Optimal quasar classification combining color and variability selection

some permutation of JB[1], DWH[1,2], ADM[2,3], KS[2], and others

ABSTRACT

1. Introduction

Bovy: make a point about how structure function is similar to periodograms in geology

2. Data

Might be worth introducing the Stripe 82 data samples for stars/quasars/RR Lyrae?

3. Method

There have been many recent advances in probabilistic quasar classification both in color space (Richards et al. 2004; Bovy et al. 2011a) and in the time domain (Kozłowski et al. 2010a,b; MacLeod et al. 2011b; Schmidt et al. 2010; Butler & Bloom 2011; Palanque-Delabrouille et al. 2011; Wu et al. 2011). But, in most cases, variability selection has been studied by placing hard cuts on samples in color space—or vice versa. This discards joint information, as some quasars vary significantly—but are confused with non-quasars in color space, and some objects are obvious quasars in color space in coadded data—but do not significantly vary above the noise for the, necessarily less precise, imaging extracted from a single epoch.

The goal of this paper is to formally combine and test color- and time-domain-based measures of quasar classification for the first time. This requires a probabilistic description of quasar classification in both color and time. Using Bayes' theorem our full description of the probability that an object of observed brightnesses is a quasar can be written as

[1] Center for Cosmology and Particle Physics, Department of Physics, New York University, 4 Washington Place, New York, NY 10003

[2] Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany

[3] Department of Physics and Astronomy, University of Wyoming, Laramie, WY 82071, USA

$$\begin{aligned} p(\text{quasar}|\text{fluxes}, \text{var}) &\propto p(\text{fluxes}, \text{var}|\text{quasar})p(\text{quasar}) \\ &= p(\text{fluxes}|\text{quasar})p(\text{var}|\text{quasar})p(\text{quasar}), \end{aligned} \quad (1)$$

where we seek the probability that an object is a quasar given its observed color and variability characteristics, $p(\text{quasar}|\text{fluxes}, \text{var})$, based on the likelihood of variability and color characteristics given that the object is a quasar. We use "fluxes" instead of "color" in our notation to specify that fluxes, not colors, are the actual empirical data we use to classify quasars. Historically, object classification in single-epoch data across fluxes measured in multiple bands has been associated with "color" selection for the very reason that *multiple* bands are necessary. Fluxes are simpler to use with mathematical rigor—as fluxes across multiple bands are more likely to be independent than colors (e.g., see Bovy et al. 2011a).

To separate the likelihoods in Equation 1, we adopted the weak assumption that color and variability characteristics of astrophysical objects are independent, such that we can "chain" the likelihoods together. This is certainly a simplification—but we endure it as the probabilistic approach allows us to make progress in properly testing the relative power of color and time-domain classification. In particular, the probabilistic approach can facilitate a natural reversion to color-based classification in the absence of significant variability information, and vice versa. By extension we are also assuming that quantities that are known to have a strong correlation with color—such as redshift—have a weak correlation with variability. This is the case when considering single-band variability (e.g. MacLeod et al. 2011a), but not when considering multiple bands (e.g. Schmidt et al. 2010).

For our classification in color space, we rely on likelihoods, $p(\text{fluxes}|\text{quasar})$, derived using the recent approach of Bovy et al. (2011a,b). We also adopt the prior $p(\text{quasar})$ from that series of papers, which (although not explicitly shown in our notation) is conditioned on the i-band quasar luminosity function. We refer the reader to Bovy et al. (2011a,b) for more extensive information on these points.

For the time-domain component of our classification, we need to describe the likelihood $p(\text{var}|\text{quasar}) = p(\boldsymbol{x}|\text{quasar})$. We express this as

$$p(\boldsymbol{x}|\text{quasar}) = \int \mathrm{d}A\mathrm{d}\gamma p(\boldsymbol{x}|A, \gamma)p(A, \gamma|\text{quasar}), \quad (2)$$

where the $\boldsymbol{x}$ represent *fluctuations* in quasar brightness (e.g., magnitude differences around some mean). The parameters A and γ are a (currently arbitrary) model description of how quasars vary. As we shall explore in the next few sections an appropriate fully descriptive model for $p(\boldsymbol{x}|A, \gamma)$ is a Gaussian process (Rasmussen & Williams 2006).

