49	2.2	2.2

Note. — Column definitions follow those in Table 4.1. The second row lists results for a sample satisfying both the B&B criteria as well as our criteria ($\tau \geq 10^{1.5}$ days and $\Delta L_{\mathrm{noise}} > 10$, after omitting light curve outliers). The third row lists results for a sample satisfying either the B&B criteria or ours. The selection in the fourth row is rejected by both methods, while the selections in the fifth and sixth rows only satisfy one of the methods.

we define our selection criteria as $\tau \geq 10^{1.5}$ days and $\Delta L_{\mathrm{noise}} > 10$ (after omitting light curve outliers). We define the B&B selection criteria as listed above. We then show the characteristics of the samples defined by passing both criteria, at least one of the criteria, failing both, and passing only one or the other. It is beneficial to use both methods to select quasars. One can achieve $E = 90\%$ in the absence of color selection when requiring both criteria, but $C = 97\%$ can be maintained when using either criteria for an E of 82%. In essence, some quasars have true τ and $\hat{\sigma}$ values too far from the prior defined by the apparent magnitude and so have reduced B&B likelihoods, while for some other objects, the addition of the strong B&B prior leads to better rejection of false positives and negatives.

4.6 Effect of Cadence and Photometric Accuracy on Best-fit Variability Parameters: Prospects for LSST and Pan-STARRS

One clear lesson from our comparison of our proposed SDSS selection criteria (applied to sparsely sampled, effectively shorter duration light curves in a wide area, extragalactic field)

and the Kozł10 selection criteria (applied to densely sampled, longer duration light curves in a small area with high stellar densities) is that the appropriate selection criteria will depend on both the properties of the light curves (duration, cadencing, and uncertainties) and the nature of the contamination. Here we consider the effects that the survey length, time sampling, and photometric errors have on the best-fit DRW parameters. These are important issues to consider for upcoming surveys such as Pan-STARRS (Kaiser et al. 2002), which is already taking data, and LSST (Ivezić et al. 2008), which will obtain accurate, well-sampled light curves for millions of quasars. To probe the time scales as short as 0.1τ, and assuming a rest-frame timescale of $\tau = 10$ days and a characteristic redshift of 2, the light curves should be sampled every 3 days in the observer's frame, which is in good agreement with the baseline cadence of LSST. The LSST photometric errors in the r band will be < 0.02 mag for $r < 22$, and there are roughly 2-3 million AGN with $r < 22$ in the 20,000 sq. deg. covered by the main LSST survey (see Table 10.2 in the LSST Science Book; LSST Science Collaborations and LSST Project 2009). Each of these objects will be observed about 1000 times (summed over all bands), yielding a database of over 2 billion photometric measurements. The characteristic timescale τ will be well-constrained if the survey length is sufficiently longer than τ. It is therefore necessary to evaluate the accuracy of fitted variability parameters as a function of light curve length.

4.6.1 The Impact of Light Curve Length

As a first test, we generate 10,000 DRW light curves using an input of $\tau = 575$ days and $\mathrm{SF}_\infty = 0.2$ mag, which are the median observed-frame values for the S82 quasar sample. The light curves are sampled once every 10 days ($\Delta t = 10$ days) and span a length of 40 years. Photometric errors are drawn from a Gaussian distribution with a standard deviation of $\sigma = 0.01$ mag. To test the impact of light curve length, we truncate the light curves to 10 years, 3 years, and 1 year and compare the resulting variability parameters. For data lengths shorter than τ, only the driving amplitude of short-term fluctuations, $\hat{\sigma} = \mathrm{SF}_\infty/\sqrt{\tau}$, is well constrained [recall that $SF(\Delta t << \tau) = \hat{\sigma}\sqrt{|\Delta t|}$]. That is, on a grid of $\log(\tau)$ versus $\log(\mathrm{SF}_\infty)$, the best-fit values will be scattered due to fitting errors along lines of constant

$\hat{\sigma}$, but much less perpendicular to it (i.e., along lines of constant $K = \tau\sqrt{\mathrm{SF}_\infty}$). Therefore, as the light curve length decreases, the mean best-fit $\hat{\sigma}$ will not vary significantly, while the best-fit parameters τ and SF_∞, and K will become biased (see Kozł10; MacLeod et al. 2010). Note that only two of these four parameters are independent.

Figure 4.15 shows the distribution of best-fit variability parameters $\hat{\sigma}$, K, τ, and SF_∞ normalized by the input values for simulated light curves with length 40 years, as well as when truncated to 10 years, 3 years, and 1 year. As the length increases, the widths of the output parameter distributions decrease, roughly as the square root of the light curve length. While the mean output value of $\hat{\sigma}$ is unaffected by the light curve length, the mean values of K, τ, and SF_∞ decrease as the length shortens, and the number of "run-away" time scales ($\tau \geq 10^5$ days) increases due to the fact that τ is now longer than the light curve length and cannot be constrained. We find that for typical variable quasars, the best-fit distributions of τ and SF_∞ are strongly biased for surveys significantly shorter than 10 years (by about 70% and 40%, respectively, for a length of 3 years). Nevertheless, the $\hat{\sigma}$ parameter is unbiased and already well-measured (rms = 20%) after 1 year.

4.6.2 The Impact of Realistic Cadence

As a second test, we generate 10,000 light curves using a simulated r-band LSST cadence (∼200 observations spread over 10 years; Delgado et al. 2006). An example light curve with this cadence is shown in Figure 4.16. Figure 4.17 compares the resulting parameters for the LSST cadence and a cadence of $\Delta t = 10$ days over 10 years as in the first test. The only significant change is an increased width for the best-fit $\hat{\sigma}$ distribution, where the accuracy of the best-fit $\hat{\sigma}$ drops from 5% to 13%.

4.6.3 The Impact of Photometric Accuracy

As a third test, we evaluate the effect that the photometric accuracy has on the fitted parameters by generating 10,000 light curves with the r-band LSST cadence and photometric errors drawn from a Gaussian of width $\sigma = 0.03$ mag. The same is done for $\sigma = 0.1$ mag. The output parameters for these two sets of light curves are compared to those from the

second test, where $\sigma = 0.01$ mag, in Figure 4.18. As expected, the distributions become broader as the photometric error bars increase and the fits become more uncertain. While a photometric accuracy of $\sigma = 0.1$ mag leads to a significant increase in the $\hat{\sigma}$ errors compared to the $\sigma = 0.01$ mag case (a factor of 4 dex), an accuracy of $\sigma = 0.03$ mag leads to only a marginally wider distribution (40%). We conclude that a photometric accuracy of $\sigma = 0.03$ mag is sufficient to obtain well-constrained variability parameters for typical quasars.

4.6.4 Selection Completeness for a Realistic LSST Simulation

As our final test, we estimate the completeness of quasar selection based on the $\tau > 100$ days criterion alone for three different light curve lengths (1, 3, and 10 years), the r-band LSST cadence, and a photometric accuracy of $\sigma = 0.03$ mag (achieved at $i \approx 22$). Now, instead of using fixed input τ and SF_∞ values, we use the full observed r-band distributions for the S82 quasar sample as presented in MacLeod et al. (2010). The distribution of input τ is shown as the gray histogram in the top panel of Figure 4.19. The output distributions for each light curve length are also shown. We compute the completeness as the fraction of the output distribution with a best-fit τ exceeding 100 days. The completeness for the input distribution is 91% (different than the 94% reported in Table 4.1 because all S82 quasars are considered, without rms > 0.05 or $i < 19$ mag limits). For light curve lengths of 10, 3, and 1 year(s), C is estimated to be 88%, 75%, and 51%, respectively. Notably, with 10 years of observations, imperfectly measured τ causes only $\sim$3% of the sample to "leak" below the $\tau = 100$ day boundary. Because it is not known what the contamination will be like for a typical LSST field, we cannot estimate the efficiency for the $\tau > 100$ days selection. With the aid of multi-band imaging, colors can be used along with a relaxed τ cut to obtain highly complete samples early in the survey ($\sim$1 year).

While the light curve length clearly has an impact on the best-fit timescale distribution, we also wish to estimate the effect of different cadences in general. This is motivated by the fact that Kozł10 found a large fraction of quasars concentrated at smaller τ when applying the DRW model to OGLE-II and OGLE-III light curves, which are more densely-sampled than the SDSS S82 data. Also, the Pan-STARRS1 3π survey will obtain 6 epochs in a given

band over 3 years, as opposed to 60 epochs for LSST. Therefore, we wish to determine whether a τ selection can still deliver decently complete samples of quasars in a typical 3 year-long Pan-STARRS survey. We use Table 4 in Schm10 to simulate a typical Pan-STARRS cadence and assume that the *griz* bands can be combined to give 24 total epochs over 3 years (we omit the Y band because of its lower sensitivity).

To make the comparison, we use a uniform input distribution of $\log(\tau)$ and $\log(\mathrm{SF}_\infty)$ within the region defined by $2 < \log(\tau/\mathrm{days}) < 3.5$ and $-1.4 < \log(\mathrm{SF}_\infty/\mathrm{mag}) < -0.2$. We do not use the observed τ distribution for SDSS quasars for the input, as in the previous test, in order to ignore any potential effects of a biased τ distribution caused by the SDSS sampling. We compare the fits at a fixed photometric uncertainty (0.03 mag), therefore assuming that the absolute depth of each survey does not play a role in these comparisons[4]. In the middle and bottom panels of Figure 4.19, we compare the completeness results for the proposed LSST samplings to the typical 10 year-long S82, 7 year-long OGLE-III, and 3 year-long (24-epoch) Pan-STARRS samplings. For each cadence, we compute the completeness as the fraction of best-fit τ that does not "leak" below the $\tau = 100$ day boundary. The results show that the denser time sampling of the OGLE-III survey leads to a higher completeness and a more compact spread in output τ when compared to S82, and a similar completeness when compared to the 10-year r-band LSST cadence ($\sim$93%). We note that the timescales for the minority of OGLE-III objects that also have OGLE-II light curves will be even better constrained, since for these the total duration is $\sim$10 years. For Pan-STARRS1 3π, C is estimated to be 70% for $\tau > 100$ days after combining the four *griz* bands. This is comparable to the result for a 3-year r-band LSST cadence. For a single Pan-STARRS band (6 epochs over 3 years), C will likely be significantly worse, so in this case an alternate selection method (such as one based on $\hat{\sigma}$) may be optimal in the absence of color selection.

[4]Because the LSST survey has a fainter magnitude limit than OGLE, SDSS, or Pan-STARRS, it will have a higher completeness for quasar selection. Since we wish to compare differences due to cadencing only, we compare at a fixed photometric uncertainty. We note that a fixed uncertainty will induce some differences in the τ distribution between surveys from what would realistically be observed. However, that is a higher order question than what is examined here.

Comparison to Butler & Bloom (2010)

We carried out similar simulations for the B&B method, using the same uniform input distribution of $\log(\tau)$ and $\log(\mathrm{SF}_\infty)$. Based on the mean trend of $\hat{\sigma}^2$ with g magnitude found for the S82 quasar sample in MacLeod et al. (2010), we assigned each light curve an apparent magnitude of

$$g = 0.334 \log \hat{\sigma}^2 + 21.38 + G(0.8), \tag{4.6}$$

where $G(0.8)$ is a Gaussian random deviate of width 0.8 mag, matching the observed scatter about the mean trend.

Figure 4.20 shows the output distributions of $\chi^2_{\mathrm{FALSE}}/\nu$ and $\chi^2_{\mathrm{QSO}}/\nu$, as well as the B&B selection limits for a 5σ significance cut against false positives. The figure legends list the cadences and the resulting completeness C. As survey length increases, C increases because of an overall increase in $\chi^2_{\mathrm{FALSE}}/\nu$. For the single-year LSST cadence, C is only 20%, and we find a similar result for the 3-year single-band Pan-STARRS 3π cadence. When combining the Pan-STARRS *griz* bands, C increases to 52%, slightly lower than the 3-year LSST cadence. The LSST 10-year completeness is lower than what was found when choosing $\tau \geq 100$ days in the previous section (86% vs. 92%), but it could be increased by weakening the cut on $\chi^2_{\mathrm{FALSE}}/\nu$. However, because the contamination rate is not known for LSST, it is unclear as to how much this cut could be varied without a major decrease in E. Again, the OGLE cadence leads to the highest C due to the large number of epochs. We also find a significant decrease in C for the S82 cadence from what is observed in the eleventh row of Table 4.1 (81% vs. 96%). This may be due to the prescription for generating mean magnitudes. A larger C is expected if we were to use the scaling in Table 1 in B&B, because the $\chi^2_{\mathrm{FALSE}}/\nu$ would likely be larger overall (i.e., white noise would be less likely).

