Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed 3 June 2008 (MN LATEX style file v2.2)

Clustering of Low-Redshift ($z \leq 2.2$) Quasars from the Sloan Digital Sky Survey

Nicholas P. Ross,[1⋆] Yue Shen,[2] Daniel E. Vanden Berk,[1] Andrew J. Connolly,[3] Gordon T. Richards,[4] Donald P. Schneider,[1] Michael A. Strauss,[2] Patrick B. Hall[5] Neta A. Bahcall[2] et al.[6]

[1] *Department of Astronomy and Astrophysics, The Pennsylvania State University, 525 Davey Laboratory, University Park, PA 16802, U.S.A.*
[2] *Princeton University Observatory, Princeton, NJ 08544, U.S.A.*
[3] *Department of Astronomy, University of Washington, Box 351580, Seattle, WA 98195, U.S.A.*
[4] *Department of Physics, Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104, U.S.A.*
[5] *Department of Physics and Astronomy, York University, Toronto, ON M3J 1P3, Canada*

3 June 2008

ABSTRACT

We present results for the Quasar 2-Point Correlation Function, ξ_Q, over the redshift range $0.3 \leq z \leq 2.2$, using data from the Sloan Digital Sky Survey (SDSS). Using nearly 50,000 quasars with spectroscopic redshifts from the Data Release 5 Quasar Catalogue, our study represents the largest sample used for this type of investigation to date. With our redshift range and the areal coverage of ≈5,700 deg^2, we sample nearly 40 h^{-3} Gpc^3 (comoving) of the Universe in volume, assuming the current ΛCDM cosmology.

Over our redshift range, we find that the redshift-space correlation function, $\xi(s)$, is not described well by a traditional single-power law, and instead fit a double-power law. Here, for the redshift-space distortions, parametrized by a line-of-sight *rms* velocity dispersion, $\langle w_z^2 \rangle^{1/2} = 800$ km s^{-1} and the dynamical infall parameter $\beta(z = 1.4) = 0.32$, provides a good fit to the data, over the scales 1 h^{-1} Mpc $\leq s \leq 80$ h^{-1} Mpc. We present the 2-D redshift-space correlation function, $\xi(\sigma, \pi)$, for SDSS quasars for the first time, though acknowledge that accuracy of quasar redshift determination will limit the amount of cosmological information available here. We test for potential systematics in our data and find that we are unbiased to various effects due to the high completeness and efficient colour-radio selection of our quasar sample.

Using the projected correlation function, $w_p(\sigma)$ we calculate the real-space correlation length, r_0, and the linear bias, b, where the bias is given by the simple model, $\xi_Q = b^2 \xi_{\rm matter}$. Using our results with simple halo-mass:bias models, we find that the bias evolves from $b(z \simeq 2) \sim 4$ to $b(z \simeq 0.30) = 2.25$. These values suggest that, if the $M_{\rm BH} - M_{\rm DMH}$ and $M_{\rm BH} - \sigma$ relationships are *redshift independent*, then quasars inhabit dark matter haloes of mass $M_{\rm halo} \sim 3 \times 10^{12} h^{-1} M_\odot$ at redshifts around $z = 2$ (the peak of quasar activity) and similar mass haloes at redshift ~ 0, i.e. those of luminous, Red Sequence galaxies today. This result continues to add to recent evidence and suggest that all galaxies which harbour black holes at the centre of their bulges, went through an AGN/quasar phase in their past.

Key words: clustering – quasars: general – cosmology: observations – large-scale structure of Universe. observations – Sky Surveys

1 INTRODUCTION

Understanding how and when the structures we see in the local Universe formed from the initial conditions present at the time of Inflation, is one of the fundamental goals of modern observational cosmology. By tracing the evolution of clus-

⋆ Email: npr@astro.psu.edu

tering, we have the potential to understand the growth of structure and its relation to the energy and matter content of the Universe, including the relationship between the dark matter and the luminous galaxies and quasars that we observe.

As such, one of the primary science goals of the Sloan Digital Sky Survey (SDSS; York et al. 2000) was to measure the properties of large-scale structure (LSS) and, in particular, one of the key science goals of the SDSS Quasar survey, was to determine the spatial clustering of quasars as a function of redshift. Shen et al. (2007) report on the clustering of high, $z \geq 2.9$ quasars from the SDSS and in this paper, we shall report on the spatial clustering from redshift $z = 2.2$ to the present day, i.e. the evolution of quasar clustering over nearly 80% of the age of the Universe.

Due to their high intrinsic luminosity, quasars are seen to large cosmological distances, and are thus good probes to investigate LSS and its evolution. However, until recently, quasar studies we plagued by low-number statistics, leading to shot noise, and investigations over small areas of sky that could be subject to sample variance. With the advent of large-area ($\gtrsim 1000$ deg^2) surveys, with efficient selection techniques, these limitations have been overcome and the number of known quasars has increased by over a magnitude in the last decade, thanks in the main part to the 2dF QSO Redshift Survey (2QZ; Boyle et al. 2000; Croom et al. 2004) and the SDSS. At present, there are over 100 000 objects spectroscopically classified as quasars in the SDSS, with the latest quasar catalogue from Schneider et al. (2007) being nearly 80 000 objects strong. Thus, using the data from these large surveys, we are now in a position to make high-precision measurements of quasar clustering properties.

The 2-Point Correlation Function, ξ, is a simple but powerful statistic commonly employed to describe LSS and is used to quantify the clustering properties of a given object (Peebles 1980). The observed value of ξ for a given, luminous, tracer is related to the underlying (dark) matter density distribution via

$$\xi(r)_{\rm quasar} = b_Q^2 \; \xi(r)_{\rm matter} \qquad (1)$$

where $\xi(r)_{\rm matter}$ is the mass correlation function and b_Q is the linear bias parameter for quasars. Although equation 1 is only a model and we do not know *a priori* if the bias has e.g a scale-dependence, under certain assumptions, the measurement and interpretation of this bias can lead to determination of the dark matter halo properties of quasars and potentially quasar lifetimes (t_q, Martini & Weinberg 2001).

This is due to the link in the standard scenario, that quasar activity is triggered by accretion on to a central, supermassive black hole (SMBH, e.g. Hopkins et al. 2006). Then, given that the growth of the SMBH relates to that of the underlying dark matter halo (Baes et al. 2003; Wyithe & Loeb 2005; Wyithe & Padmanabhan 2006) and the halo properties are correlated with the local density contrast, clustering measurements provide an insight into quasar and black hole physics (da Ângela et al. 2008). This includes constraining η, the fraction of the Eddington Luminosity at which quasars shine, and the constant ϵ, which is related to the halo density profile (Wyithe & Loeb 2005).

Early measurements of the quasar 2PCF, (e.g. Arp 1970; Hawkins & Reddish 1975; Osmer 1981; Shanks et al. 1983) found signal for an excess of clustering of quasars on separations $\sim$ a few h^{-1} Mpc scales with both other quasars (the auto-correlation) and surrounding galaxies (cross-correlation). This result has been confirmed with data from the recent surveys, e.g. Croom et al. (2005); Porciani et al. (2004) with the Quasar 2PCF being traditionally described by a single power-law of the form,

$$\xi(r) = (r/r_0)^{-\gamma} \qquad (2)$$

over the range 1 h^{-1} Mpc $\leq r \leq$ 100 h^{-1} Mpc, where r_0 is the correlation length and γ the power-law slope.