3.1. Quasar variability and the structure function

The structure function is the description of the mean squared difference (or, sometimes, root mean squared difference) between pairs of observations of some object's brightness (or other property) as a function of the time lag between the observations. The structure function is a description of a second-order statistic, or two-point function, of the brightness history of the source. It does not, in itself, describe how to generate such measurements, nor does it, in itself, provide a prescription for writing down the likelihood of structure function parameters. That is, it does not provide a specific description of the probability distribution function (PDF) for any particular set of observations.

Information on how objects vary can be derived from just two observations. As quasars—unlike stars—vary on decade-long timescales, there is (potentially) particularly useful empirical information in long time-baselines, e.g. the decade-long intervals between surveys such as Pan-STARRS and the SDSS. So, we choose to introduce a method to derive likelihoods from time-domain information, $p(\mathrm{var}|\mathrm{QSO})$, that can handle two-point descriptions. To also build a simple probabilistic description, we seek a generative model—we want to be able to generate brightness observations consistent with any (reasonable) set of structure function parameters, and also to assess the probability of any particular data set given those parameters. Because the Gaussian is the simplest two-point distribution function with a non-trivial variance, we build this model from a Gaussian process.

Hogg: One advantage: Can deal with the data responsibly, unlike methods that use only magnitude differences.

Hogg: One disadvantage: This model will make assumptions about the higher-point functions, which probably aren't true in detail. The "apply arithmetic operations to the data" makes no such assumptions explicitly (though of course it does not use the data optimally).

3.2. The structure function and the Gaussian process

Imagine a set of N measurements m_i, which can be for our purposes calibrated magnitudes all taken in a single photometric bandpass of a single source. We will consider multiple sources below. Each measurement m_i is taken at time t_i, and has a (presumed known) uncertainty variance σ_i^2.

The structure function $V(|\Delta t|)$ is defined as follows: The expectation value $E[\cdot]$—the expected mean value over a large number of hypothetical similar experiments—for the difference between observation m_i and m_j (when $i \neq j$) is

$$E[(m_i - m_j)^2] \quad = \quad \sigma_i^2 + \sigma_j^2 + V(|t_i - t_j|) \quad , \tag{3}$$

where the observations are presumed to be (from a measurement noise perspective) independent, and the structure function $V(\cdot)$ effectively describes the variance in excess of that expected from

the observational noise. In the literature, the structure function has occasionally been defined in terms of the root-mean-square—the square-root of the above definition—and sometimes in terms of the mean absolute difference, which is slightly different again. But we will stick with variance for notational simplicity.

A Gaussian process is characterized by a function describing the mean $\boldsymbol{m}(t)$ magnitude as a function of time t and a function $\boldsymbol{\Sigma}(t,t')$ describing the covariances between magnitudes observed at different epochs t and t'. We will assume that the mean magnitude is constant and that the process is stationary such that $\boldsymbol{\Sigma}(t,t') \equiv \boldsymbol{\Sigma}(t-t')$. The probability of a set of N observations $\{x_i\}_{i=1}^N$ is then given by that of the N-dimensional Gaussian with mean $[m,m,\ldots,m]^{\mathrm{T}}$ and N by N dimensional variance matrix $\boldsymbol{\Sigma}$ with elements $\boldsymbol{\Sigma}_{ij} = \boldsymbol{\Sigma}(t_i - t_j)$.

It is possible to define a Gaussian process that generates data in accordance with any (reasonable) structure function. The structure function os the expectation of the squared magnitude difference between observations separated by a time Δt. Thus, we can re-write it as

$$V(|\Delta t|) = E[(m(t)-m(t+\Delta t))^2] = 2\,E[(m(t)-E[m])^2] - 2\,E[(m(t)-E[m])(m(t+\Delta t)-E[m])]\,. \quad (4)$$

The variance function of the Gaussian process corresponding to the structure function V is then given by

$$\boldsymbol{\Sigma}(\Delta t) = 2\,V(\infty) - 2\,V(\Delta t)\,. \quad (5)$$

Variability models can therefore be expressed either in terms of the variance function or equivalently in terms of a structure function.