4.7 Discussion and Summary

We used the light curves for variable point sources from SDSS Stripe 82 to determine how well quasars can be separated from stars with the aid of photometric variability. The Stripe 82 data set, designed to search for supernovae, is excellent for quasar variability studies since it provides light curves for a large statistical sample of objects to a faint limiting

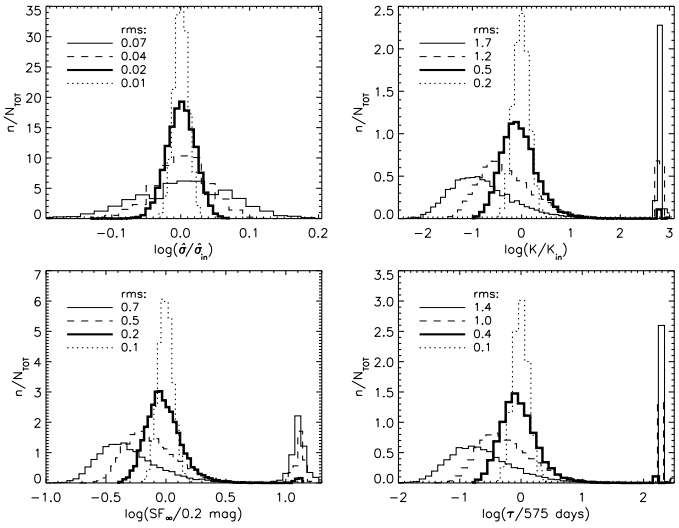

Figure 4.15 Distributions of $\hat{\sigma} = \mathrm{SF}_\infty/\sqrt{\tau}$, $K = \tau\sqrt{\mathrm{SF}_\infty}$, τ, and SF_∞ normalized by the input values for 10,000 simulated light curves with length 40 years (dotted), 10 years (thick), 3 years (dashed), and 1 year (thin solid), all using $\Delta t = 10$ days and photometric accuracy of 0.01 mag. n indicates the number of points in a bin divided by the bin width, and N_{TOT} is the total number of points used for each histogram. The rms of each distribution is listed in the legend.

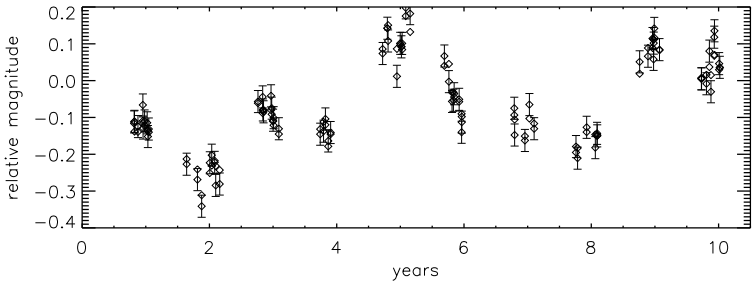

Figure 4.16 A model light curve generated for $\tau = 575$ days, $\mathrm{SF}_\infty = 0.2$ mag, the r-band LSST cadence, and a photometric accuracy of 0.03 mag.

magnitude (see Boutsia et al. 2010 for another example of using supernova surveys to search for quasars). The data set used here contains precise photometry (using recalibrated light curves by Ivezić et al. 2007 and Sesar et al. 2007), as well as extensive spectroscopy, which has enabled us to develop an efficient method for selecting quasars while maintaining high completeness. The variability model used here (a damped random walk, see KBS09; Kozł10) is robust enough to handle sparse or irregularly-sampled data. By modeling the variability as a stochastic process described by the exponential covariance matrix $S_{ij} =$

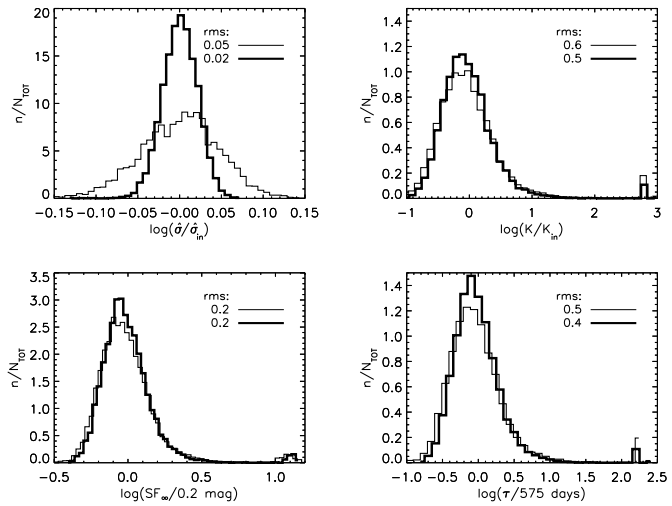

Figure 4.17 Distributions of $\hat{\sigma} = \mathrm{SF}_\infty/\sqrt{\tau}$, $K = \tau\sqrt{\mathrm{SF}_\infty}$, τ, and SF_∞ normalized by the input values for 10,000 simulated light curves with a photometric accuracy of 0.01 mag and the r-band LSST cadence with $\sim$200 data points (thin solid). The thick line shows the distribution from Figure 4.15 with a cadence of $\Delta t = 10$ days over 10 years (365 data points). The rms of each distribution is listed in the legend.

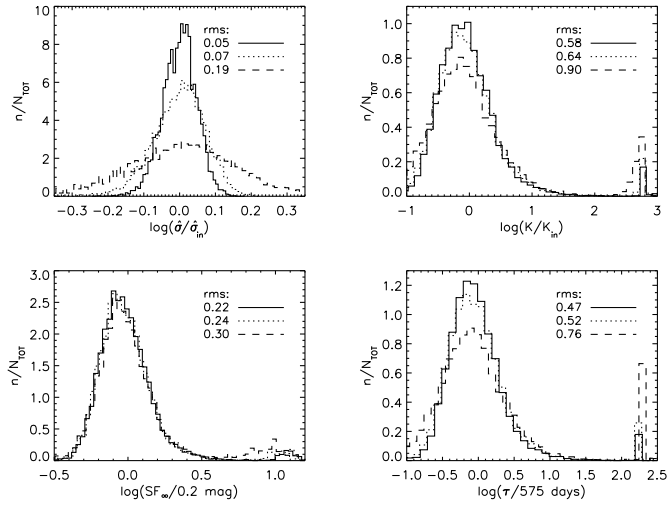

Figure 4.18 Distributions of $\hat{\sigma} = \mathrm{SF}_\infty/\sqrt{\tau}$, $K = \tau\sqrt{\mathrm{SF}_\infty}$, τ, and SF_∞ normalized by the input values for 10,000 simulated light curves with a 10-year r-band LSST cadence and a photometric accuracy of 0.01 mag (solid), 0.03 mag (dotted), and 0.1 mag (dashed). The rms of each distribution is listed in the legend.

$\sigma^2 \exp(-|t_i - t_j|/\tau)$, only O($N$) operations are required to determine the model parameters for a light curve with N data points. This is an important algorithmic feature in the context of upcoming massive synoptic surveys, such as LSST.

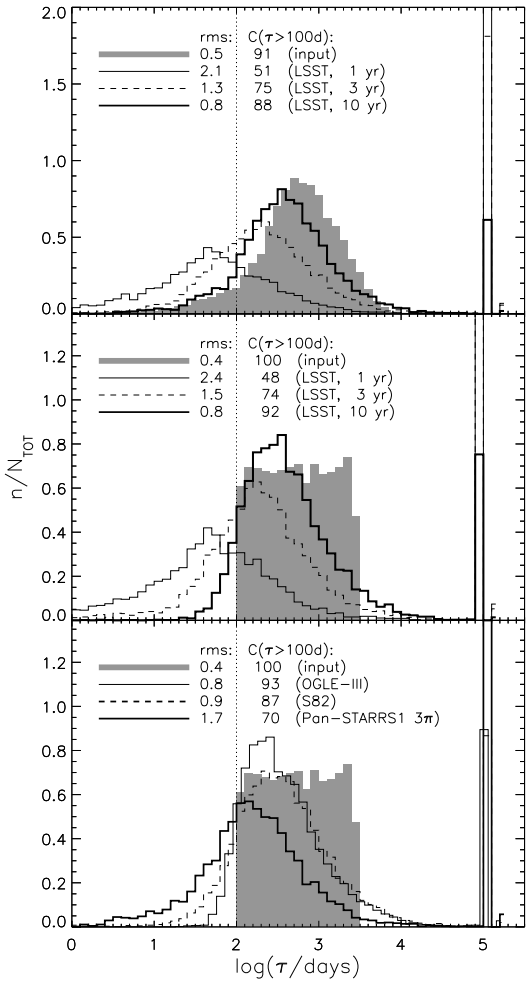

Figure 4.19 *Top:* Distributions of τ as compared to the input distribution (shaded histogram) for ∼7,000 simulated light curves with the r-band LSST cadence, a photometric accuracy of 0.03 mag, and lengths of 1 year (thin solid), 3 years (dashed), and 10 years (thick solid). The rms of each distribution is listed in the legend, as well as the completeness of the quasar selection criteria $\tau \geq 100$ days (dotted line). Note that light curves significantly longer than 1 year are required to achieve $> 50\%$ completeness. *Middle:* As in top panel but for a uniform input distribution of τ within the range $100 < \tau < 10^{3.5}$ days. *Bottom:* As in middle panel but for OGLE-III (thin solid), S82 (dashed), and Pan-STARRS1 3π (thick solid) cadences, which span 7, 10, and 3 years, respectively. The denser time sampling of the OGLE-III survey leads to a higher completeness than S82.

The DRW model provides unique information on the characteristic timescale, τ, for which a range has been observed for quasars (Sesar et al. 2006; KBS09; Kozł10; MacLeod et al. 2010). Compared to selecting solely on the short-term behavior of the SF (the parameter $\hat{\sigma}$), the inclusion of τ information boosts C from 90% to 97% while maintaining an efficiency

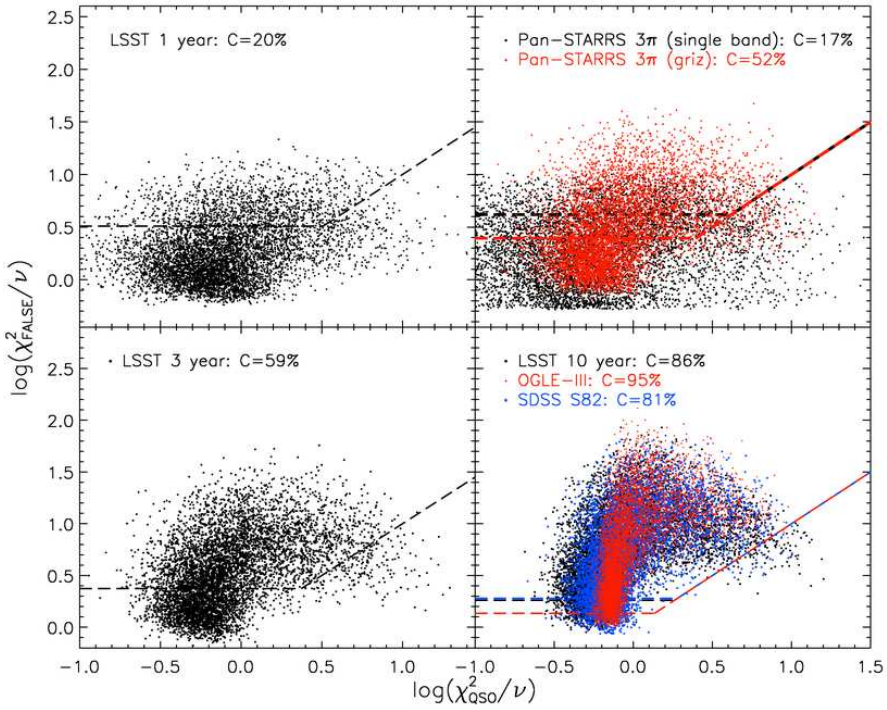

Figure 4.20 Distributions of $\chi^2_{\rm FALSE}/\nu$ and $\chi^2_{\rm QSO}/\nu$ resulting from fitting ∼7,000 simulated DRW light curves with various cadences (listed in the legends) using the B&B technique. The completeness of the quasar selection criteria (the fraction above the dashed lines) for each cadence is listed.