However, the evolution of the quasar correlation function is more disputed, with some authors claiming that quasar clustering either decreases or only weakly evolves with redshift (Iovino & Shaver 1988; Croom & Shanks 1996) while others say it increases with redshift, (Kundic 1997; La Franca et al. 1998; Porciani et al. 2004; Croom et al. 2005; Shen et al. 2007). Shen et al. (2007), the high-redshift companion study to this work, found that redshift $2.9 \leq z \leq 5.4$ quasars are significantly more clustered than their $z \sim 1.5$ counterparts, having a real-space correlation length and power-law slope of $r_0 = 15.2 \pm 2.7$ h^{-1} Mpc and $\gamma = 2.0 \pm 0.3$, respectively, over the scales 4 h^{-1} Mpc $\leq \sigma \leq$ 150 h^{-1} Mpc (where σ is the separation from the projected correlation function, $w_p(\sigma)$). Shen et al. (2007) also find that bias increases with redshift, with, $b_Q \sim 8$ at $z = 3.0$ and $b_Q \sim 16$ at $z = 4.5$. These values suggest quasar lifetimes of $t_q \sim$0.4-5$\times 10^7$ yr at $2.9 \leq z \leq 3.5$ and potentially longer, $t_q \sim$3-60$\times 10^7$ yr, for quasars at $z \geq 3.5$.

Myers et al. (2006, 2007) examined the clustering of quasar candidates over ~ 50 h^{-1}kpc to ~ 20 h^{-1} Mpc scales using photometrically identified catalogues and found that the linear bias, b_Q increases with redshift, from $b_Q = 1.93$ at redshifts $0.4 \leq z < 1.0$ to $b_Q = 2.84$ at $2.1 \leq z < 2.8$. This increase in bias with redshift, suggests that ξ_Q is either constant, or more weakly evolving with redshift, than $\xi_{\rm m}$, since $\xi_{\rm m}$ is always expected to grow as z decreases due to continued gravitational growth of structure. Padmanabhan et al. (2008) measure the clustering of photometrically selected luminous red galaxies (LRGs) around a low, $0.2 < z < 0.6$, redshift sample of quasars, with both LRG and quasar samples coming from the SDSS. They determine a large-scale quasar bias $b_Q = 1.09 \pm 0.15$ at a median redshift of 0.43, and observe no redshift or luminosity evolution. This bias value corresponds to a mean halo mass of $M_{\rm halo} \sim 10^{12} h^{-1} M_\odot$, Eddington ratios from 0.01 to 1 and quasar lifetimes of less than 10^7 yrs. After taking into account measurement and interpretation subtleties, the results from Padmanabhan et al. (2008), are in good agreement with those from Serber et al. (2006), who find that $M_i \leq -22$, $z \leq 0.4$ quasars are located in higher local overdensitites than typical L^* galaxies. This result however, is a local, $\lesssim 100$ h^{-1}kpc, phenomenon suggesting a picture in which quasars typically reside in L^* galaxies, but have a local excess of neighbors within $\sim 0.15 - 0.7$ h^{-1} Mpc. This local density excess is likely to contribute to the triggering of quasar activity through mergers and other interactions.

Recently, Ivashchenko (2007) used just under 47,000 Quasars from the SDSS DR5 to investigate luminosity-dependent clustering. The results from this work were that quasar clustering strength is only a weak function of luminosity, with brighter quasars tending to be more clustered.

However, it is not immediately clear from this work whether the subtleties of the SDSS Quasar survey (selection functions and so forth) were properly taken into account. Also, these authors did not investigate the evolution of the SDSS Quasar 2PCF. Both these issues we deal with in detail here.

Finally, we note that data from the SDSS Quasar survey can also be used in constraining cosmological parameters, such as in Yahata et al. (2005). These authors use just under 20,000 quasars and measure the 2PCF over $8\ \ h^{-1}$ Mpc $< s < 500\ \ h^{-1}$ Mpc scales and see distortions in the overall shape of the 2PCF, due to the non-zero baryon content of the Universe. From Assuming scale-independent linear bias, they find that 0.80 - $2.8\Omega_{\rm b} < \Omega_\Lambda < 0.90 - 1.4\Omega_{\rm b}$ where $\Omega_{\rm b}$ and Ω_Λ are the baryon density and value of the Cosmological Constant in terms of the critical density, $\Omega_{\rm critical} = \rho/\rho_{\rm critical} = 8\pi G\rho/3H^2$, respectively.

In this paper, we shall measure quasar clustering properties for redshifts $z \leq 2.2$, via the 2-point correlation function, using the largest sample of spectroscopically identified quasars to date. We do this in order to investigate whether quasar clustering evolves with redshift, and to test current galaxy formation and evolution theoretical models, which make strong predictions for the evolution of e.g. linear bias, with redshift. Furthermore, with this information we will aim to potentially constrain cosmological parameters via redshift-space distortions (discussed in Section 5).

As such, this paper is organised as follows. In Section 2 we describe our data sample in detail, investigating several effects that could cause potential systematic measurement issues. In Section 3 we describe the techniques involved in measuring the 2-point correlation function and the simple models that can be used to parametrize its evolution with redshift. In Section 4 we present our results. In Section 5 we discuss the redshift-space distortions apparent in our clustering measurement and in Section 6 we compare and contrast our evolutionary results with contemporary studies. We conclude in Section 7 and add two detailed appendices for supplementary material. This paper is a compliment to Shen et al. (2007), a follow-up to Yahata et al. (2005), a potential check on Ivashchenko (2007) and a basis for Shen et al. (2008, in prep.).

We assume the currently preferred flat, ΛCDM cosmology where $\Omega_{\rm b} = 0.042$, $\Omega_{\rm m} = 0.237$, $\Omega_\Lambda = 0.763$ (Sánchez et al. 2006; Spergel et al. 2007) and quote distances in h^{-1} Mpc to aid in ease of comparisons with previous results in the literature, with all distances herein are given in comoving coordinates. Where a value of Hubble's Constant is assumed e.g. for Absolute Magnitudes, this will be quoted explicitly and our magnitudes are based on the AB zero-point system.

2 DATA

In this section we shall describe our data. Much care has to be taken when constructing a dataset that is valid for a statistical analysis and we use this section, along with Appendix A and Appendix B to describe the various samples used to investigate potential systematic effects in our clustering measurements.

2.1 The Sloan Digital Sky Survey

The SDSS uses a dedicated 2.5m wide-field telescope (Gunn et al. 2006) to collect light for 30 2k×2k CCDs (Gunn et al. 1998) over five broad bands - *ugriz* (Fukugita et al. 1996) - in order to image $\sim \pi$ steradians of the sky. The imaging data are taken on dark photometric nights of good seeing (Hogg et al. 2001) and are calibrated photometrically (Smith et al. 2002; Ivezić et al. 2004; Tucker et al. 2006), astrometrically (Pier et al. 2003), and object parameters are measured (Lupton et al. 2001; Stoughton et al. 2002).