For example, imagine that the two quantities m_i and m_j of equation (3) are not observations of a quasar but instead random numbers drawn from an N-dimensional Gaussian

$$\begin{aligned}
p(\boldsymbol{x}) &= \mathscr{N}(\boldsymbol{x}|\boldsymbol{m},\boldsymbol{\Sigma}) \\
\boldsymbol{x} &= \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ \cdots \\ m_N \end{bmatrix}, \quad \boldsymbol{m} = \begin{bmatrix} m \\ m \\ m \\ \cdots \\ m \end{bmatrix} \\
\boldsymbol{\Sigma} &= \begin{bmatrix} (V_\infty + \sigma_1^2) & (V_\infty - V_{12}) & (V_\infty - V_{13}) & \cdots & (V_\infty - V_{1N}) \\ (V_\infty - V_{21}) & (V_\infty + \sigma_2^2) & (V_\infty - V_{23}) & \cdots & (V_\infty - V_{2N}) \\ (V_\infty - V_{31}) & (V_\infty - V_{32}) & (V_\infty + \sigma_3^2) & \cdots & (V_\infty - V_{3N}) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ (V_\infty - V_{N1}) & (V_\infty - V_{N2}) & (V_\infty - V_{N3}) & \cdots & (V_\infty + \sigma_N^2) \end{bmatrix} \\
V_{ij} &\equiv \frac{1}{2} V(|t_i - t_j|) \quad , \qquad (6)
\end{aligned}$$

where we have assembled the observations into a column vector $\boldsymbol{x}$, $\mathscr{N}(\cdot|\boldsymbol{m},\boldsymbol{\Sigma})$ is the general normal or Gaussian PDF given mean $\boldsymbol{m}$ and variance tensor $\boldsymbol{\Sigma}$, m is an arbitrary parameter, V_{ij} is the

structure function evaluated at time lag $|t_i - t_j|$, and V_∞ is $V(\infty)$. If we make many draws from this Gaussian, the expectation values of $(m_i - m_j)$ and $(m_i - m_j)^2$ for any pair of measurements m_i and m_j (with $i \neq j$), are just

$$\begin{aligned} E[(m_i - m_j)] &= 0 \\ E[(m_i - m_j)^2] &= \sigma_i^2 + \sigma_j^2 + 2V_{ij} \quad , \end{aligned} \tag{7}$$

which, by design, is equivalent to the description of the structure function in Equation (3).

Two additional points arise from this description. Although $\boldsymbol{m}$ does not enter into the prediction of the mean or variance of the magnitude *difference*, it does affect the magnitudes themselves, so it is, in principle, an observable property of the model. However, in what follows we will fix the mean magnitude of each object to the mean of the lightcurve. When fit independently the best-fit mean magnitude are typically $\lesssim 0.1$ mag different from the mean of the lightcurve and, more importantly, inferences about the structure function parameters are unaffected. Also, although V_∞ is not, in practice, measurable, it can be approximated by evaluating the structure function at large time lag.

3.3. Model Expressions for the Structure Function

So far, we have expressed the structure function V as arbitrary. In fact we expect the structure function to have—or model it as having—a simple parameterized form. In this paper we deal with the two most popular recent forms in the literature. A power-law form

$$V(|\Delta t|) = A^2 \left(\frac{\Delta t}{1\mathrm{yr}} \right)^\gamma \quad , \tag{8}$$

(e.g., Schmidt et al. 2010) for dimensionless parameters A and γ. We also consider a damped random walk model, for which the variance function of the Gaussian process has an exponential form

$$\boldsymbol{\Sigma}(|\Delta t|) = \frac{\sigma^2}{2} e^{-\left(\frac{|\Delta t|}{\tau}\right)} \quad , \tag{9}$$

(e.g., Butler & Bloom 2011) where τ is a damping timescale and σ^2 is the intrinsic variance of the process (equivalently, the structure function for infinite time lags).

After parameterizing the structure function, the complete model—the Gaussian process with mean $\boldsymbol{m}$ and variance $\boldsymbol{\Sigma}$ introduced in §3.2 —for any set of observations is specified by only three parameters: either (m, A, γ) or (m, σ, τ). Thus, in turn, the likelihood $p(\boldsymbol{x}|A, \gamma)$ in Equation 2 can be described by 3 parameters as

$$p(\boldsymbol{x}|A, \gamma) = \mathscr{N}(\boldsymbol{x}|\boldsymbol{m}, \boldsymbol{\Sigma}) \quad , \tag{10}$$

with V here expressed as a function of (A, γ) as that is the parameterization we chose for Equation 2—but $(A, \gamma) \rightarrow (\tau, \sigma)$ without loss of generality. The term $\mathscr{N}(\boldsymbol{x}|\boldsymbol{m}, \boldsymbol{\Sigma})$ is the Gaussian process outlined in Equation 6.