(purity, E) at 80%, based on variability alone. Conversely, for $C = 98\%$, the τ information boosts E from about 60% to 75%, while at $C = 90\%$, E rises from about 80% to 85%. This method enables the selection of $z \sim 3$ quasars, where color selection largely fails. With the aid of $u - g$ and $g - r$ color selection, the efficiency can be further boosted to 96% with a reasonably high completeness of 93%. It is worth reiterating that the E determined here is only a lower limit because the spectroscopically confirmed quasar sample is not 100% complete. This performance is accomplished simply by imposing a lower limit on τ and an upper limit on $\hat{\sigma}$. The separation might improve when using more advanced techniques such as kernel density estimation (e.g., Richards 2008) rather than imposing simple limits.

By using the parametrization $SF(\Delta t) = A(\Delta t/1{\rm yr})^{\gamma}$, Schm10 was able to select quasars from UVX objects with $E \simeq 95\%$ at 90% completeness when selecting on A and γ. This selection is similar to one using $\hat{\sigma}$ alone, and provides a reasonable method for selecting

quasars from short ($\lesssim$ 1 year) light curves. With longer light curves, the inclusion of τ information significantly improves the performance. The differences in C and E are summarized in the bottom panels of Figure 4.12 and illustrated in Figure 4.13. If we restrict our initial sample to UVX objects and simply use $\tau \geq 100$ days to select quasars, we find that C improves from the 90% found by Schm10 to 95%, while E improves from about 95% to 97%. When combining the selection criteria $\tau \geq 10^{1.6}$ days and $\hat{\sigma} \leq 10^{-0.2}$ mag yr$^{-1/2}$, we find that C improves further to 98% with $E = 97\%$. We note that the DRW model has benefits beyond the improved selection efficiency compared to Schm10. In addition to being a faster algorithm [O(N) rather than O(N^2)], the DRW model also produces information on the physics of the accretion disk since τ and SF$_\infty$ are physical parameters that show correlations with luminosity, black hole mass, and wavelength (see MacLeod et al. 2010) and provides an approach for classifying other aperiodic sources as well (see Kozł10). The latter two points are also advantages of our approach over that in B&B. For quasar selection, both methods yield similar results for completeness and efficiency.

Our selection criteria identify 255 objects with $i < 19$ and $-35^\circ < \mathrm{RA} < 50^\circ$ that vary on long time scales ($\tau > 100$ days with $\Delta L_{\mathrm{noise}} > 2$) and do not presently have SDSS spectra; the majority of them are likely quasars. Based on their light curves, several are expected to be interesting long-timescale variable objects, such as AGB stars. We target these objects for follow-up spectroscopy with the goal of either confirming their quasar nature or identifying them as major contaminants in quasar selection based on variability.

The DRW provides a complete statistical model for quasar variability. This is especially useful for evaluating the success of quasar selection in upcoming surveys, and enables us to estimate the completeness for different survey lengths, cadences, and photometric accuracy. We cannot estimate E, however, since it is dependent on an unknown makeup of the faint variable optical sky. For a typical (simulated) LSST cadence over 10 years and a photometric accuracy of 0.03 mag (achieved at $i \approx 22$), C is expected to be 88% for a sample selection criterion of $\tau > 100$ days. For typical variable quasars, the best-fit distributions of τ and SF$_\infty$ are biased for surveys shorter than 10 years (by about 70% and 40%, respectively, for a length of 3 years). However, $\hat{\sigma}$ is well-constrained, and with the aid of multi-band imaging, colors can be used along with a relaxed τ cut to obtain highly complete samples

early in the LSST survey ($\sim$1 year). For the Pan-STARRS1 3π survey, C is estimated to be 70% for a $\tau > 100$ day selection after combining the *griz* bands, which is comparable to the result for a 3-year r-band LSST cadence. The dense cadence of the LSST will provide tighter constraints on τ for a survey length of 10 years. This improvement is also important for relating observations to the physics of accretion disks (see MacLeod et al. 2010, and references therein). The most important conclusion of this work is that, given an adequate survey cadence, photometric variability provides an even better method than color selection for separating quasars from stars.

4.8 Appendix: Spectroscopic Follow-up with Apache Point Observatory DIS

Recent sky surveys have identified quasars using spectroscopy, optical color selection (Richards et al. 2009), X-ray detections (Young et al., 2012), mid-IR colors (Kozłowski et al., 2012), and optical variability (Kozłowski et al., 2010b; Ross et al., 2012), among other methods. Using color-based selection, the SDSS has provided a spectroscopically complete survey to $i < 19$ for the quasar region in SDSS color space, providing large quasar samples which are valuable in many studies, including the construction of luminosity functions (Croom et al., 2009a) and composite spectra (Vanden Berk et al., 2001) as a function of redshift. However, the color-based selection method misses many quasars at intermediate redshift, with a completeness of $C \sim 42\%$ at $2.5 < z < 3$ (whereas $C \sim 94\%$ for $z < 2.5$), where the optical colors resemble those of stars (B&B). The SDSS-III Baryon Oscillation Spectroscopic Survey (BOSS) has now confirmed 100,000 quasars at high enough redshift for studies of the Lyman-α forest. The upcoming large-area sky surveys such as Pan-STARRS and LSST plan to use their time-domain information rather than spectroscopy to separate quasars from stars, making it important to explore and optimize selection methods using current samples.

The S82 data set, designed to search for supernovae, is excellent for quasar variability studies since it provides light curves for a large statistical sample of objects to a faint limiting magnitude. The application of a DRW model to the light curves of ∼10,000 variable point sources with $i < 19$ from SDSS S82 yielded 255 quasar candidates based on their optical variability that lacked SDSS spectra (Table 1 in MacLeod et al. 2011b). In order to estimate the completeness of SDSS spectroscopic quasars, and to test DRW variability as a selection method, many of these candidates were targeted for follow-up spectroscopy using either the 2.5m or 3.5m telescope at Apache Point Observatory (APO). A portion of these objects were targeted as part of a BOSS ancillary program. Using the priority system described below, we obtained spectra using the Dual Imaging Spectrograph (DIS) on the ARC 3.5m at APO for a subsample which were either outside the R.A. range, too bright ($i < 16.2$), or too red for the BOSS ancillary program. The results of both follow-up programs for our 255 quasar candidates are presented in Section 4.8.4.

4.8.1 Target Selection Based on Variability

We began by selecting all S82 variable point sources with $i < 19$ in the RA range $-35^\circ < \alpha < 50^\circ$ that have $\tau \geq 100$ days and $\Delta L_{noise} > 2$ ($\sim$1700 sources total). Of these, 1300 were already spectroscopically confirmed as quasars and 255 lacked SDSS spectra. However, visual inspection revealed a single outlier at a significantly fainter magnitude in many of their light curves which resulted in a large best-fit τ, especially for objects in the stellar locus on a $(u-g,\ g-r)$ diagram. When rejecting the most outlying data point in each light curve and then refitting the models, 30% of the 255 (41% of the stellar locus objects) drop out of the $\tau \geq 100$ days and $\Delta L_{noise} > 2$ selection. Therefore, we excluded this 30% from the target selection. We also excluded another 45% which were already targeted for spectroscopic follow-up in an ancillary program for SDSS-III BOSS. The remaining 55% were not targeted by BOSS due to a limit in either RA, -20 deg $<$ RA $<$ 45 deg, or in i magnitude, $i > 16.2$. This resulted in 98 targets for follow-up with DIS.

4.8.2 Selection Based on 2MASS, GALEX, X-ray, and Radio Data

We then subdivided these 98 targets into three groups, assigning higher priority in our observation queue to those candidates most likely to be quasars based on optical SDSS colors and other publicly available multiwavelength data. In the optical, we divide the $(u-g)$ vs. $(g-r)$ plane into five regions (see Fig. 4.21). Quasars generally fall within regions II and IV, while region I contains the stellar locus, and region III denotes RR Lyrae colors. The I_{off} region contains objects that seem consistent with reddened quasars, as their light curves resemble those of quasars, and they are rather faint and clustered at redder $(r-i,\ i-z)$ colors. We expect that 2–5% of quasars are missed by the SDSS color selection due to intrinsic reddening. Our 98 variability-selected candidates are spread throughout all five regions.

Were we to rely exclusively on optical data, priority would be assigned, in decreasing order, to regions II and IV, region I_{off}, region III, and lastly region I. However, we have incorporated multiwavelength observations (IR, UV, X-ray and radio) to identify targets that have SEDs suggestive of quasar emission. In particular, we produced a list (Table 4.4)

of those objects among the 98 targets possessing one or more of the following:

1. A 2MASS match (to within $1.5''$) with IR quasar colors $J - K > 1$
2. A ROSAT X-ray counterpart within $30''$ (BSC and FSC; Schneider et al. 2010)
3. A GALEX GR6 counterpart (to within $3''$; Agüeros et al. 2005) with UV-optical colors satisfying either or both of:
   (a) $n - u < 1.5$, $u - g < 0.8$ (n is near-UV magnitude, see Fig. 2a)
   (b) $f - n > 0$, $g - r < 0.5$ (f is far-UV magnitude, see Fig. 2b)
4. A FIRST radio counterpart within $2''$ (Kimball & Ivezić, 2008)
5. A NVSS radio counterpart within $7''$ (Kimball & Ivezić, 2008)

Of our 98 targets, 23 satisfied one or more of conditions 1–5; we assigned highest priority to these promising quasar candidates. Table 4.5 lists how many of the 98 targets in each optical color region satisfy the above conditions. Three of these (ID 1900925, 1122082, 1690257) dropped out of a $\tau > 10^{1.5}$ days and $\Delta L_{noise} > 2$ selection when excluding light curve outliers (including a flare for ID 1900925). However, since each of these three satisfies one or more of the conditions above, we include them in our top priority list, as they may be interesting objects. Among the candidates which fail the above criteria, second priority was assigned to objects with similar colors to quasars in the optical (regions II and IV), and lowest priority was assigned to those on the stellar locus.

4.8.3 Data and Analysis

In Autumn (Quarter 4) 2010, we were awarded two half-nights, one bright and one dark (September 26, B-half, and December 11, A-half, respectively), for spectroscopic follow-up using APO DIS. We were able to obtain 44 spectra and confirm several as quasars despite the small amount of dark time awarded. Twenty-five spectra were obtained during bright time, including the brightest high-priority and many of the lower-priority targets. We used exposure times that should yield spectra with signal-to-noise ratios greater than 7. If we were interested in obtaining high-resolution spectra to perform detailed classifications

Table 4.4. Highest Priority Targets

ID	α (°)	δ (°)	i	Color	Criteria	t_{obs} (min)	Class	Redshift
433936	349.534760	0.197694	18.73	II	3 (a)	23.2	QSO	1.24
493214	348.344788	−1.052154	18.04	II	3 (a)	12.3	QSO	1.79
744044	343.897919	0.220414	18.51	II	3 (a), 3 (b)	18.9	QSO	1.90
1070095	336.462799	1.264313	17.74	II	1	9.3	N/A	-
1077378	339.876740	−1.260627	18.10	II	1, 3 (b)	13.0	QSO?	-
1122082	336.992218	0.618182	17.99	II	1, 4, 5	11.8	BL Lac[a]	-
1165315	338.794312	0.274968	18.09	I_{off}	1	12.9	QSO[a]	0.32
1169739	338.322815	−0.633036	17.99	II	3 (a), 3 (b)	11.7	QSO	1.27
1403402	334.656219	−0.292136	17.97	I_{off}	1	11.5	QSO[a]	0.44
1424605	331.310760	0.146080	15.26	I	1	2.0	AGB	-
1431246	333.090973	0.723634	18.70	II	3 (a), 3 (b)	22.5	QSO?[a]	-
1437582	334.604340	0.407353	18.46	II	3 (a)	18.1	QSO[a]	0.91
1496900	330.369263	0.786810	15.98	I	3 (b)	2.0	STAR (G)	-
1545011	334.009674	0.974059	17.68	III	1	8.8	QSO[a]	2.86
1690257	34.692432	−1.049126	15.69	I	5	2.0	STAR (K)	-
1743419	325.969055	−0.254378	17.32	II	3 (a), 3 (b)	6.3	QSO	1.52
1850942	328.062378	−0.864421	17.88	II	1, 3 (b)	10.6	QSO?	-
1900925	328.822662	−0.763361	13.29	I	2	2.0	STAR (M)	-
2002954	36.638287	−0.888411	13.52	I	1	2.0	STAR (M)	-
2072959	35.382450	0.663692	18.66	I_{off}	1	21.7	QSO?	-
2533277	44.002426	−1.262077	18.17	II	3 (a)	13.8	QSO (+CV?)	1.25
3104096	45.299786	−0.553857	16.80	I	3 (b)	3.9	SNe[a]	0.0618
3995101	357.069672	1.095822	17.87	I	1	10.5	N/A	-

[a]No DIS spectrum; classification found elsewhere (either in SDSS BOSS or SIMBAD catalogues).