Using the imaging data, quasar target candidates are then selected for spectroscopic follow-up as described in Richards et al. (2002). A tiling algorithm then assigns these candidates to specific spectroscopic plates, in order to maximise target completeness (Blanton et al. 2003). Each 3° spectroscopic plate holds 640 fibres and quasar candidates are allocated approximately 18 fibers deg^{-2}.

The primary quasar candidates (where we define the primary sample below) were given targetting priority over the MAIN and Luminous Red Galaxy (LRG) survey targets (Strauss et al. 2002; Eisenstein et al. 2001, respectively). Targetting priority is required since no two fibres can be placed physically closer than 55" (corresponding to $\sim 0.7\ \ h^{-1}$ Mpc at $\langle z \rangle = 1.27$, the mean redshift of our sample). Thus, excluding subtle effects due to gravitational lensing, (Scranton et al. 2005; Mountrichas & Shanks 2007), the LSS 'footprint' of these foreground galaxies should not affect our LSS quasar measurements. We also note, however, that some targets e.g. brown dwarf candidates, were given priority ahead of even the primary quasar candidates. However, since the surface density of these Galactic objects was very low ($\ll$ a few deg^{-2}), this should not have any significant impact on our results. We do, however, check to see how other fibre placements issues, i.e. fibre collisions, affect the quasar 2PCF.

For our analysis, we use the Data Release Five (DR5; Adelman-McCarthy et al. 2007), which has an imaging footprint of 8 000 deg^2. There are 77 429 confirmed quasars over the 5 740 deg^2 spectroscopic DR5 footprint, reported in the latest version of the quasar catalogue (DR5Q; Schneider et al. 2007). The 65 660 DR5Q quasars with redshifts $z \leq 2.2$ will be the parent sample we use in this investigation. The $z \leq 2.2$ limit is set due to the fact that at this redshift, the "ultra-violet excess" (UVX) method of selecting quasars begins to fail due to the Lyα emission-line moving into the SDSS u-band. This number of quasars represents a factor of > 2 increase in objects over the previous largest quasar survey, the 2QZ (Boyle et al. 2000; Croom et al. 2005).

2.2 Target Selection

Full details of quasar selection are given in Richards et al. (2002). There are several ways a quasar candidate can be selected for spectroscopic follow-up, but here we are concerned with only those quasars selected as primary targets, i.e. the main quasar selection. Quasar candidates are selected via their non-stellar colours in the *ugriz* imaging and by matching unresolved sources to the FIRST radio catalogues. Low-redshift, $z \lesssim 3$, quasar targets are selected based on their location in *ugri*-colour space and the high-redshift, $z \gtrsim 3$, objects in *griz*-colour space. The quasar candidate sample

Sample Description (area / deg^2)	Number in sample	$z_{\rm min}$	$z_{\rm max}$
DR5Q ($\approx$ 5740)	77 429	0.078	5.414
" $z \leq 2.9$	71 375	0.078	2.900
" $0.3 \leq z \leq 2.9$	69 692	0.300	2.900
" $z \leq 2.2$	65 660	0.078	2.200
" $0.3 \leq z \leq 2.2$	63 977	0.300	2.200
PRIMARY (5713)	55 577	0.080	5.414
" $z \leq 2.9$	50 062	0.080	2.900
" $0.3 \leq z \leq 2.9$	48 526	0.300	2.900
" $z \leq 2.2$	46 272	0.080	2.200
" $0.3 \leq z \leq 2.2$	**44 736**	**0.300**	**2.200**
UNIFORM (4013)	38 208	0.084	5.338
" $z \leq 2.9$	33 699	0.084	2.900
" $0.3 \leq z \leq 2.9$	32 648	0.300	2.900
" $z \leq 2.2$	31 290	0.084	2.200
" $0.3 \leq z \leq 2.2$	**30 239**	**0.300**	**2.200**

Table 1. The SDSS Spectroscopic Quasar Samples. We include upper $z \leq 2.9$ limits simply as a guide, working from the low-z cut of Shen et al. (2007).

is flux-limited to $i = 19.1$ but since high-redshift quasars are rare, objects lying in regions of colour-space corresponding to quasars at $z > 3$ are targetted to $i = 20.2$. Furthermore, if the radio identification of an SDSS object is matched to within 2" of a source in the FIRST catalogue, (Faint Images of the Radio Sky at Twenty centimetres; Becker et al. 1995), it is included in the quasar selection.

A quasar is selected as a primary target if its 'PrimTarget' flag (Table 6, Stoughton et al. 2002) gets set to one, or more, of the following; `TARGET_QSO_CAP`; `TARGET_QSO_SKIRT`; `TARGET_HIZ`; `TARGET_FIRST_CAP`or `TARGET_FIRST_SKIRT`. Note that the selections for `TARGET_QSO_CAP` and `TARGET_QSO_SKIRT` are identical, as are the selections for `TARGET_FIRST_CAP` and `TARGET_FIRST_SKIRT`, and will be referred to as just `TARGET_QSO` and `TARGET_FIRST` respectively.

In total there are 55 577 quasars in the DR5Q that had their target flags set to one (or more) of these primary flags. The difference between this number and the 74 429 total is due in a large part to the "serendipitous" objects that make it into the DR5Q. Serendipitous targets were assigned to unused fibres once the primary quasar, LRG and galaxy targets were allocated fibres. Serendipitous targets are those that lie away from the stellar locus but fall below the quasar survey magnitude limits or have matches to *ROSAT* sources. The use of the primary sample is motivated by the fact that the SDSS quasar was designed to be complete in this selection, but no attempt was made at completeness for the other categories.

2.3 Quasar Samples

We use quasars selected from the DR5 Quasar catalogue of Schneider et al. (2007). This catalog consists of spectroscopically identified quasars that have luminosities larger than $M_i = -22.0$ (measured in the rest frame) and at least one emission line with FWHM larger than 1000 km s^{-1}. Every object in the DR5Q had its spectrum manually inspected and the catalogue does not contain Type 2 QSOs, Seyferts or BL Lac objects.

We propose to use two subsamples from DR5Q. The first we shall call the "PRIMARY" Sample, which will include those objects in the DR5Q which were targetted as primary quasar candidates, having satisfied one, or more, of the `TARGET_QSO` , `TARGET_HIZ` or `TARGET_FIRST` selections (see Schneider et al. (2007), Section 4, for more details on these flags). There are a total of 55 577 objects that are in this sub-sample, which is reduced to 46 272 with the $z \leq 2.2$ redshift cut applied and further reduced to 44 736 with the redshift limits, $0.3 \leq z \leq 2.2$ (Table 1). The PRIMARY sample covers a total of 5 713 deg^2, with a total of 48 927 objects across 4 977 deg^2 for the NGC, and 6 650 objects over 736 deg^2 for the SGC, giving sky densities of 9.73, 9.83 and 9.03 deg^{-2}, respectively.

Our second sample will be a "UNIFORM" sample. This is the same data that is used in the clustering analysis by Shen et al. (2007) for the high redshift, $z \geq 2.9$ SDSS Quasars, and full details of the selection is given in that paper. As a general pointer, Shen et al. (2007) follow the method of Richards et al. (2006) and use a uniform algorithm (hence the name) to select their quasars. In practice, this uniform algorithm selects against data obtained from the earlier part of the SDSS, particularly in the EDR (Stoughton et al. 2002) and DR1 (Abazajian et al. 2003). This was because the final quasar target selection algorithm was not in place when the data were taken for the EDR or DR1. There are 38 208 objects in total that are in this UNIFORM sub-sample, dropping to 31 290 when a redshift cut of $z \leq 2.2$ is applied and dropping again to 30 239 when the redshift cuts of $0.3 \leq z \leq 2.2$ are imposed. The UNIFORM sample covers 4013deg^2.