3.4. Generating fake lightcurves

The formulation in terms of a Gaussian process allows one to generate a lightcurve for an object with a model structure function. The fundamental property of a Gaussian process is that all of its marginal distributions—marginalizing over unobserved times—are Gaussian, with the mean and variance of this Gaussian given by the Gaussian process' mean and variance functions. Generating a fake lightcurve is then as simple as sampling from the appropriate Gaussian distribution (Rasmussen & Williams 2006).

Given a structure function $V(|\Delta t|)$ and the parameters m and A, we can generate a lightcurve at N times $\{t_i\}_{i=1}^N$ by sampling from the N-dimensional Gaussian with mean and variance given by equation (6). In the left panels of Figure **??** five lightcurves are generated for the power-law structure function of equation (8) with parameters that represent the variability of the quasar SDSSJ203817.37+003029.8. The lower panel shows that these lightcurves have the right power-law structure.

BOVY: UPDATE THESE FIGURES

In addition to making fake data that is consistent with a given structure function, we can also generate lightcurves constrained by existing observations $\{(t_j^e, m_j^e, \sigma_j^{2,e})\}_{j=1}^M$ while still keeping the desired structure function. To do this, we write down the joint distribution of the magnitudes of the fake data and the existing data. This distribution is an $N+M$-dimensional Gaussian with mean and variance with structure

$$\begin{aligned} \boldsymbol{m} &= [\boldsymbol{m}_i\, \boldsymbol{m}_j]^{\mathrm{T}}\,, \\ \boldsymbol{\Sigma} &= \left[\begin{array}{cc} \boldsymbol{\Sigma}_{ii} & \boldsymbol{\Sigma}_{ij} \\ \boldsymbol{\Sigma}_{ij}^{\mathrm{T}} & \boldsymbol{\Sigma}_{jj} \end{array}\right], \end{aligned} \tag{11}$$

where each element of both $\boldsymbol{m}_i$ and $\boldsymbol{m}_j$ is m. $\boldsymbol{\Sigma}_{ii}$ is calculated as in equation (6) with times $\{t_i\}_{i=1}^N \otimes \{t_i\}_{i=1}^N$, $\boldsymbol{\Sigma}_{ij}$ with times $\{t_i\}_{i=1}^N \otimes \{t_j^e\}_{j=1}^M$, and $\boldsymbol{\Sigma}_{jj}$ with times $\{t_j^e\}_{j=1}^M \otimes \{t_j^e\}_{j=1}^M$. The fake and existing data are similarly placed in a vector

$$\boldsymbol{m} = [\boldsymbol{m}^e\, \boldsymbol{m}^f] = [m_1^f \ldots m_N^f\, m_1^e \ldots m_M^e]^{\mathrm{T}}\,, \tag{12}$$

where the m_i^f are the fake magnitudes. To generate the fake data constrained by the existing observations, we then draw from the N-dimensional Gaussian obtained by conditioning the $N+M$-dimensional Gaussian in equation (11) on the M observations. This constrained Gaussian has mean and variance (e.g., Bovy, Hogg, & Roweis 2009)

$$\begin{aligned} \boldsymbol{m} &= \boldsymbol{m}_i + \boldsymbol{\Sigma}_{ij}\, \boldsymbol{\Sigma}_{jj}^{-1}\, (\boldsymbol{m}^e - \boldsymbol{m}_j)\,, \\ \boldsymbol{\Sigma} &= \boldsymbol{\Sigma}_{ii} - \boldsymbol{\Sigma}_{ij}\, \boldsymbol{\Sigma}_{jj}^{-1}\, \boldsymbol{\Sigma}_{ij}^{\mathrm{T}}\,. \end{aligned} \tag{13}$$

In the right panels of Figure **??** five fake lightcurves are shown that are constrained by the data for the quasar SDSSJ203817.37+003029.8. The lower panel shows again that these lightcurves have the right power-law structure.