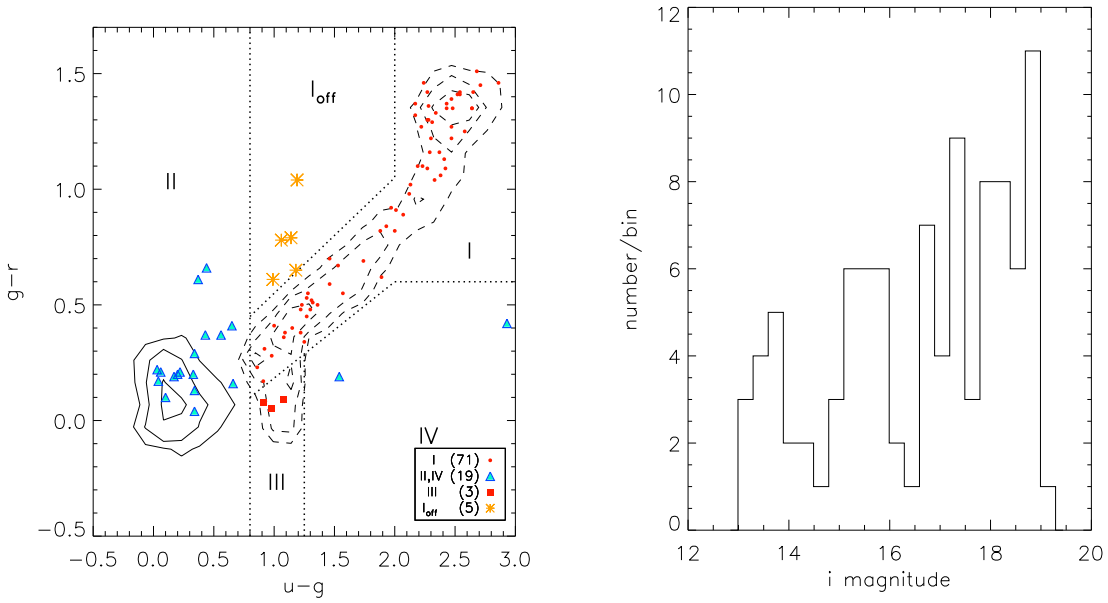

Figure 4.21 Left: The solid contours show the distribution in $u-g$ versus $g-r$ colors of variable point sources spectroscopically confirmed as quasars, while the dashed contours show that for spectroscopically confirmed stars. The dotted lines divide the color space into four characteristic regions: the stellar locus (I), the low-redshift ($z \lesssim 2$) quasar region (II), the RR Lyrae region (III), the high-redshift quasar region (IV), and the region above the stellar locus (I_{off}). All 98 targets are overplotted. Right: Distribution of i magnitudes for the 98 targets.

such as stellar subtypes, we would have used an order of magnitude longer exposure times. However, we used the minimum exposure times necessary to resolve the continuum and any broad emission lines, since we are only interested in identification. However, we did enforce a lower limit of 2 minutes for the exposure time in order to account for slew time.

The spectra were then reduced and calibrated, using the standard star G191-B2B for the flux calibration. A project was established as part of the Pre-Major in Astronomy Program (Pre-MAP[5]) for the spectroscopic classification. In order to classify the spectra, we first attempted to fit stellar template spectra from McGurk et al. (2010) using the first Eigen-component for each template. For the objects which were not well-fit by the stellar spectra according to visual inspection of the absorption lines and continuum, we then attempted to fit quasar composite spectra from Vanden Berk et al. (2001) as a function of redshift,

[5]http://www.astro.washington.edu/users/premap/

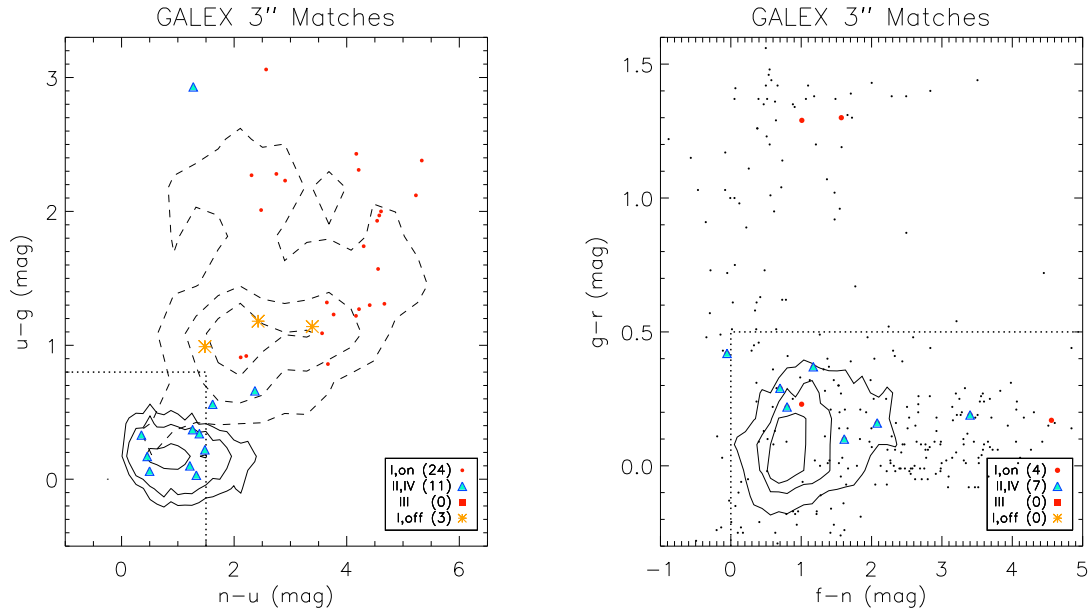

Figure 4.22 GALEX-SDSS UV-optical color-color diagrams for variable quasars (solid contours) and stars (dashed contours or black dots). Our variability-selected quasar candidates with GALEX counterparts are shown as symbols (according to the optical color regions in Fig. 4.21). The dotted lines show our selection regions for follow-up spectroscopy.

looking to match the wavelength ratios of broad emission lines.

4.8.4 Results

Of the 255 targets, 107 were observed by BOSS with 61% as confirmed quasars, or 67% when excluding objects with outlier flags set. In total, 144 $i < 19$ candidates now have existing SDSS or DIS spectra or pre-existing spectra in SIMBAD[6], and 90% of these do not have outlier flags set. Of the objects with outlier flags set which now have spectra, only 2/16 are quasars. This result justifies our criteria for excluding light curves with problematic outliers. Of the sample without outlier flags set, 54% are quasars, or 57% when excluding objects with colors similar to late-type stars ($g - i > 1.8$). The contaminants include 6 supernovae (according to SIMBAD), 54–62 stars, 3 emission line galaxies, and 3–11 normal galaxies. Of the stars with detailed classifications available, most are variable K-type stars

[6]http://simbad.u-strasbg.fr/simbad/

Table 4.5. Multiwavelength-selected Targets by Optical Color

Color Region	N_{total}	X-ray	Radio	UV	IR	Any
I	71	1	1	2	3	7
I_{off}	5	0	0	0	3	3
II	17	0	1	10	4	12
III	3	0	0	0	1	1
IV	2	0	0	0	0	0

Note. — The last column lists the number of targets that satisfy any of the conditions 1–5.

(17/52), but there are also several variable F- and G-type stars, and several M-dwarfs.

The classifications and redshifts of our top priority candidates are presented in the two right-hand columns of Table 4.4, where available. In this Table, ∼50–70% are quasars, one is a BL Lac found in Croom et al. (2009b), one is an AGB star, one is a supernova, two are M-dwarfs (one is a flaring M-dwarf also found by Kowalski et al. 2009), one is a variable G-type star, and one is a K-type star showing a slowly declining flux over time and associated with a radio counterpart. Four objects are possibly quasars but are not certain due to low signal-to-noise in their spectra. There are two confirmed quasars and likely a third in the I_{off} region of optical color space that have quasar-like near-IR colors in 2MASS. Targeting objects in this color regime with quasar-like variability seems promising for finding obscured or reddened quasars (although mid-IR data would be useful here as well).

As part of our DIS follow-up program, we discovered a quasar at redshift 1.52 which was a top-priority target (ID 1743419) lacking any spectrum in BOSS or SIMBAD. The light curve and DIS spectrum for this quasar are shown in Figure 4.23. We discovered another quasar at redshift 3.07 (SDSS J031052.13+005303.4) which was likely missed by the SDSS spectroscopic targeting due to its stellar-like colors ($u-g = 1.54$, $g-r = 0.2$). Its light curve and DIS spectrum are also presented in Figure 4.23. We also identified an object in our high-priority list (ID 2533277) to be a potentially interesting composite. The light curve is

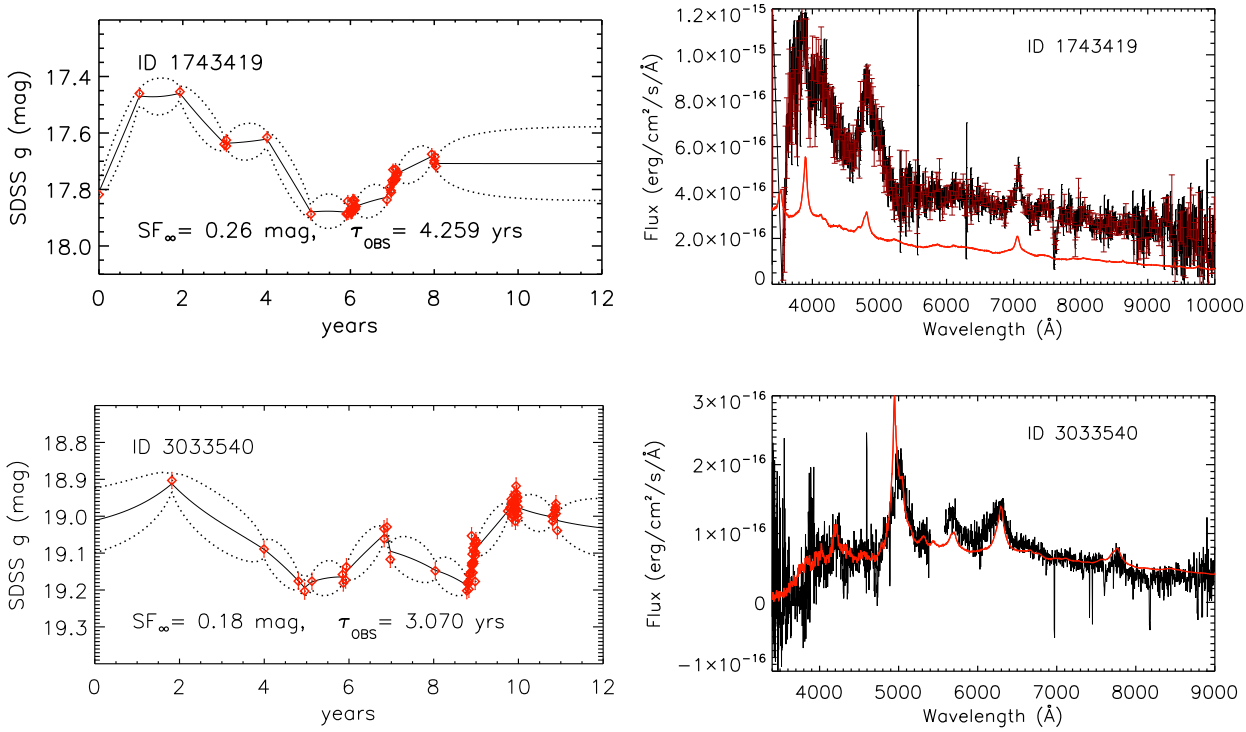

Figure 4.23 Light curves, best-fit DRW models, and DIS spectra for two newly confirmed quasars (cosmic rays were not masked out). The error bars in flux are shown in the top spectrum as an example. The best-fit quasar composite spectra from Vanden Berk et al. (2001) are overplotted in red with redshifts 1.52 (top) and 3.07 (bottom), and with arbitrary normalization.

shown in the top panel of Figure 4.24, and the DIS spectrum obtained on our dark half-night (December 11) is shown in the middle panel. The large emission feature spanning from 7000 to 8300Å was too difficult to identify without obtaining another spectrum. Therefore, two months later, another DIS spectrum of the source was obtained at our request during the observing time of another graduate student, Adam Kowalski (bottom panel). Given the disappearance of the large emission feature in the second spectrum, the source appears to be a composite of a quasar at a redshift of 1.25 and another variable system, perhaps a low accretion rate magnetic binary system (or LARB CV) with variable cyclotron emission (Schmidt et al., 2005). Further spectroscopy for this object would be valuable in determining the nature of the system(s).