[YS: Can you fill in the split in area for the UNIFORMS for the NGC and SGP?]

We motivate the use of both the PRIMARY and UNIFORM sub-samples in our analysis for several reasons. Since PRIMARY is generally a more 'liberally' selected quasar sample, it will be interesting to note, what affect, if any,

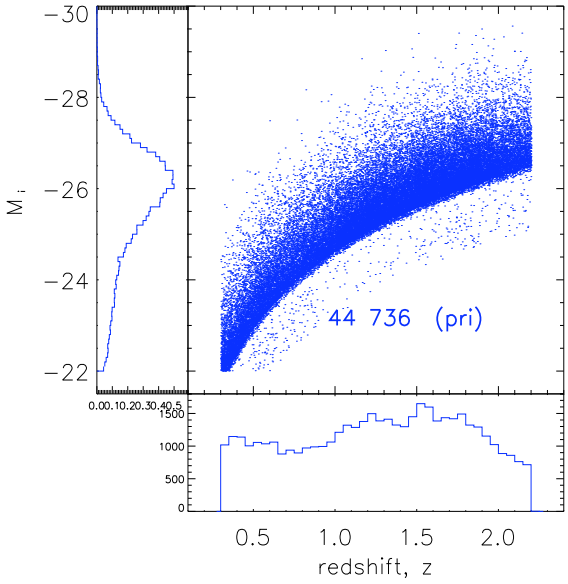

Figure 1. The SDSS DR5 Quasar $L-z$ plane for the PRIMARY sample. Absolute Magnitudes, M_i, are taken to be at $z = 0$. The affect of the apparent $i = 19.1$ magnitude limit can clearly be seen, and the objects that are fainter than this were those selected to be high, $z > 3$ quasars, but turned out not to be. Please see `http://www.astro.psu.edu/users/npr/DR5/` for further $L-z$ plots at full resolution.

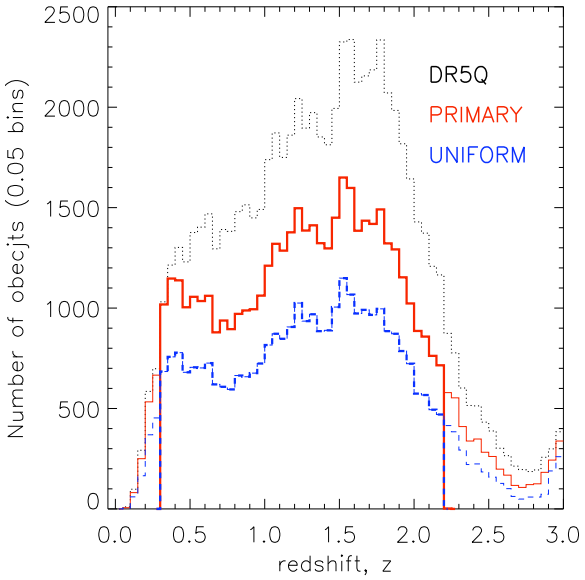

Figure 2. The SDSS DR5 Quasar $N(z)$. The solid (red) histogram show the quasar redshift distribution for the PRIMARY sample, while the dashed (blue) histogram show the redshift distribution for the UNIFORM sample. The thin lines for both PRIMARY and UNIFORM, do not include the $0.3 \leq z \leq 2.2$ cuts. As a comparison, the full DR5Q is given by the dotted (black) histogram.

this selection has when compared to results from the UNIFORM sample, the latter being specifically designed to be a statistical sample for correlation and luminosity function investigations. Further, if we find the difference in results generated from the PRIMARY and UNIFORM samples to be negligible (which should be the case for redshifts $z \leq 2.2$), then the $\sim 50\%$ increase in quasar numbers that comes from the PRIMARY data will boost the Signal-to-Noise when we come to divide the sample for our 2PCF redshift-evolution studies

We also aim to split the "$L - z$ degeneracy" that is a common issue to quasar studies (e.g. da Ângela et al. 2008). Due to the evolution of the quasar luminosity function, and the flux-limited nature of the SDSS Quasar sample, a luminosity-redshift degeneracy is introduced to our sample. This is simply due to the fact that one is biased towards more luminous quasars at high redshift and fainter objects at low-redshifts. Thus, in order to separate the affects of any potential luminosity-dependent clustering and redshift-dependent clustering, one would like to have as large a sample of objects, *at a fixed luminosity*, over the redshift-range of interest. The additional PRIMARY quasars will thus help break the $L - z$ degeneracy, while keeping the S/N per redshift bin suitably high. We plot our redshift-luminosity range for our PRIMARY sample in Figure 1.

2.4 Reddening signatures present within the clustering

While all selection for the quasar sample is undertaken using dereddened colors (Richards et al 2001), if there remain systematic errors in the reddening model they can induce excess power into the clustering in a number of different ways. We present results in Appendix B to demonstrate how the effects described above do/do not seriously affect our $\xi(s)$ measurements and the interpretations based thereon.
TO BE COMPLETED...

3 TECHNIQUES

In this section we shall describe the techniques we shall use to calculate the Quasar $z \leq 2.2$ 2PCF.

3.1 Estimating the 2-Point Quasar Correlation Function

The 2PCF measures the excess (or decrement) in probability for finding a *pair* of given objects at a given spatial separation, with respect to a random, Poisson distribution. The quasar 2PCF, $\xi(x)$, is defined by the joint probability that a pair of quasars are found in the two volume elements dV_1 and dV_2 placed at separation x,

$$dP_{1,2} = n_1 n_2 [1 + \xi(x)] dV_1 dV_2 \quad (3)$$

where x has dimensions of length, and n is the number density. In practice, ξ is measured by comparing the actual quasar distribution to a catalogue of "random" quasars, which have the same selection function, angular mask and radial distribution as the data, but are spatially distributed in a "random" manner - i.e. have no LSS. The construction of this random sample shall be described in Section 3.2.

We note at this point the difference between the *redshift*-space correlation function, denoted $\xi(s)$, and the *real*-space

correlation function, $\xi(r)$. The measurement of a quasar redshift, will not only have a (large) component due to the Hubble expansion, but also a component due to the intrinsic peculiar velocities associated with the individual quasar. As we shall see, the peculiar velocities can been seen in the redshift-space correlation function, both at small- and large-scales. The real-space correlation function is what would be measured in the absence of any redshift-space distortions. However, as noted in Shen et al. (2007, (Appendix A)), and Schneider et al. (2007), subtle effects creep into quasar redshift determination and redshift-errors of order $\Delta z = 0.001$, will potentially dominate our peculiar velocity signal.

MAS: Is further discussion needed here re: redshift-space distortions?