3.5. The Structure Function of known objects

So far, we have concerned ourselves with the likelihood $p(\boldsymbol{x}|A,\gamma)$ in Equation 2. Namely, generating samples for a range of structure function parameters to test against observed light curves for an (ultimately unclassified) varying object. We demonstrate how these samples populate the structure function plane in Fig. 1 where we plot the posterior probability distribution function $p(A,\gamma|\boldsymbol{x})$ for two illustrative quasars. We use both a slice sampling MCMC technique (Neal 2003) and a faster affine invariant ensemble MCMC sampler (Goodman & Weare 2010). We assume an uninformative prior that is flat in γ and $\log A$ to make this figure—so $p(\log A,\gamma|\boldsymbol{x}) \propto p(\boldsymbol{x}|logA,\gamma)$. These posterior distribution functions show that the uncertainty in the inferred structure function parameters is large, even for the relatively well-sampled stripe-82 lightcurves.

The other term in Equation 2 $p(A,\gamma|\mathrm{quasar})$ is the likelihood of a set of structure function parameters given that an object is known to be a quasar. This term can be derived by using our formalism to determine the distribution of structure function parameters for all known quasars. In Figures 2 and 3 we display this distribution for known quasars in Stripe 82 (see §2). Also plotted are $p(A,\gamma|\mathrm{star})$ and $p(A,\gamma|\mathrm{RR\,Lyrae})$. In the right-hand panel of each figure we show the underlying, deconvolved distribution of $p(A,\gamma)$ for each class, which is the likelihood in Equation 2. This deconvolution is obtained by approximating each object's posterior distribution function for A and γ as a Gaussian distribution in $\log A$ and $\log\gamma$ and using the extreme deconvolution technique to model the underlying distribution as a sum of 5 Gaussian components for quasars, 1 component for stars, and 2 components for RR Lyrae. As is already known (Schmidt et al. 2010; Butler & Bloom 2011) quasars, stars and RR Lyrae separate well in variability space—given the ~ 50 epochs of observations available in Stripe 82—although some overlap can remain, depending on the errors on brightness fluctuations in a particular set of observations. The deconvolved distributions are similar to the observed distributions, except that the low-γ tail for quasars as good as disappears.

3.6. A simple model to help distinguish quasars from RR Lyrae

Part of the power of combining color- and variability-based selection for quasars (and other astronomical objects) is that ambiguous sources that overlap multiple classes of object in the time domain may not overlap in color space, and vice versa. This is not just an effect of position in color or variability space, but also the relative sizes of measurement errors in these spaces. For instance, colors derived from fluxes co-added over multiple-epochs will always be more precise than individual, single-epoch fluxes used for classification in the time domain.

One illustrative example is RR Lyrae. As shown in Figures 2 and 3 some fraction of RR Lyrae overlap the quasar locus in both (A,γ) and (σ,τ) space. But, RR Lyrae are red in (e.g.) $u-g$ color whereas the vast majority of quasars at low redshift ($z < 2.5$) are blue. So, with correct probabilistic classification, far fewer quasars should be confused with RR Lyrae if both colors and variability are considered. There will still, always, be a small number of RR Lyrae that cannot be

discriminated from quasars (around redshift $z \sim 3$) in both variability and color space. But, this number is certainly smaller than the overlap in variability space alone.

As noted near Equation 1, our intention in this paper is to use the color classification scheme of Bovy et al. (2011a,b). That series of papers considered all stars in bulk, i.e., did not separate out RR Lyrae as a special case in color space. To improve our model in this paper, we adopt a very simple likelihood description $p(\text{fluxes}|\text{RR Lyrae})$ for RR Lyrae in color space using the u and g bands

$$\begin{aligned} p(u-g \geq 1|\text{RR Lyrae}) &= 1 \\ p(u-g < 1|\text{RR Lyrae}) &= 0 \quad , \end{aligned} \tag{14}$$

essentially reverting to the variability-based probability in regions of color space RR Lyrae can occupy and reverting to 0 in regions of color space that RR Lyrae do not populate.

ADM: This might not quite be the correct cut. Update when I determine it.