Of the 15 candidates with $16.2 < i < 19$, $(g - i) < 1.8$, no outlier flag set, and still no current spectroscopic information, 80% are located outside the R. A. range followed-up by BOSS, and thus lie in regions of enhanced stellar density. Based on our results, $\sim$70% of these 15 candidates are likely to be quasars. Further spectroscopic follow-up will not only confirm more quasars but also may discover new interesting variable objects. Aside from quasar selection, our method identifies some stars showing quasi-periodic variations such as AGBs or variable white dwarfs (DAVs, e.g., SDSS J014023.98+001542.7), which provide excellent probes of stellar structure.

Finally, the redshift distribution of the newly confirmed quasars in our sample of $i < 19$ candidates is shown in Figure 4.25. Given that 45% of the newly confirmed quasars in our list have redshifts greater than 2.2, it is clear that SDSS missed a considerable fraction of quasars because of their stellar-like colors. This illustrates the need for a variability-based selection, as is currently used in SDSS-III in the S82 region of the sky. In future large time-domain surveys such as LSST, variability selection may be more efficient in extending quasar samples to fainter objects at any given redshift than multiple-fiber spectrographs such as used in BOSS.

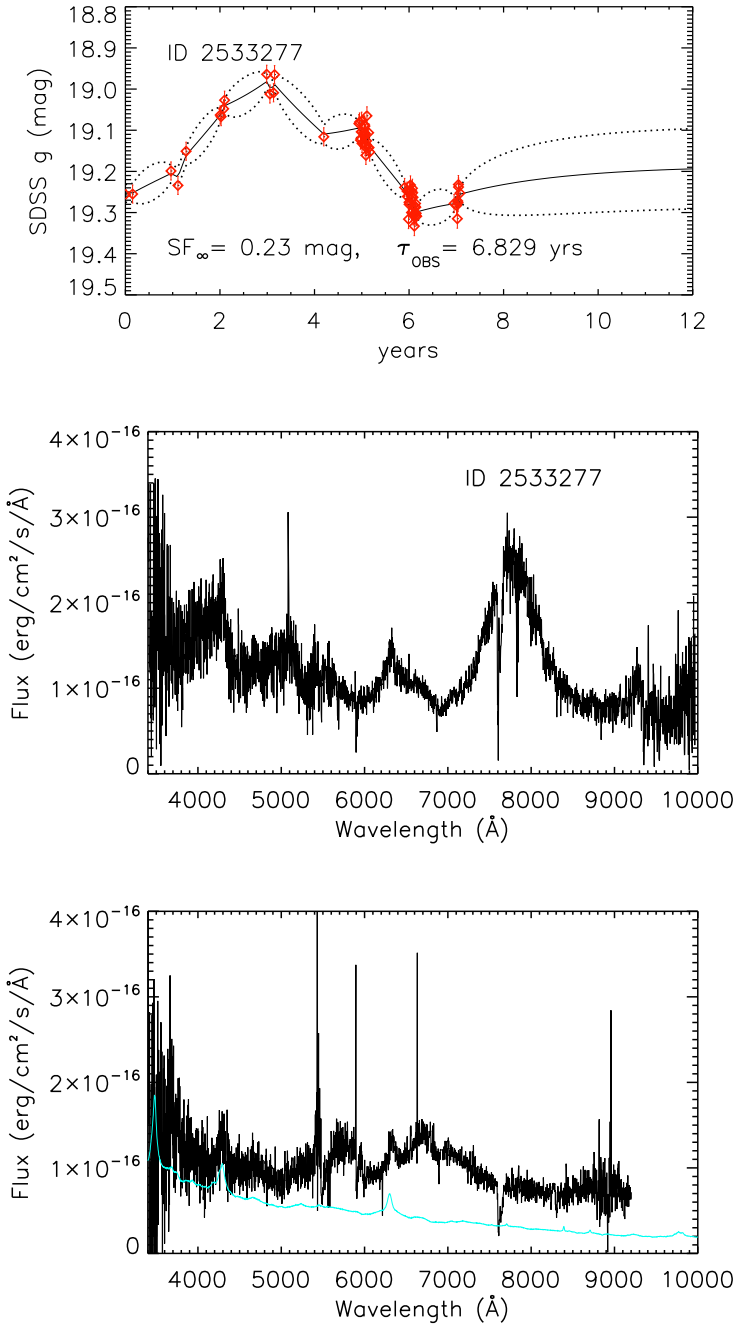

Figure 4.24 Light curve, best-fit DRW model, and DIS spectra for an interesting, possible quasar+CV composite. The best-fit quasar composite spectrum from Vanden Berk et al. (2001) is overplotted in cyan with redshift 1.25 and arbitrary normalization. The broad emission feature at 8000Å is present in the middle panel, but has disappeared 2 months later as seen in the bottom panel.

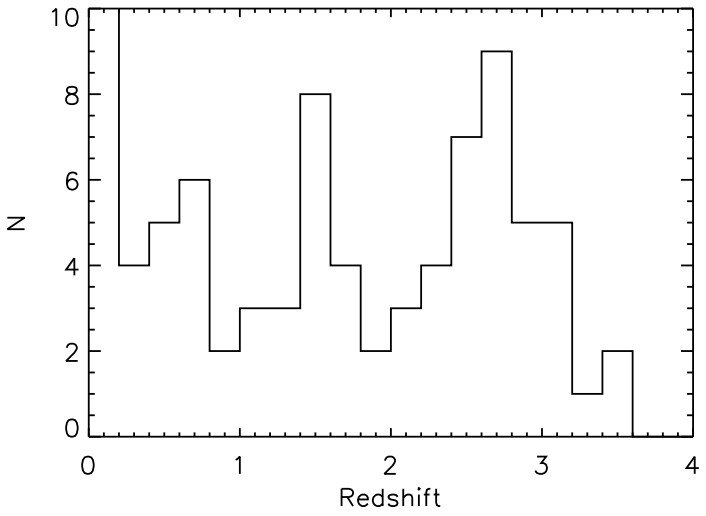

Figure 4.25 Redshift distribution for all newly-confirmed quasars in our list of 255 $i < 19$ quasar candidates based on variability. Roughly half are above redshift 2.2.

Chapter 5

SUMMARY AND CONCLUSIONS

Before the Sloan Digital Sky Survey (SDSS) began producing large amounts of data to study quasar variability, samples were limited to ∼100 objects with sparse light curves. Nevertheless, studies were able to compare the observed variability to phenomenological models including starbursts and accretion disk instabilities (Kawaguchi et al. 1998). Significant progress in the description of quasar variability has been made by utilizing SDSS data (e.g., Vanden Berk et al. 2004) and by including longer-term data available from the Palomar Observatory Sky Survey (POSS; de Vries et al. 2005; Sesar et al. 2006). With such large quasar samples provided early on by the SDSS (∼20,000) with few epochs of observation, variability studies analyzed the magnitude changes of all quasars simultaneously as a function of time lag, wavelength, and other properties. However, this ensemble variability is only an average over individual quasar behavior. Individual light curves have, until recently, been very limited in either sample size or time resolution. Analyses have traditionally used non-parametric characterizations of variability such as structure functions and PSDs, and it has thus been difficult to connect the variability to theory.

For my thesis, I have worked towards revolutionizing this field on three levels. First, I compiled the largest sample of quasar light curves, based on re-calibrated SDSS Stripe 82 (S82) data extended to 50-year timescales by matching to POSS. For now, this data set is uniquely powerful for studying quasar variability. Second, I have shown that quasar variability is well-characterized by the damped random walk (DRW) model of Kelly et al. (2009) and Kozłowski et al. (2010b) with two parameters, a timescale (τ) and asymptotic amplitude (SF_∞), and that these parameters are well-correlated with the physical properties of the quasar (wavelength, luminosity, Eddington factor, and black hole mass). Third, I have shown how to use these tools to variability-select quasars in large surveys. My method overcomes many of the traditional problems of quasar selection, such as the similar colors

of stars and quasars at intermediate redshift, and odd colors due to absorption troughs or extinction. Furthermore, I demonstrated that this selection technique works in a new SDSS redshift survey of variability-selected quasars, and I reconciled these variability models of individual quasars with their ensemble statistics.

By applying DRW models to the light curves of ∼10,000 SDSS quasars in S82, we found that τ increases with increasing wavelength and increases with increasing black hole mass. The amplitude of variability is anti-correlated with the Eddington ratio, which suggests a scenario where optical fluctuations are tied to variations in the accretion rate. The radio-loudest quasars have systematically larger variability amplitudes by about 30%, when corrected for the other observed trends, while the distribution of their characteristic timescale is indistinguishable from that of the full sample. We did not detect any statistically robust differences in τ and SF_∞ between the full sample and the small subsample of quasars detected by ROSAT. These results provide a simple quantitative framework for generating mock quasar light curves and are currently used in LSST image simulations.

By constructing the ensemble structure function for ∼25,000 quasars repeatedly imaged by SDSS over the Northern Galactic Cap, we have fully explained the ensemble variability statistics of the non-S82 quasars such as the exponential distribution of large magnitude changes. All available data are consistent with the DRW model as a viable description of the optical continuum variability of quasars on timescales of ∼5–2000 days in the rest frame. We have used these models to predict the incidence of quasar contamination in transient surveys such as those from PTF and LSST.

With the aid of an SDSS spectroscopically confirmed quasar sample, we demonstrated that variability selection in typical extragalactic fields with low stellar density can deliver complete samples with reasonable purity. Compared to a selection method based solely on the slope of the structure function, the inclusion of τ information boosts E from 60% to 75% while maintaining a highly complete sample (98%) even in the absence of color information. Hence, selection methods based on variability will play an important role in the selection of quasars with data provided by upcoming large sky surveys. Given an adequate survey cadence of several years in length, variability selection will provide an even better method than color selection for separating quasars from stars.

We tested our variability selection on a sample of unidentified objects in S82 with $i < 19$ using results from the SDSS BOSS spectroscopic survey and our own follow-up program with APO DIS. We find that by selecting objects with $\tau > 32$ days and a DRW likelihood relative to white noise exceeding the 10σ level ($\Delta L_{noise} > 10$), we achieve a completeness of 95% with 91% purity. By targeting objects that also resemble quasars in other wavelength regimes, we discovered that while it can be used to find quasars, the selection criteria also identify other interesting variable sources such as BL Lacs, supernovae, AGBs, flaring M-dwarfs, and possibly CVs.

Aside from quasar selection, my work thus far has had the overarching goal of connecting observations of quasar variability to theory. First, given the measured timescales of roughly 1 year in the rest frame, we are able to conclude that the characteristic timescale for quasar variability is on the order of the thermal timescale. The DRW mathematical model supports accretion disk instabilities as the dominant variability mechanism, with instabilities growing on a timescale similar to the thermal timescale for accretion disks. However, it should be noted that accretion rate perturbations do not seem to be the sole driver of the variability, given the results of Schmidt et al. (2012), which indicate that an extra source such as hot spots is necessary to account for the optical color changes.

The measured timescales are also similar to the dynamical timescales for accretion disks (on the order of months). However, it is more likely that the variability we are seeing is associated with the thermal timescale rather than the dynamical timescale because of several reasons. First, the results form the larger two-epoch sample of quasars indicate that τ is actually longer than one year, and closer to 2 years in the rest frame. Secondly, magnetohydrodynamic (MHD) simulations of radiation-dominated AGN accretion disks have predicted shorter thermal timescales than implied by the standard α prescription (Turner 2004), leading to a better agreement with the observed timescales. Third, while the characteristic timescale for turbulence in MHD simulations is on the order of the dynamical timescale, one would expect dissipation of magnetic and kinetic energy density on the thermal timescale (Hirose et al. 2009).