We use the estimator of Landy & Szalay (1993) as this has been found to be the most reliable estimator for 2PCF studies (Kerscher et al. 2000). We do, however, check our calculations using the 'standard' estimator of Davis & Peebles (1983) and the Hamilton estimator from Hamilton (1992). The three estimators are given by,

$$\xi_{Std}(s) = \left(\frac{N_{rd}}{N}\frac{DD(s)}{DR(s)}\right) - 1, \tag{4}$$

$$\xi_{\rm Ham}(s) = \frac{DD(s)\cdot RR(s)}{DR(s)^2} - 1 \tag{5}$$

and

$$\xi_{LS}(s) = 1 + \left(\frac{N_{rd}}{N}\right)^2 \frac{DD(s)}{RR(s)} - 2\left(\frac{N_{rd}}{N}\right)\frac{DR(s)}{RR(s)} \tag{6}$$

$$\equiv \frac{\langle DD\rangle - \langle 2DR\rangle + \langle RR\rangle}{\langle RR\rangle}, \tag{7}$$

for the 'Standard', Hamilton and Landy-Szalay (LS) estimators respectively. Here N and $N_{\rm rnd}$ are the number of data and random points in your sample, DD(s) is the number of data-data pairs in the given catalogue (within the survey window, W), DR(s) are the number of data-random pairs and RR(s) the number of random-random pairs. The angled brackets denote the suitably normalised pair counts. We employ at least twenty times more random points than data in order to reduce Poisson noise and choose our bin widths to be $\Delta\log(s/\ h^{-1}\ {\rm Mpc}) = 0.1$ in size.

One can also resolve the redshift-space separation, s, into two components, σ and π, where σ is the separation between two objects *perpendicular* to the line-of-sight and π is the separation *parallel* to the line-of-sight. Thus,

$$s^2 = \sigma^2 + \pi^2, \tag{8}$$

where $r_p \equiv \sigma$ is also found in the literature. The '2-D' redshift-space correlation function, $\xi(\sigma,\pi)$, can be calculated as before,

$$\xi_{\rm LS}(\sigma,\pi) = \frac{\langle DD(\sigma,\pi)\rangle - \langle 2DR(\sigma,\pi)\rangle + \langle RR(\sigma,\pi)\rangle}{\langle RR(\sigma,\pi)\rangle} \tag{9}$$

where the bin sizes are now chosen to be $\Delta\log(\sigma\ \ h^{-1}\ {\rm Mpc}) = \Delta\log(\pi\ \ h^{-1}\ {\rm Mpc}) = 0.2$.

3.1.1 The Projected Correlation Function, $w_p(\sigma)$

For ease of comparison against e.g. theoretical models, we desire to know not only the redshift-space correlation function, but also the real-space correlation function and indeed the real-space correlation function length, r_0 and slope, γ. As such, we use the fact that redshift-space distortions affect only the radial component of $\xi(\sigma,\pi)$, and thus by integrating along the line-of-sight direction, π, we obtain the projected correlation function,

$$w_{\rm p}(\sigma) = 2\int_0^\infty \xi(\sigma,\pi)\, d\pi. \tag{10}$$

In practice we set the upper limit on the integral to be $\pi_{\rm max} = 10^{1.8} = 63.1\ \ h^{-1}$ Mpc and show that although varying this limit does cause some difference to the $w_p(\sigma)$, it does not cause significant changes to the 2PCF over the scales of interest for our studies (Appendix B9).

Due to $w_p(\sigma)$ now describing the real-space clustering, the integral in equation 10 can be rewritten in terms of $\xi(r)$ (Davis & Peebles 1983),

$$w_{\rm p}(\sigma) = 2\int_0^{\pi{\rm max}} \frac{r\,\xi(r)}{\sqrt{(r^2-\sigma^2)}} dr \tag{11}$$

If we then assume that $\xi(r)$ is a power-law of the form, $\xi(r) = (r/r_0)^{-\gamma}$ (which, as we shall find later, is a fair assumption), then equation 11 can be integrated analytically, such that with $\pi_{\rm max} = \infty$,

$$w_p(\sigma) = r_0^{\gamma_r}\sigma^{1-\gamma_r}\left[\frac{\Gamma(\frac{1}{2})\,\Gamma(\frac{\gamma-1}{2})}{\Gamma(\frac{\gamma}{2})}\right] = r_0^{\gamma_r}\sigma^{1-\gamma_r}A(\gamma_r), \tag{12}$$

where $\Gamma(x)$ is the Gamma function calculated at x. We now have a method for calculating the real-space correlation length and power-law slope, denoted r_0 and γ_r respectively (Magliocchetti et al. 1999; Hawkins et al. 2003).

The redshift-space and real-space correlation function can be related via

$$\xi(s) = \xi(r)\left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right), \tag{13}$$

where the β parametrizes the 'flattening' at large-scales of the correlation function due to the infall of matter from underdense to overdense regions. We again note that equation 13 is only in the limit of linear peculiar velocities and the value of β has traditionally been measured via fits to observed data (e.g. Kaiser 1987; Peacock et al. 2001; Hawkins et al. 2003; Ross et al. 2007; Guzzo et al. 2008).

3.1.2 2PCF Evolution

The 2PCF can also evolve with redshift,

$$\xi(r,z) = (r/r_0)^{-\gamma}(1+z)^{\gamma-(3+\epsilon)} \tag{14}$$

$$= (r/r_0)^{\gamma-(3+\epsilon)} \tag{15}$$

where r again is the comoving distance, r_0 is the comoving correlation length at $z = 0$ and where the choice of ϵ determines the redshift evolution. Thus, the comoving correlation length at redshift z, can be expressed as

$$r_0(z) = r_0 f(z), \quad f(z) = (1+z)^{\gamma-(3+\epsilon)} \tag{16}$$

(Phillipps et al. 1978; Peacock 1999; Overzier et al. 2003; Farrah et al. 2006).

Following these authors, we postulate that there are essentially three important model cases, that use this ϵ-parametrization. The first is a "Comoving clustering" model where dark matter haloes expand with the Universe and

$\epsilon = \gamma - 3$ implies that clustering remains constant with redshift, $(1+z)^0$. The second model is the "Stable" clustering mode, where $\epsilon = 0$. In this case, the size of the dark matter haloes is frozen in *proper* coordinates, and hence, the comoving expansion of the background mass distribution makes the clustering of the haloes grow *stronger*. Put another way, if (quasar) clustering is gravitationally bound at small scales, then e.g. clusters have fixed physical sizes. This would then imply a redshift dependence of $(1+z)^{-1.2}$, if one assumes a canonical $\gamma = 1.8$ power-law. The third model is "linear growth" clustering model, where clustering grows as expected under linear perturbation theory and $\epsilon = \gamma - 1$ (and not $\epsilon = 1$ as in Farrah et al. 2006). This now implies a redshift dependence of $(1+z)^{-2}$. For those models which have $\epsilon \geq 0$, clustering increases with time or, as redshift drops).

We note however, that these are relatively simplistic and mearly rough guides to the underlying physics (e.g. Moscardini et al. 1998). As such, to fully exploit the observational results and constrain theoretical models, we note only compare the ϵ-models to our data, but also make use of the results from recent N-body and Semi-Analytic Models. We shall return to this in Section 6.

3.2 Construction of the Random Catalogue

. As mentioned above, in order to work out the ξ in practice, one needs a random catalogue of points, that mimics the data in every way, bar its clustering signal. Here we describe the construction of such a "random" catalogoue.