3.7. A probabilistic combination of color and variability classification

Let us define the likelihood that an object is a quasar based on its colors to be $\mathcal{L}^c_{\rm QSO} = p(\text{fluxes}|\text{quasar})$, and define the likelihood that an object is a quasar based on its variability to be $\mathcal{L}^v_{\rm QSO} = p(\text{var}|\text{quasar})$. Using equivalent notation for stars and RR Lyrae, our final classification probability that on object is a quasar is

$$p(\text{quasar}|\text{fluxes, var}) = \frac{\mathcal{L}^c_{\rm QSO}\mathcal{L}^v_{\rm QSO}p(\text{QSO})}{\mathcal{L}^c_{\rm QSO}\mathcal{L}^v_{\rm QSO}p(\text{QSO}) + \mathcal{L}^c_{\rm star}\mathcal{L}^v_{\rm star}p(\text{star}) + \mathcal{L}^c_{\rm RR}\mathcal{L}^v_{\rm RR}p(\text{RR})} \quad . \tag{15}$$

Most of the terms in Equation 15 have already been defined. The likelihood of being a quasar in the time domain $\mathcal{L}^v_{\rm QSO} = p(\text{var}|\text{quasar})$ is expressed in Equation 2, and an identical formalism holds for stars and RR Lyrae. Defining these terms was the main purpose of §3.

The color-based likelihoods $\mathcal{L}^c_{\rm QSO} = p(\text{fluxes}|\text{quasar})$ and $\mathcal{L}^c_{\rm star} = p(\text{fluxes}|\text{star})$ are direct outputs of the models in Bovy et al. (2011a). We defined a simple color-based likelihood for RR Lyrae stars $\mathcal{L}^c_{\rm RR}$ in Equation 14.

The priors $p(\text{QSO})$ and $p(\text{star})$ are also outputs of the Bovy et al. (2011a) model, based respectively on the i-band quasar luminosity function and the i-band number counts of stars. All that remains to define is the prior on RR Lyrae, which we define as

$$p(\text{RR}) = \alpha \text{p(star)} \quad . \tag{16}$$

Where α is simply the fraction of stars that are RR Lyrae.

As a final simplification, we note that we only consider our variability model in a single band, choosing the r band. Until quasars drop out of the r band at very high redshift, variability in just r should adequately represent how stars and quasars vary in every optical band.

4. Results

Simple discussion of how well we re-classify the stripe 82 quasars.

5. Discussion

5.1. Which is "better", color or variability selection?

include Y-band so we can look at grizY.

5.2. How many epochs are necessary to separate quasars and stars?

5.3. τ, σ or A, γ?

6. Conclusions and Future Directions

ADM: Now we have Pan-STARRS long-baselines, we could augment the structure function using a set of samplings of the two-epoch SDSS-pan-STARRS variability. 120,000 quasars at a a range of redshifts, so large number of rest-frame time lags. Could use hierarchical Bayes to probabilistically infer the structure function parameters.

ADM: We're ignoring information, because we determine P(color) and P(variability) and multiply them. The assumption that variability and color are independent is probably inaccurate. Should really invert the full matrix that includes cross-correlation band-matrix.

ADM: Star-galaxy separation and continuing to "chain likelihoods" in Eqn. 1. Note again that this is not optimal (you really want a joint probability without the assumption of independence).

ADM: Cadence redshifts. Slight shifts in line positions scatter in and out of filters as a function of color, break degeneracy in some photometric redshifts (essentially redo this paper but multiply by xdqsoz instead of xdqso). There's a little bit of a g-sigma relation in Butler & Bloom, but I think we're talking about something different.

ADM: Temporal-spatial structure functions. Variability as a function of position as well as time so that a galaxy with a resolved AGN source at the center looks different from a more passive galaxy and looks different from, e.g., a pair of QSOs that are merged in imaging.

Thanks to people. Thanks to *SDSS*. Thanks to *Pan-STARRS*. Thanks to *Python*. Thanks to grants.

REFERENCES

Bovy, J., et al. 2011a, ApJ, 729, 141

Bovy, J., et al. 2011b, arXiv:1105.3975

Bovy, J., Hogg, D. W., & Roweis, S. T. 2009, arXiv:0905.2979v1 [stat.ME]

Butler, N. R. & Bloom, J. S., 2010, AJ, 141, 93

Goodman, J. & Weare, J., 2010, Comm. App. Math. and Comp. Sci., 5, 65

Kozłowski, S., et al., 2010a, ApJ, 708, 927

Kozłowski, S., et al., 2010b, ApJ, 716, 530

MacLeod, C. L., et al., 2011a, ApJ, 721, 1014

MacLeod, C. L., et al., 2011b, ApJ, 728, 26

Neal., R. M., 2003, Ann. Stat., 31, 705

Palanque-Delabrouille, N., et al. 2011, A&A, 530, A122

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

Rasmussen, C. E. & Williams, C., 2006, Gaussian Processes for Machine Learning (MIT Press)