If MHD models could be improved to reproduce the observed characteristics of variability, they should be able to shed light on the source of stochasticity in quasar light curves.

Also, relating variability parameters to physical parameters is only a first step to understanding the nature of quasar variability, since it is now necessary to connect the observed trends to theories of disk structure as well as with other probes of their structure, such as reverberation mapping, variability in other wavebands, and gravitational microlensing, as discussed below. In addition to improving the DRW model and using it in other contexts, future work is necessary for connecting the model to quasar physics and to other probes of quasar structure.

5.1 Expanding and Improving on the DRW Model

In order to connect observations of quasar variability to physical theories, continued effort is necessary in testing whether the DRW model is correct, particularly on long timescales. While our analysis of the S82 quasars showed that τ is similar to the thermal timescale of accretion disks, the X-ray variability of AGNs and galactic black holes shows evidence for multiple τ, corresponding to various physical timescales (Kelly et al. 2011). Due to the limited cadence of the S82 survey, it is necessary to obtain better light curves to test such long-term behavior in the optical (see Section 2.5.4). Furthermore, the τ values measured for S82 light curves seem to be biased low when comparing to the magnitude changes in a larger two-epoch sample of SDSS quasars, and it is not clear whether the bias is due to an inaccurate model description or the limited cadence. The Dark Energy Survey (DES; Honscheid et al. 2008) supernova program will greatly expand many of the S82 quasar light curves with *griz* sampling once per week for $\sim$3 months per year over 5 years. With a longer time baseline, the accuracy of the DRW model, as well as the intrinsic properties of quasars which are revealed through variability such as the τ distribution, can be reassessed.

On shorter timescales, recent results from the Kepler satellite (Mushotzky et al., 2011) showed that the PSD for quasars on timescales of a day is significantly steeper than that for a DRW model (PSD $\propto f^{-3}$). Well-sampled optical light curves from the OGLE survey for a larger sample of quasars are in support of this result (Zu et al., 2012). Interestingly, this steepening of the power spectrum is not present in the short-term X-ray variability of quasars. This difference presents a puzzle in connecting the X-ray and optical variability of quasars and in understanding the underlying physics. While the mathematical models are

now in place, clearly there is more work to be done in obtaining data of sufficient quality, such as a larger sample of well-sampled, simultaneous optical and X-ray light curves for quasars, and in connecting theoretical models of X-ray and optical variability.

Finally, in reverberation mapping, the lag times between the continuum and broad lines can be improved by modeling the AGN continuum variability as a DRW, assuming that the emission line variability originates from variations in the continuum (Zu et al., 2011). The resulting BLR sizes are of paramount importance in extragalactic studies because they are used to calibrate the mass–σ relation of black holes. Such measurements are feasible even in the absence of spectra, using broad band photometric light curves, which will be useful in large time-domain surveys (Chelouche & Daniel, 2011). Here, and in the recent development of a cosmological tool where AGN luminosity distances to $z \sim 4$ can be estimated from reverberation mapping (Watson et al., 2011), the DRW model would be valuable, since a better quantitative model for the continuum variability is needed.

5.2 Connecting to Physics: The Disk–Jet Connection

The sub-mm and radio emission of quasars can potentially shed light on the jet–disk connection in quasar variability. For example, by matching the S82 quasar sample to the FIRST and NVSS radio catalogs of Kimball & Ivezić (2008), we found that radio-loud (RL) quasars in S82 have larger optical variability amplitudes than their radio-quiet counterparts. Furthermore, in a work currently in progress (Ruan et al.), we find that the optical variability amplitudes and timescales differ significantly between Fermi blazars and quasars. The sub-mm variability of blazars can also be well-described by the same process as for quasars in the optical and also exhibits long-term variability (Strom et al., 2010). These results suggest a large influence from jet emission, but nevertheless, it seems that the main mechanism responsible for the optical variability operates irrespective of the radio flux.

While the viewing angle of the jet may play a role in the variability of quasars versus blazars, the evolution of jets over cosmological timescales is still a largely unexplored subject and may also play a role in quasar time variability. Recently, the influence of jets was taken into consideration for the first time in a cosmological simulation (Dubois et al., 2011). In their simulation, quasars transition from a radio-quiet "quasar mode" at higher redshift to

a "radio mode" with a prevalent jet at lower redshift, where the transition is determined by the Eddington ratio (L/L_{Edd}). The simulation is able to reproduce the bright cutoff in the quasar luminosity function while accounting for the increased number of radio galaxies at low redshift. A next step would be to see if the trends between variability and radio-loudness support this picture. For example, in this cosmological context, L/L_{Edd} is lower for RL quasars. Therefore, given that variability amplitude is anti-correlated with L/L_{Edd}, it follows that RL quasars have larger variability amplitudes. However, for the S82 quasars, we find that the trend between the variability amplitude and radio-loudness is independent of the anti-correlation between the amplitude and L/L_{Edd}. Therefore, the measured variability of RL quasars is larger than expected based simply on the accretion rate, suggesting that the mechanical energy output from a jet during a late evolutionary stage of AGNs may be revealed through their optical and radio variability.

5.3 Connecting to Other Probes: Lensed Quasars and Flux Ratios

Other probes of quasar structure, such as gravitational lenses, provide a way to test the origins of variability and connect to theories of disk structure. In a strong lens system, microlensing of the quasar by stars in the foreground galaxy effectively resolves the quasar accretion disk, since the stellar Einstein radii subtend angles comparable to the disk (Pooley et al., 2007; Morgan et al., 2010). Assuming that the intrinsic quasar variability must alter the surface brightness structure of the disk, the microlensing magnifications must in turn be affected, and the spatial patterns of variability should be detectable in the microlensing signals. Interestingly, the accretion disk sizes inferred from microlensing light curves are in disagreement with the sizes of standard thin accretion disks (Blackburne et al., 2011). An inhomogeneous disk model, where the temperature fluctuations throughout the disk are driven by a DRW process, can explain the disk sizes derived from microlensing light curves while matching the observed level of optical variability, and predicts SEDs which are in better agreement with observations than standard thin disk models (Dexter & Agol, 2011). This model is testable using microlensing light curves, wherein the variations due to inhomogeneities are estimated to exceed 1% (Dexter & Agol, 2011).

However, in order to study the source quasar through lensing magnifications, the mass

distribution of the lens needs to be understood. One problem in existing studies is that the microlensing is convolved with "millilensing" by substructure in the lens. In the mid-IR, the dust torus in the quasar subtends a significantly larger angle than do stars in the lens galaxy, making microlensing unimportant. Thus, in the mid-IR, any discrepancy between the image flux ratios and those expected in a smooth lens model (a so-called "flux ratio anomaly") points to substructure in the lens galaxy (Mao & Schneider, 1998; MacLeod et al., 2009). This is also cosmologically important, because the number of substructures found in cosmological simulations significantly exceeds the number of observed systems with flux ratio anomalies (Dalal & Kochanek, 2002; Metcalf & Amara, 2010). The apparent deficiency of substructures is not yet a strong challenge to the current cosmological model (CDM), in part because the sample of observed lenses is extremely small. The LBTI instrument on the Large Binocular Telescope (LBT), now being commissioned, will be the best tool for increasing the sample size of flux ratio anomalies until the advent of the James Webb Space Telescope or a 30-meter telescope.

BIBLIOGRAPHY

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543

Agüeros, M. A., Ivezić, Ž., Covey, K. R., et al. 2005, AJ, 130, 1022

Ai, Y. L., Yuan, W., Zhou, H. Y., et al. 2010, ApJ, 716, L31

Antonucci, R. 1993, ARA&A, 31, 473

Aretxaga, I., Cid Fernandes, R., & Terlevich, R. J. 1997, MNRAS, 286, 271

Arévalo, P., McHardy, I. M., & Summons, D. P. 2008a, MNRAS, 388, 211

Arévalo, P., Uttley, P., Kaspi, S., et al. 2008b, MNRAS, 389, 1479

Bauer, A., Baltay, C., Coppi, P., et al. 2009, ApJ, 696, 1241

Belloni, T., ed. 2010, Lecture Notes in Physics, Berlin Springer Verlag, Vol. 794, The Jet Paradigm

Bhatti, W. A., Richmond, M. W., Ford, H. C., & Petro, L. D. 2010, ApJS, 186, 233

Blackburne, J. A., & Kochanek, C. S. 2010, ApJ, 718, 1079

Blackburne, J. A., Pooley, D., Rappaport, S., & Schechter, P. L. 2011, ApJ, 729, 34

Blaes, O., Hubeny, I., Agol, E., & Krolik, J. H. 2001, ApJ, 563, 560

Bongiorno, A., Zamorani, G., Gavignaud, I., et al. 2007, A&A, 472, 443

Boutsia, K., Leibundgut, B., Trevese, D., & Vagnetti, F. 2009, A&A, 497, 81

Boutsia, K., Leibundgut, B., Trevese, D., & Vagnetti, F. 2010, in Astronomical Society of the Pacific Conference Series, Vol. 424, 9th International Conference of the Hellenic Astronomical Society, ed. K. Tsinganos, D. Hatzidimitriou, & T. Matsakos, 277

Butler, N. R., & Bloom, J. S. 2011, AJ, 141, 93

Cenko, S. B., Butler, N. R., Ofek, E. O., et al. 2010, AJ, 140, 224

Chelouche, D., & Daniel, E. 2011, ArXiv e-prints

Churazov, E., Gilfanov, M., & Revnivtsev, M. 2001, MNRAS, 321, 759

Collier, S., & Peterson, B. M. 2001, ApJ, 555, 775

Croom, S. M., Richards, G. T., Shanks, T., et al. 2009a, MNRAS, 399, 1755

—. 2009b, MNRAS, 392, 19

Dalal, N., & Kochanek, C. S. 2002, ApJ, 572, 25

de Vries, W. H., Becker, R. H., & White, R. L. 2003, AJ, 126, 1217

de Vries, W. H., Becker, R. H., White, R. L., & Loomis, C. 2005, AJ, 129, 615

Delgado, F., Cook, K., Miller, M., Allsman, R., & Pierfederici, F. 2006, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 6270, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series

Denney, K. D., Peterson, B. M., Dietrich, M., Vestergaard, M., & Bentz, M. C. 2009, ApJ, 692, 246

Dexter, J., & Agol, E. 2011, ApJ, 727, L24

Djorgovski, S. G., Gal, R. R., Odewahn, S. C., et al. 1998, in Wide Field Surveys in Cosmology, ed. S. Colombi, Y. Mellier, & B. Raban, 89

Dubois, Y., Devriendt, J., Slyz, A., & Teyssier, R. 2011, ArXiv e-prints

Emmanoulopoulos, D., McHardy, I. M., & Uttley, P. 2010, MNRAS, 404, 931

Frank, J., King, A., & Raine, D. J. 2002, Accretion Power in Astrophysics: Third Edition

Frieman, J. A., Bassett, B., Becker, A., et al. 2008, AJ, 135, 338

Fukugita, M., Ichikawa, T., Gunn, J. E., et al. 1996, AJ, 111, 1748

Gal-Yam, A., Ofek, E. O., Filippenko, A. V., Chornock, R., & Li, W. 2002, PASP, 114, 587

Gaskell, C. M., & Klimek, E. S. 2003, Astronomical and Astrophysical Transactions, 22, 661

Geha, M., Alcock, C., Allsman, R. A., et al. 2003, AJ, 125, 1

Giveon, U., Maoz, D., Kaspi, S., Netzer, H., & Smith, P. S. 1999, MNRAS, 306, 637

Gotz, D., Mereghetti, S., von Kienlin, A., & Beck, M. 2009, GRB Coordinates Network, 9649, 1

Gunn, J. E., Carr, M., Rockosi, C., et al. 1998, AJ, 116, 3040

Hawkins, M. R. S. 1983, MNRAS, 202, 571

—. 1993, Nature, 366, 242

—. 2002, MNRAS, 329, 76

—. 2007, A&A, 462, 581

Hawkins, M. R. S., & Veron, P. 1995, MNRAS, 275, 1102

Hawley, J. F., & Krolik, J. H. 2001, ApJ, 548, 348

Hennawi, J. F., & Prochaska, J. X. 2007, ApJ, 655, 735

Hirose, S., Krolik, J. H., & Blaes, O. 2009, ApJ, 691, 16

Honscheid, K., DePoy, D. L., & for the DES Collaboration. 2008, ArXiv e-prints

Hook, I. M., McMahon, R. G., Boyle, B. J., & Irwin, M. J. 1994, MNRAS, 268, 305

Horne, J. H., & Baliunas, S. L. 1986, ApJ, 302, 757

Hughes, P. A., Aller, H. D., & Aller, M. F. 1992, ApJ, 396, 469

Ivezić, Ž., Menou, K., Knapp, G. R., et al. 2002, AJ, 124, 2364

Ivezić, Ž., Lupton, R. H., Anderson, S., et al. 2003, Mem. Soc. Astron. Italiana, 74, 978