3.2.1 Angular Mask

The angular mask of the SDSS has a describable, but non-trivial geometry. As such, one of the main challenges involved in producing a catalogue of "random" points, is to mimic the angular geometry completeness of the observational survey. Here again we use our two samples, PRIMARY and UNIFORM, as our data, and thus aim to build to angular completeness masks for both. Full technical details of the construction of our angular mask are given in Appendix A.

3.2.2 Radial Distribution

Figure 2 shows the N(z) distribution of the DR5Q Quasars from our samples. We fit a high-order polynomial to both the PRIMARY and UNIFORM samples which we use to generate the random sample redshift distribution. This method has proved reliable in previous quasar clustering studies (Croom et al. 2005; da Ângela et al. 2008).

3.2.3 Fibre Collisions and the Impact on Small-Scale Structure

Due to the design of the SDSS fibres and plates, no two spectroscopic fibres can be positions more closely than a separation on the sky of 55". Thus, if two objects are closer than this separation, they will not be fibred in one plate observation. We use the PhotoObjAll catalogue (Appendix A) to investigate which potential quasars would not have been observed due to fibre collisions, and find that on average ~ 1 object deg^{-2} is affect by this limitation. Thus, although we acknowledge that we are potentially missing quasar pairs (or triples, or quads etc.) on small, $\leq 1\ h^{-1}$ Mpc scales, we do not currently correct for this.
TBC...

3.3 Errors and Covariances

Recent correlation studies (Scranton et al. 2002; Zehavi et al. 2002; Myers et al. 2006; Ross et al. 2007, e.g.) have employed three main methods, *Poisson, Field-to-Field* and *Jackknife* to estimate errors associated with correlation function measurements. The 'simplest' of these is a Poisson error, defined as

$$\sigma_{\rm Poi} = \frac{1+\xi(s)}{\sqrt{\rm DD(s)}}, \tag{17}$$

where $\xi(s)$ is the measured 2PCF at separation s, and DD(s) is the number of data-data pairs also at separation s. However, as reported in Myers et al. (2005); Ross et al. (2007, e.g.), the Poisson error can under-estimate a measurements error when compared to e.g. the Field-to-Field or Jackknife error. For this work, we will not report any Field-to-field errors, but instead concentrate on a jackknife resampling procedure in order to calculate the full covariance matrix.

The jackknife procedure divides the sample into 20 sub-samples, each sub-sample covering approximately 200 deg^2 of the sky, corresponding to a spatial extent of $\approx 300\ h^{-1}$ Mpc at the mean redshift of $\bar{z} = 1.27$. The number of subdivisions is chosen such that each represents a cosmologically significant volume, while retaining sufficient numbers of objects that shot noise will not dominate any subsequent analysis. We note that $\sim$200 deg^2 is slightly larger in area than the entire area of recently complete 2SLAQ survey (Cannon et al. 2006; Croom et al. 2008). Then the covariance matrix is estimated by removing a single sub-sample, re-calculating the correlation function, and repeating this procedure for all the sub-samples. The covariance matrix is thus,

$$\mathrm{Cov}(\xi_i, \xi_j) = \frac{N-1}{N} \sum_{k=1}^{N} (\xi_i^k - \bar{\xi_i})\,(\xi_j^k - \bar{\xi_j}) \tag{18}$$

where ξ_i^k and ξ_j^k are the correlation functions for each sub-sample and $\bar{\xi_i}$ is the mean value of ξ_i from the sub-samples (Zehavi et al. 2002). We present full details of the areas removed and the Covariance matrix recovered, in Appendix B.

4 RESULTS

4.1 SDSS Quasar Redshift-Space 2-Point Correlation Function, $\xi(s)$ $(0.30 \leq z \leq 2.2)$

The 2-Point redshift-space correlation function for the UNIFORM sample over is given in Figure 3. We follow da Ângela et al. (2008) and fit a model which has

$$\xi(r) = \begin{cases} (r/6.00)^{-1.45}, & r < 10\ \ h^{-1}\ \mathrm{Mpc} \\ (r/7.25)^{-2.30}, & r > 10\ \ h^{-1}\ \mathrm{Mpc} \end{cases} \tag{19}$$

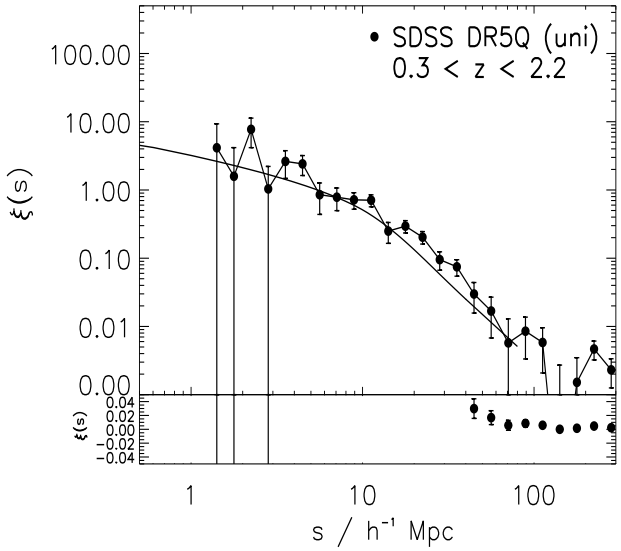

Figure 3. The SDSS Quasar redshift-space 2PCF, $\xi(s)$, with our double power-law model. The lower panel shows the $\xi(s)$ behaviour near zero on a linear scale.

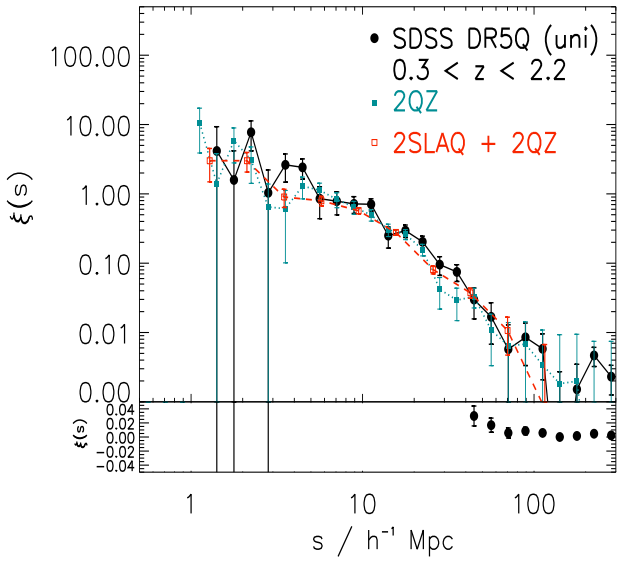

Figure 4. The Quasar redshift-space 2PCF, $\xi(s)$, for the SDSS, 2QZ and 2SLAQ QSO surveys.

and a *rms* velocity dispersion, $\langle w_z^2 \rangle^{1/2} = 800$ km s^{-1} and the dynamical infall parameter $\beta(z = 1.4) = 0.32$. Although we acknowledge this model fit could be improved, e.g. higher correlation length at large-scales, we do strongly suggest that a double power-law fit is a much more acceptable description of the quasar redshift-space correlation function that the traditional single power-law model.

[NOTE TO CO-AUTHORS: IMPROVED FIT FORTHCOMING...]