Richards, G. T., et al., 2004, ApJS, 155, 257

Rybicki, G. B. & Kleyna, J. T., 1994, in Reverberation Mapping of the Broad-Line Region in Active Galactic Nuclei, P. M. Condhalekar, K. Horne, & B. M. Peter son (eds.), ASPCS, 69, 85

Schmidt, K. B., et al., 2010, ApJ, 714, 1194

Schmidt, K. B., et al., 2011, ApJ, submitted

Schneider, D. P., et al., 2010, AJ, 139, 2360

Wu, X.-B., Wang, R., Schmidt, K. B., Bian, F., Jiang, L., & Fan, X. 2011, arXiv:1107.0646

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

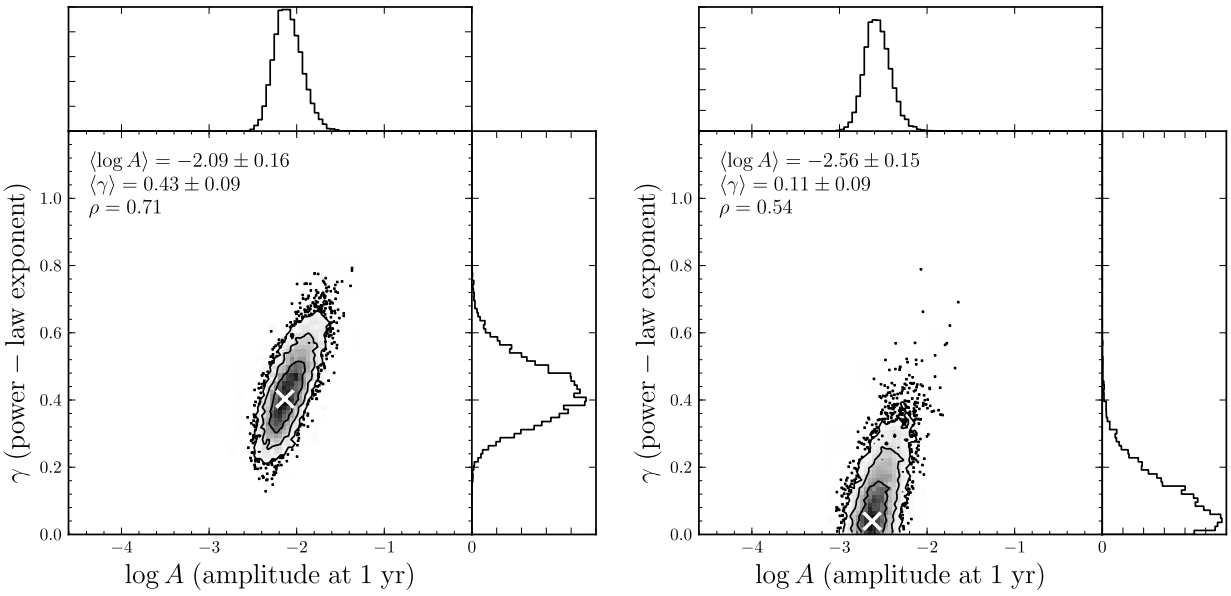

Fig. 1.— Example posterior probability distribution functions for the parameters of the r-band power-law structure-function quasar variability parameters for a typical (SDSSJ013435.82-000511.3; *left*) and low-variability (SDSSJ033317.79+005429.6; *right*) quasar on *SDSS* imaging stripe-82. The mean ± standard deviation for each parameter and correlation ρ between the parameters are given in the upper-left corner. The white cross indicates the best-fit parameters. Grayscales are linear and contours contain 68, 95, and 99 percent of the distribution; outliers beyond 99 percent are individually shown.