Ivezić, Z., Lupton, R., Johnston, D., et al. 2004, in Astronomical Society of the Pacific Conference Series, Vol. 311, AGN Physics with the Sloan Digital Sky Survey, ed. G. T. Richards & P. B. Hall, 437

Ivezic, Ž., Lupton, R. H., Juric, M., et al. 2004, in IAU Symposium, Vol. 222, The Interplay Among Black Holes, Stars and ISM in Galactic Nuclei, ed. T. Storchi-Bergmann, L. C. Ho, & H. R. Schmitt, 525–526

Ivezić, Ž., Lupton, R. H., Schlegel, D., et al. 2004, Astronomische Nachrichten, 325, 583

Ivezić, Ž., Smith, J. A., Miknaitis, G., et al. 2007, AJ, 134, 973

Ivezic, Z., Tyson, J. A., Acosta, E., et al. 2008, ArXiv e-prints

Jiang, L., Fan, X., Cool, R. J., et al. 2006, AJ, 131, 2788

Kaiser, N., Aussel, H., Burke, B. E., et al. 2002, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4836, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, ed. J. A. Tyson & S. Wolff, 154–164

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., & Siemiginowska, A. 2009, ApJ, 698, 895

Kelly, B. C., Sobolewska, M., & Siemiginowska, A. 2011, ApJ, 730, 52

Kimball, A. E., & Ivezić, Ž. 2008, AJ, 136, 684

Kniffen, D. A., Bertsch, D. L., Fichtel, C. E., et al. 1993, ApJ, 411, 133

Kollmeier, J. A., Onken, C. A., Kochanek, C. S., et al. 2006, ApJ, 648, 128

Koptelova, E., Chen, W. P., Chiueh, T., et al. 2012, ArXiv e-prints

Kowalski, A. F., Hawley, S. L., Hilton, E. J., et al. 2009, AJ, 138, 633

Kozłowski, S., & Kochanek, C. S. 2009, ApJ, 701, 508

Kozłowski, S., Kochanek, C. S., & Udalski, A. 2011, ApJS, 194, 22

Kozłowski, S., Kochanek, C. S., Stern, D., et al. 2010a, ApJ, 716, 530

Kozłowski, S., Kochanek, C. S., Udalski, A., et al. 2010b, ApJ, 708, 927

Kozłowski, S., Kochanek, C. S., Jacyszyn, A. M., et al. 2012, ApJ, 746, 27

Krolik, J. H. 1999, Active galactic nuclei : from the central black hole to the galactic environment, ed. Krolik, J. H.

Kundic, T., Turner, E. L., Colley, W. N., et al. 1997, ApJ, 482, 75

Law, N. M., Kulkarni, S. R., Dekany, R. G., et al. 2009, PASP, 121, 1395

Li, S.-L., & Cao, X. 2008, MNRAS, 387, L41

Liu, H. T., Bai, J. M., Zhao, X. H., & Ma, L. 2008, ApJ, 677, 884

Lomb, N. R. 1976, Ap&SS, 39, 447

LSST Science Collaborations, Abell, P. A., Allison, J., et al. 2009, ArXiv e-prints

Lupton, R., Gunn, J. E., Ivezić, Z., et al. 2001, in Astronomical Society of the Pacific Conference Series, Vol. 238, Astronomical Data Analysis Software and Systems X, ed. F. R. Harnden, Jr., F. A. Primini, & H. E. Payne, 269

Lupton, R. H., Gunn, J. E., & Szalay, A. S. 1999, AJ, 118, 1406

Lyubarskii, Y. E. 1997, MNRAS, 292, 679

Lyutyi, V. M. 1972, AZh, 49, 930

MacLeod, C., Ivezić, Ž., de Vries, W., Sesar, B., & Becker, A. 2008, in American Institute of Physics Conference Series, Vol. 1082, American Institute of Physics Conference Series, ed. C. A. L. Bailer-Jones, 282–286

MacLeod, C. L., Kochanek, C. S., & Agol, E. 2009, ApJ, 699, 1578

MacLeod, C. L., Ivezić, Ž., Kochanek, C. S., et al. 2010, ApJ, 721, 1014

MacLeod, C. L., Ivezic, Z., Sesar, B., et al. 2011a, ArXiv e-prints

MacLeod, C. L., Brooks, K., Ivezić, Ž., et al. 2011b, ApJ, 728, 26

Mahabal, A. A., Djorgovski, S. G., Graham, M. J., et al. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert, 604

Mao, S., & Schneider, P. 1998, MNRAS, 295, 587

Marconi, A., Axon, D. J., Maiolino, R., et al. 2008, ApJ, 678, 693

Markwardt, C. B., Gavriil, F. P., Palmer, D. M., Baumgartner, W. H., & Barthelmy, S. D. 2009, GRB Coordinates Network, 9645, 1

Martini, P., & Schneider, D. P. 2003, ApJ, 597, L109

Matthews, T. A., & Sandage, A. R. 1963, ApJ, 138, 30

McDonald, P., & Eisenstein, D. J. 2007, Phys. Rev. D, 76, 063009

McGurk, R. C., Kimball, A. E., & Ivezić, Ž. 2010, AJ, 139, 1261

McHardy, I. 2010, in Lecture Notes in Physics, Berlin Springer Verlag, Vol. 794, Lecture Notes in Physics, Berlin Springer Verlag, ed. T. Belloni, 203

McHardy, I. M., Koerding, E., Knigge, C., Uttley, P., & Fender, R. P. 2006, Nature, 444, 730

Metcalf, R. B., & Amara, A. 2010, ArXiv e-prints

Meusinger, H., Hinze, A., & de Hoon, A. 2011, A&A, 525, A37

Morgan, C. W., Kochanek, C. S., Morgan, N. D., & Falco, E. E. 2010, ApJ, 712, 1129

Morley, J. C., Nelson, C. R., & Zivot, E. 2003, The Review of Economics and Statistics, 85, pp. 235

Mushotzky, R. F., Edelson, R., Baumgartner, W., & Gandhi, P. 2011, ApJ, 743, L12

Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713

Pereyra, N. A., Vanden Berk, D. E., Turnshek, D. A., et al. 2006, ApJ, 642, 87

Peterson, B. M. 1993, PASP, 105, 247

Peterson, B. M., Bentz, M. C., Desroches, L.-B., et al. 2005, ApJ, 632, 799

Pier, J. R., Munn, J. A., Hindsley, R. B., et al. 2003, AJ, 125, 1559

Pooley, D., Blackburne, J. A., Rappaport, S., & Schechter, P. L. 2007, ApJ, 661, 19

Press, W. H., Rybicki, G. B., & Hewitt, J. N. 1992, ApJ, 385, 404

Reimann, J. D. 1994, PhD thesis, UNIVERSITY OF CALIFORNIA, BERKELEY.

Rengstorf, A. W., Brunner, R. J., & Wilhite, B. C. 2006, AJ, 131, 1923

Rengstorf, A. W., Mufson, S. L., Andrews, P., et al. 2004, ApJ, 617, 184

Richards, G. T. 2008, in American Institute of Physics Conference Series, Vol. 1082, American Institute of Physics Conference Series, ed. C. A. L. Bailer-Jones, 22–28

Richards, G. T., Fan, X., Newberg, H. J., et al. 2002, AJ, 123, 2945

Richards, G. T., Strauss, M. A., Fan, X., et al. 2006, AJ, 131, 2766

Richards, G. T., Myers, A. D., Gray, A. G., et al. 2009, ApJS, 180, 67

Ross, N. P., Shen, Y., Strauss, M. A., et al. 2009, ApJ, 697, 1634

Ross, N. P., Myers, A. D., Sheldon, E. S., et al. 2012, ApJS, 199, 3

Rybicki, G. B., & Press, W. H. 1992, ApJ, 398, 169

—. 1995, Physical Review Letters, 74, 1060

Scargle, J. D. 1982, ApJ, 263, 835

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525

Schmidt, G. D., Szkody, P., Vanlandingham, K. M., et al. 2005, ApJ, 630, 1037

Schmidt, K. B., Marshall, P. J., Rix, H.-W., et al. 2010, ApJ, 714, 1194

Schmidt, K. B., Rix, H.-W., Shields, J. C., et al. 2012, ApJ, 744, 147

Schneider, D. P., Hall, P. B., Richards, G. T., et al. 2007, AJ, 134, 102

Schneider, D. P., Richards, G. T., Hall, P. B., et al. 2010, AJ, 139, 2360

Schweitzer, M., Lutz, D., Sturm, E., et al. 2006, ApJ, 649, 79

Sesar, B., Svilković, D., Ivezić, Ž., et al. 2006, AJ, 131, 2801

Sesar, B., Ivezić, Ž., Lupton, R. H., et al. 2007, AJ, 134, 2236

Shen, Y., Greene, J. E., Strauss, M. A., Richards, G. T., & Schneider, D. P. 2008, ApJ, 680, 169

Shen, Y., Richards, G. T., Strauss, M. A., et al. 2011, ApJS, 194, 45

Smith, J. A., Tucker, D. L., Kent, S., et al. 2002, AJ, 123, 2121

Stoughton, C., Lupton, R. H., Bernardi, M., et al. 2002, AJ, 123, 485

Strom, A. L., Siemiginowska, A., Gurwell, M. A., & Kelly, B. C. 2010, ArXiv e-prints

Timmer, J., & Koenig, M. 1995, A&A, 300, 707

Trèvese, D., Kron, R. G., & Bunone, A. 2001, ApJ, 551, 103

Trèvese, D., & Vagnetti, F. 2002, ApJ, 564, 624

Turner, N. J. 2004, ApJ, 605, L45

Udalski, A., Kubiak, M., & Szymanski, M. 1997, Acta Astron., 47, 319

Udalski, A., Szymanski, M. K., Soszynski, I., & Poleski, R. 2008, Acta Astron., 58, 69

Urry, C. M., & Padovani, P. 1995, PASP, 107, 803

Uttley, P., McHardy, I. M., & Papadakis, I. E. 2002, MNRAS, 332, 231

van den Bergh, S., Herbst, E., & Pritchet, C. 1973, AJ, 78, 375

Vanden Berk, D. E., Richards, G. T., Bauer, A., et al. 2001, AJ, 122, 549

Vanden Berk, D. E., Lee, B. C., Wilhite, B. C., et al. 2002, ApJ, 576, 673

Vanden Berk, D. E., Wilhite, B. C., Kron, R. G., et al. 2004, ApJ, 601, 692

Vestergaard, M., & Peterson, B. M. 2006, ApJ, 641, 689

Voevodkin, A. 2011, ArXiv e-prints

Voges, W., Aschenbach, B., Boller, T., et al. 1999, A&A, 349, 389

Watson, D., Denney, K. D., Vestergaard, M., & Davis, T. M. 2011, ApJ, 740, L49+

Welsh, B. Y., Wheatley, J. M., & Neil, J. D. 2011, A&A, 527, A15

White, R. L., Becker, R. H., Helfand, D. J., & Gregg, M. D. 1997, ApJ, 475, 479

Wilhite, B. C., Brunner, R. J., Grier, C. J., Schneider, D. P., & vanden Berk, D. E. 2008, MNRAS, 383, 1232

Wilhite, B. C., Vanden Berk, D. E., Brunner, R. J., & Brinkmann, J. V. 2006, ApJ, 641, 78

Wilhite, B. C., Vanden Berk, D. E., Kron, R. G., et al. 2005, ApJ, 633, 638

Wold, M., Brotherton, M. S., & Shang, Z. 2007, MNRAS, 375, 989

York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, AJ, 120, 1579

Young, M., Brandt, W. N., Xue, Y. Q., et al. 2012, ApJ, 748, 124

Zu, Y., Kochanek, C. S., Kozłowski, S., & Udalski, A. 2012, ArXiv e-prints

Zu, Y., Kochanek, C. S., & Peterson, B. M. 2011, ApJ, 735, 80

Appendix A

CATALOG FORMAT

Here we present a database of ∼3.5 million photometric measurements for 80,000 spectroscopically confirmed quasars to be used for time variability studies. The database consists of three different data sets: two with repeated SDSS imaging in five UV-to-IR photometric bands, and one with SDSS versus POSS imaging for three bands. The observed time lags span the range from 0.8 days to 10 years for the SDSS data sets, and up to 20 years for the SDSS vs. DPOSS data set. The three data sets are described below:

1. *Northern Survey:* The SDSS imaging data are obtained by drift scanning. Because of the scan overlaps, and because of the scan convergence near the survey poles, about 40% of the northern survey area (∼4000 deg^2) is surveyed at least twice. This method provides 2-epoch 5-band coverage for ∼25,000 spectroscopically confirmed quasars. We adopt the SDSS BEST photometry listed in the DR7 Quasar Catalog V (Schneider et al. 2010) for the primary observations, and we searched for all unresolved secondary observations within 1 arcsecond of the primary using CasJobs. Here, we include all observations up to three per quasar (only 0.2% of the sample had more than three observations). We also include the redshifts, absolute magnitudes, FIRST, RASS, and 2MASS photometry as listed in the DR7 Quasar Catalog V, and the black hole masses as measured from emission line widths by Shen et al. (2011). The catalog format is found in Table A.1. We also provide a list of all the objects (in the same order) with the same exact format as the DR7 Quasar Catalog V (Table 2 in Schneider et al. 2010).