In Figure 4, we show $\xi(s)$ from two other recent surveys, the 2QZ and 2SLAQ QSO survey, covering very similar redshift ranges. Here we see remarkably similar clustering strengths in the SDSS, 2QZ and 2SLAQ samples over 1 h^{-1} Mpc $\leq s \leq$ 100 h^{-1} Mpc scales.

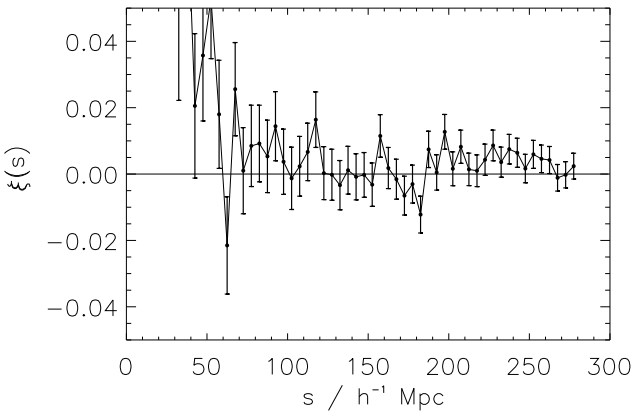

Figure 5. The SDSS Quasar redshift-space 2PCF, $\xi(s)$, large scales.

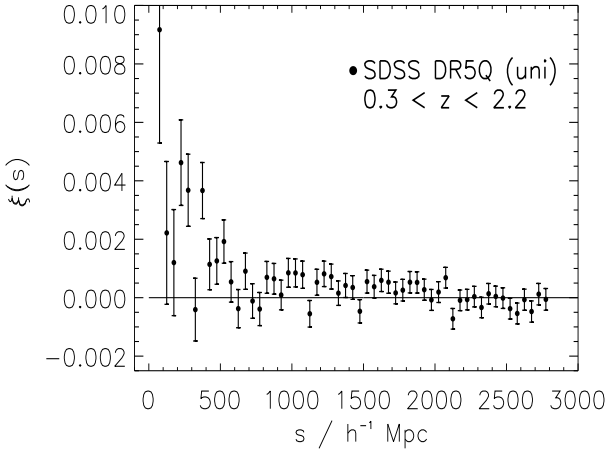

Figure 6. The SDSS Quasar redshift-space 2PCF, $\xi(s)$, at extra-large scales.

4.1.1 Small-Scale, < 1 h^{-1} Mpc

What do we see?
cf. Hennawi et al. (2006) and Myers et al. (2007)

4.1.2 Large-Scale, > 100 h^{-1} Mpc

In Figure 5 we show the large, ≤ 300 h^{-1} Mpc, behaviour of the redshift-space 2PCF.

4.2 SDSS Quasar 2-D 2-Point Correlation Function, $\xi(\sigma, \pi)$ $(0.30 \leq z \leq 2.2)$

Figure 7 shows the SDSS DR5 Quasar 2-D redshift-space correlation function $\xi(\sigma, \pi)$.

4.3 SDSS Quasar Projected 2-Point Correlation Function, $w_p(\sigma)$ $(0.30 \leq z \leq 2.2)$

In Figure 8, we show the projected 2-point correlation function, $w_p(\sigma)$ (note again we use σ for the across the line-of-sight separation, other authors use $w_p(r_p)$)

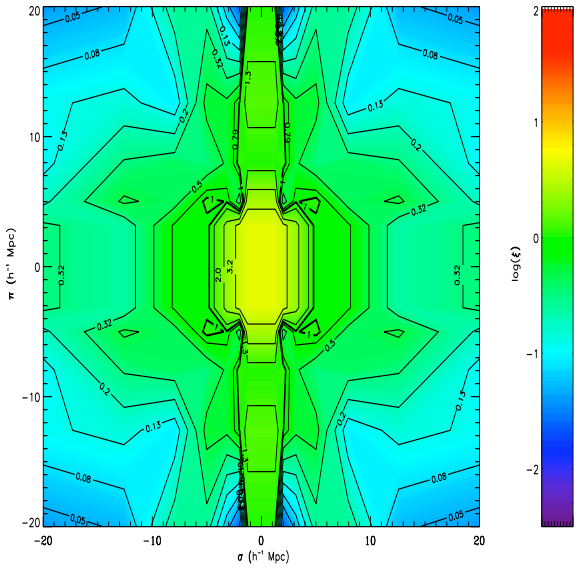

Figure 7. The SDSS DR5 Quasar $\xi(\sigma, \pi)$.

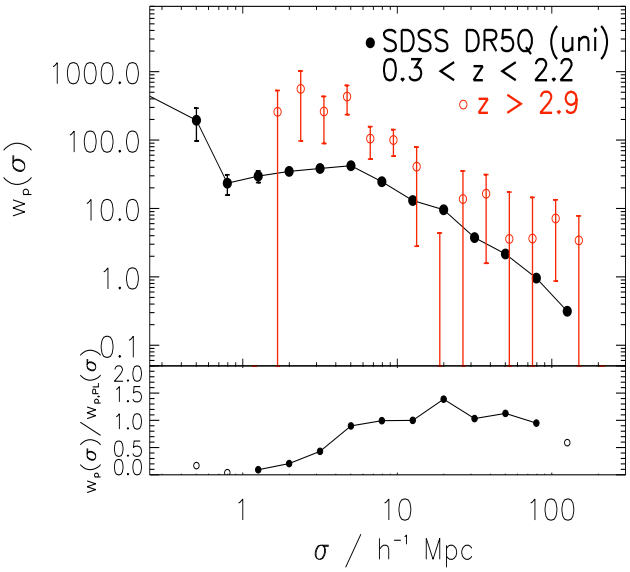

Figure 8. The SDSS Quasar redshift-space 2PCF, $w_p(\sigma)$.

4.4 Evolution of the SDSS Quasar Correlation Function

Figures 9 and 10 show the evolution of the redshift-space, $\xi(s)$, and the projected, $w_p(\sigma)$ 2PCF, using the SDSS DR5 'Uniform' Quasar sample.

Note in both plots the trend for the amplitude of the correlation function to increase, with the thin solid line in both plots is for the full DR5Q 'Uniform' sample, over $0.30 < z < 2.2$. Poisson error bars are currently quoted in these figures but should only be taken as indicative at this stage (as they are probably vastly underestimating the error in especially the $w_p(\sigma)$ results).

Figure X will show the evolution of the 2PCF, across our redshift range, but over a narrow, ~ 1 magnitude, range in luminosity, thus breaking the $L - z$ degeneracy.