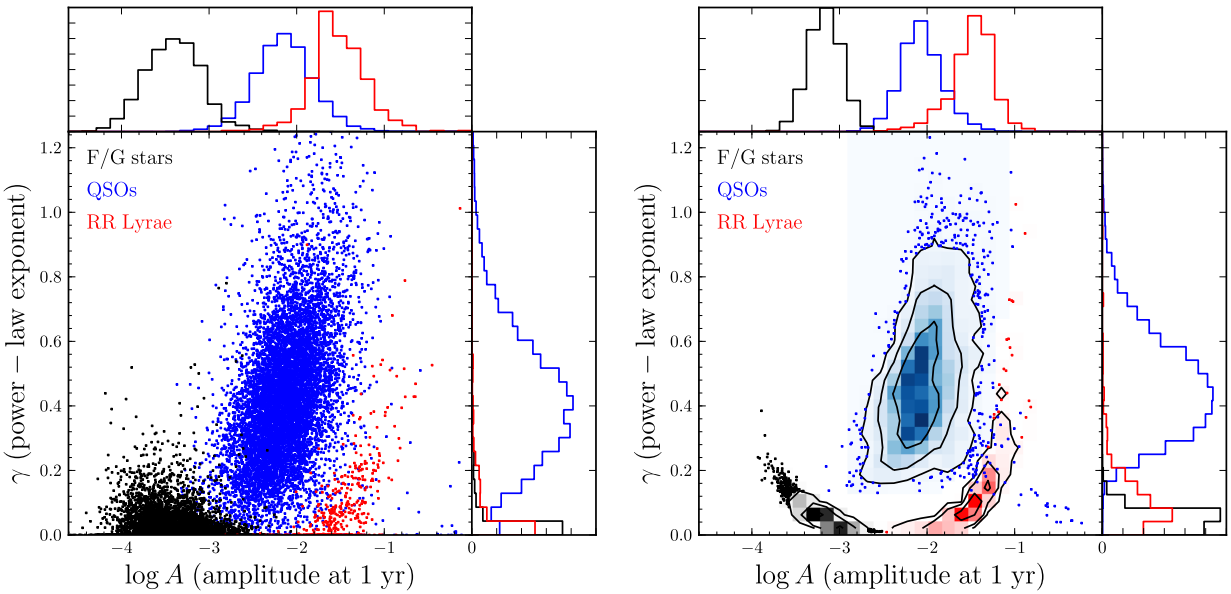

Fig. 2.— Observed (*left*) and underlying (*right*) distributions for the parameters of the r-band power-law structure-function variability model. The color scales in the right panel are linear and contours contain 68, 95, and 99 percent of the distribution; outliers beyond 99 percent are individually shown.

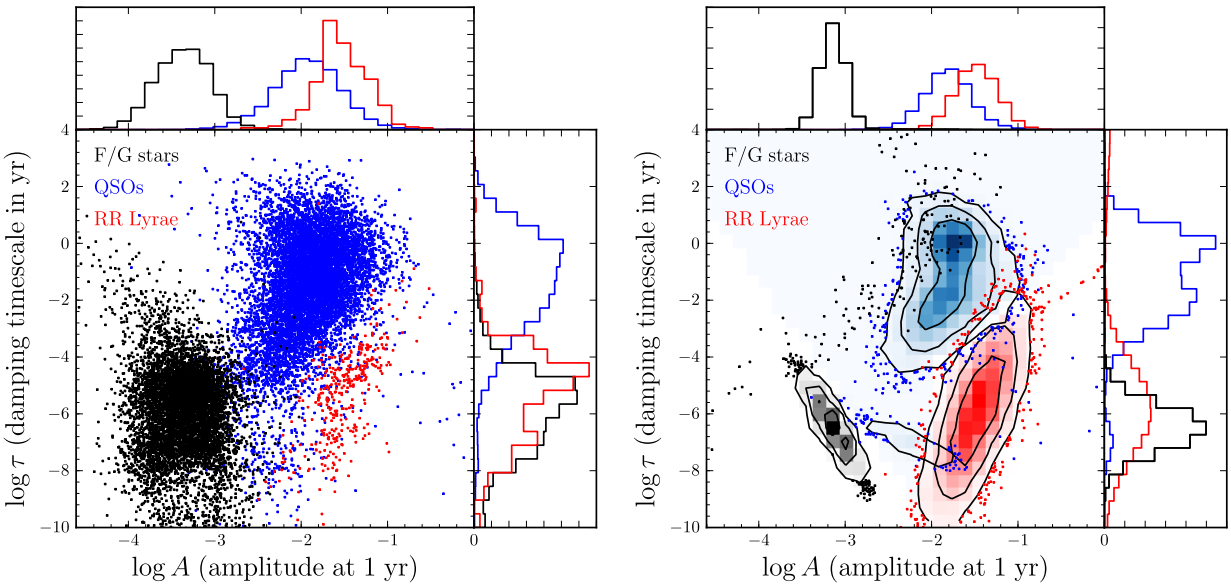

Fig. 3.— Same as Figure 2 but for the damped-random-walk variability model.