2. *Southern Survey:* About 290 square degrees of the southern survey area has already been observed ∼60 times to search for variable objects and, by stacking the frames, to go deeper. This is the SDSS Stripe 82, which is 22h 24m < R.A. < 04h 08m and

|Dec| $<$ 1.27 deg. These multi-epoch data have time scales ranging from 3 hours to almost 10 years. This method provides well-sampled 5-band light curves for an unprecedented number of quasars (9,258). The catalog format is found in Table A.2, and the light curve file format is found in Table A.3. We also provide a list of all the objects (in the same order) with the same exact format as the DR5 Quasar Catalog IV (Table 2 in Schneider et al. 2007).

3. *SDSS–DPOSS:* We also include a catalog of all SDSS DR7 quasars with DPOSS observations. Following the procedure outlined in Sesar et al. (2006), we have recalibrated DPOSS data (Djorgovski et al. 1998) in 8,000 deg^2 of sky from the SDSS Data Release 5. The main advantage of this data set, which includes 81,189 quasars, is its long time baseline of 20 years. Here, we present the SDSS–DPOSS photometry in GRI bands, where the latter is accurate to 0.10–0.15 mag. The catalog format is presented in Table A.4.

For more details, see http://www.astro.washington.edu/users/ivezic/macleod/qso_dr7/.

Table A.1: Northern Survey: Catalog Format

Column	Format	Units	Description
1	I6	—	Row of DR7 Quasar Catalog V (out of 105783)
2	F7.3	mag	Absolute i band magnitude[h]
3	F6.4	—	Redshift
4	F6.3	$[M_\odot]$	Black hole mass ("0" means none could be found) [g]
5	F6.3	[ergs/s]	Bolometric luminosity[g]
6	F6.3	mag	Galactic extinction in u filter (from Schlegel et al. 1998)[c]
7	F7.3	mag	FIRST Peak 20 cm flux density[d]
8	F8.3	—	Signal-to-noise ratio for FIRST flux density
9	F8.3	[ct/s]	RASS BSC/FSC full band count rate[e]
10	F7.3	—	Signal-to-noise ratio for RASS count rate
11	F7.3	mag	2MASS J band magnitude[f]
12	F6.3	mag	Error in J
13	F7.3	mag	2MASS H band magnitude[f]
14	F6.3	mag	Error in H
15	F7.3	mag	2MASS K band magnitude[f]
16	F6.3	mag	Error in K
17	I3	—	Total number of observations
18	I1	—	Flag for Stripe 82 object[a]
19	F9.3	days	Modified Julian Date of imaging observation
20	F6.3	mag	BEST SDSS u band PSF magnitude[b]
21	F6.3	mag	Error in u
22	F6.3	mag	BEST SDSS g band PSF magnitude[b]
23	F6.3	mag	Error in g
24	F6.3	mag	BEST SDSS r band PSF magnitude[b]

Continued on Next Page...

Table A.1 – Continued

Column	Format	Units	Description
25	F6.3	mag	Error in r
26	F6.3	mag	BEST SDSS i band PSF magnitude[b]
27	F6.3	mag	Error in i
28	F6.3	mag	BEST SDSS z band PSF magnitude[b]
29	F6.3	mag	Error in z
...	...	...	Fields 19 through 29 repeated for each additional observation

[a] A value of 1 (0) indicates a (non-)Stripe 82 source.

[b] SDSS photometric measurements are asinh magnitudes (Lupton, Gunn, & Szalay 1999) and are normalized (to ∼3% accuracy) to the AB-magnitude system (Oke & Gunn 1983). Uncorrected for Galactic extinction. A value of 0.000 indicates that the value could not be retrieved from the SDSS database.

[c] Where A_g, A_r, A_i, $A_z = 0.736$, 0.534, 0.405, $0.287 \times A_u$, respectively.

[d] In AB magnitudes; $-2.5\log(f_\nu/3631\mathrm{Jy})$.

[e] X-ray data from ROSAT All Sky Survey Bright and Faint source catalogs. A value of −9.000 indicates a non-detection.

[f] All 2MASS data are from the 2MASS All-Sky Data Release Point Source Catalog (PSC) as of 2003 March 25. Note that 2MASS measurements are Vega-based, not AB, magnitudes. A value of 0.000 indicates a non-detection.

[g] Taken from the Shen et al. (2011) catalog.

[h] For $\Omega_m = 0.300$, $\Omega_\Lambda = 0.700$, $h = 0.70$, and $\alpha_Q = -0.50$.

Table A.2. Southern Survey: Catalog Format

Column	Format	Units	Description
1	I7	—	The name of the light curve file
2	F10.6	deg	Median Right Ascension in decimal degrees (J2000)
3	F10.6	deg	Median Declination in decimal degrees (J2000)
4	I5	—	Row of DR5 Quasar Catalog IV (out of 77429)
5	F7.3	mag	Absolute i-band magnitude, K-corrected to $z = 0$[d]
6	F7.3	mag	Absolute i-band magnitude, K-corrected to $z = 2$[a]
7	F6.4	—	Redshift
8	F5.3	[$M_\odot$]	Black hole mass ("0.000" means none could be found)[a]
9	F6.3	[ergs/s]	Bolometric luminosity[a]
10	F6.3	mag	SDSS BEST u-band PSF magnitude[b]
11	F6.3	mag	SDSS BEST g-band PSF magnitude[b]
12	F6.3	mag	SDSS BEST r-band PSF magnitude[b]
13	F6.3	mag	SDSS BEST i-band PSF magnitude[b]
14	F6.3	mag	SDSS BEST z-band PSF magnitude[b]
15	F6.3	mag	Galactic extinction in u filter (from Schlegel et al. 1998)[c]

[a]Taken from the Shen et al. (2008) catalog. A value of "-1" indicates a newly confirmed DR7 quasar, and thus is not present in Shen et al. (2008) (see the latest version, Shen et al. 2011, for the values).

[b]SDSS photometric measurements are asinh magnitudes (Lupton, Gunn, & Szalay 1999) and are normalized (to ∼3% accuracy) to the AB-magnitude system (Oke & Gunn 1983). Uncorrected for Galactic extinction. A value of 0.000 indicates that the value could not be retrieved from the SDSS database.

[c]Where A_g, A_r, A_i, $A_z = 0.736$, 0.534, 0.405, $0.287 \times A_u$, respectively. This value is set to zero if it is a newly confirmed DR7 quasar (see the DR7 Quasar Catalog V for the true values).

[d]For $\Omega_m = 0.300$, $\Omega_\Lambda = 0.700$, $h = 0.70$, and $\alpha_Q = -0.50$.

Table A.3. Southern Survey: Light Curve File Format

Column	Format	Units	Description
1	D	days	Modified Julian Date for u-band observation
2	F6.3	mag	Apparent u-band magnitude[a]
3	F5.3	mag	Error in u
4	D	days	Modified Julian Date for g-band observation
5	F6.3	mag	Apparent g-band magnitude[a]
6	F5.3	mag	Error in g
7	D	days	Modified Julian Date for r-band observation
8	F6.3	mag	Apparent r-band magnitude[a]
9	F5.3	mag	Error in r
10	D	days	Modified Julian Date for i-band observation
11	F6.3	mag	Apparent i-band magnitude[a]
12	F5.3	mag	Error in i
13	D	days	Modified Julian Date for z-band observation
14	F6.3	mag	Apparent z-band magnitude[a]
15	F5.3	mag	Error in z
16	D	deg	Median Right Ascension in decimal degrees (J2000)[b]
17	D	deg	Median Declination in decimal degrees (J2000)

[a]Not corrected for Galactic absorption. Bad observations are printed as "-99.99".

[b]360 deg is subtracted from all RA values exceeding 300 deg.

Table A.4: SDSS–DPOSS: Catalog Format

Column	Format	Units	Description
1	F10.6	deg	Right Ascension in decimal degrees (J2000)
2	F10.6	deg	Declination in decimal degrees (J2000)
3	I3	—	DPOSS Plate ID
4	F9.4	yrs	Epoch of the DPOSS observation in the G band
5	F9.4	yrs	Epoch of the DPOSS observation in the R band
6	F9.4	yrs	Epoch of the DPOSS observation in the I band
7	F6.2	mag	DPOSS catalog G magnitude[d]
8	F6.2	mag	Error in G
9	F6.2	mag	G-band quality flag[e]
10	F6.2	mag	DPOSS catalog R magnitude[d]
11	F6.2	mag	Error in R
12	F6.2	mag	R-band quality flag[e]
13	F6.2	mag	DPOSS catalog I magnitude[d]
14	F6.2	mag	Error in I
15	I1	mag	I-band quality flag[e]
16	I6	—	Row of DR7 Quasar Catalog V (out of 105783)
17	F7.3	mag	Absolute i band magnitude[g]
18	F6.4	—	redshift
19	F6.3	[$M_\odot$]	Black hole mass ("0.000" means none could be found)[f]
20	F6.3	[ergs/s]	Bolometric luminosity[f]
21	F6.3	mag	Galactic extinction in u filter (from Schlegel et al. 1998)[c]
22	I1	—	Number of observations
23	I1	—	Flag for Stripe 82 object[a]
24	I5	days	Modified Julian Date for SDSS imaging observation

Continued on Next Page...

Table A.4 – Continued

Column	Format	Units	Description
25	F6.3	mag	BEST SDSS g band PSF magnitude[b]
26	F6.3	mag	Error in g
27	F6.3	mag	BEST SDSS r band PSF magnitude[b]
28	F6.3	mag	Error in r
29	F6.3	mag	BEST SDSS i band PSF magnitude[b]
30	F6.3	mag	Error in i

[a]A value of 1 (0) indicates a (non-)Stripe 82 source.

[b]SDSS photometric measurements are asinh magnitudes (Lupton, Gunn, & Szalay 1999) and are normalized (to ∼3% accuracy) to the AB-magnitude system (Oke & Gunn 1983). Uncorrected for Galactic extinction. A value of 0.000 indicates that the value could not be retrieved from the SDSS database.

[c]Where A_g, A_r, A_i, $A_z = 0.736$, 0.534, 0.405, 0.287 $\times A_u$, respectively.

[d]Not corrected for Galactic absorption. Bad observations are printed as "-99.99".

[e]Flag indicating whether the recalibrated magnitude is of good quality (a value of 1 is "good"). See Ses06 for a description of these cuts.

[f]Taken from the Shen et al. (2011) catalog.

[g]For $\Omega_m = 0.300$, $\Omega_\Lambda = 0.700$, $h = 0.70$, and $\alpha_Q = -0.50$.

VITA

Chelsea MacLeod was born July 1, 1984 in Mobile, Alabama. After graduating from high school in North Carolina, she began studying physics at the University of North Carolina in Chapel Hill as a result of her interest in astronomy. She received her Bachelor of Science degree in physics in 2006, after participating in an undergraduate research project on the follow-up imaging of gamma-ray bursts. The discovery of the most distant explosion in the universe at that time by her research group was among the many influential experiences which prompted her to pursue astronomy as a career. She joined the graduate program in astronomy at the University of Washington in 2006, which she feels is the best decision she has made yet. She hopes to become "rich and famous" one day, as promised by her advisor.