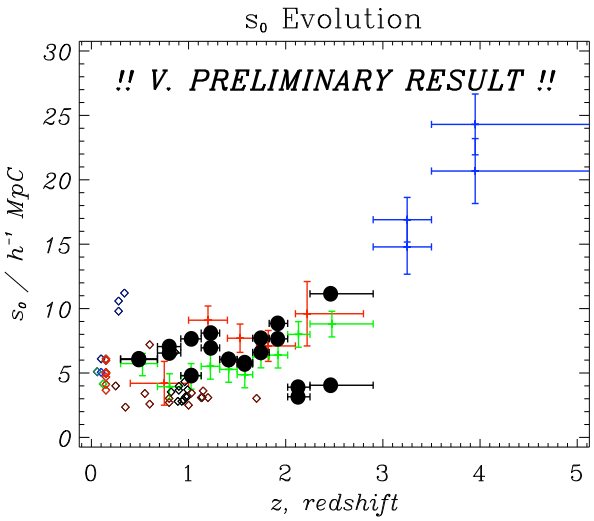

Figure 11. Evolution of the redshift-space Correlation Length, s_0. Black Points, this work; Green Points, Croom et al. (2005); Red points, Myers et al. (2006); Blue points, Shen et al. (2007). With diamonds from various other surveys incl. 2dFGRS, SDSS MAIN, SDSS LRG, 2SLAQ LRG, VVDS, DEEP2.

5 REDSHIFT-SPACE DISTORTIONS

5.1 What value of β, $\langle w_z^2 \rangle^{1/2}$...

Can we tell anything from the z-space distortions?
What value of β, $\langle w_z^2 \rangle^{1/2}$ and maybe most of all b_Q best-fit our data?
Re-visit Hoyle et al. (2002); da Ângela et al. (2005).
$\langle w_z^2 \rangle^{1/2} \equiv\equiv a$ used by Ross et al. (2007); Hawkins et al. (2003) respectively.

6 EVOLUTION OF BIAS

6.1 Comparison to Previous Results

Compare to Croom et al. (2004), using $\bar{\xi}$, ??? and Porciani & Norberg (2006) (?)
Compare to Myers et al. (2006)
Compare to Shen et al. (2007)
and various galaxy surveys, e.g. 2dFGRS, SDSS MAIN, SDSS LRGs, 2SLAQ LRGs, VVDS, DEEP2... **NB. We shall compare s_0 results to those of Croom05 (who give everything exclusively in redshift-space), but will use r_0 values (from our $w_p(\sigma)$ fits) as often as possible.**

6.2 Different Models of bias

Very simple correlation length, $r_0(z)$, linear bias, $b(z)$ and ϵ models.
$\xi(r, z) = (r/r_0)(1+z)^{\gamma-(3+\epsilon)}$

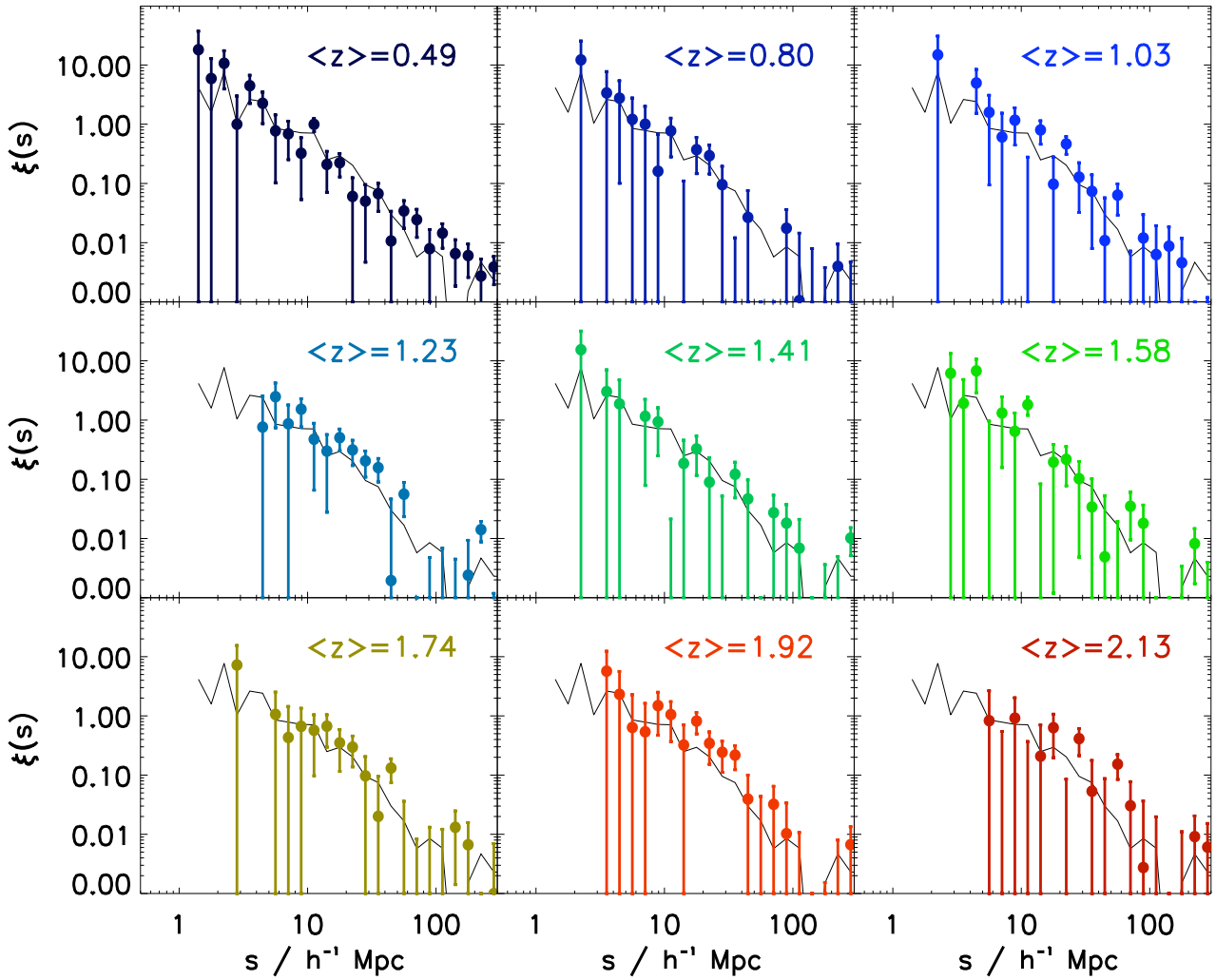

Figure 9. The SDSS DR5 Quasar redshift-space 2PCF, $\xi(s)$, and its evolution with redshift. The thin, solid line in each panel is $\xi(s)$ for the full DR5Q 'Uniform' sample, over $0.30 < z < 2.2$.

z-interval	$\bar{z}$	$N_{\rm gals}$	Range/ h^{-1} Mpc	r_0
0.30,2.20	1.269	30 239	1.0,100.0	5.28±0.52
0.30,0.68	0.488	5 404	1.0,100.0	5.80±0.54
0.68,0.92	0.801	3 001	1.0,100.0	5.81±0.73
0.92,1.13	1.029	3 365	1.0,100.0	8.20±0.96
1.13,1.32	1.228	3 623	1.0,100.0	7.31±0.87
1.32,1.50	1.412	3 333	1.0,100.0	6.78±0.77
1.50,1.66	1.577	3 405	1.0,100.0	7.71±1.06
1.66,1.83	1.744	3 240	1.0,100.0	7.34±0.79
1.83,2.02	1.917	2 970	1.0,100.0	9.72±0.91
2.02,2.20	2.104	1 899	1.0,100.0	8.83±1.21

Table 2. Evolution of the real-space correlation length, holding the and power-law slope fixed at $\gamma = 1.8$ from the UNIFORM sample, using fits to $w_p(\sigma)$.

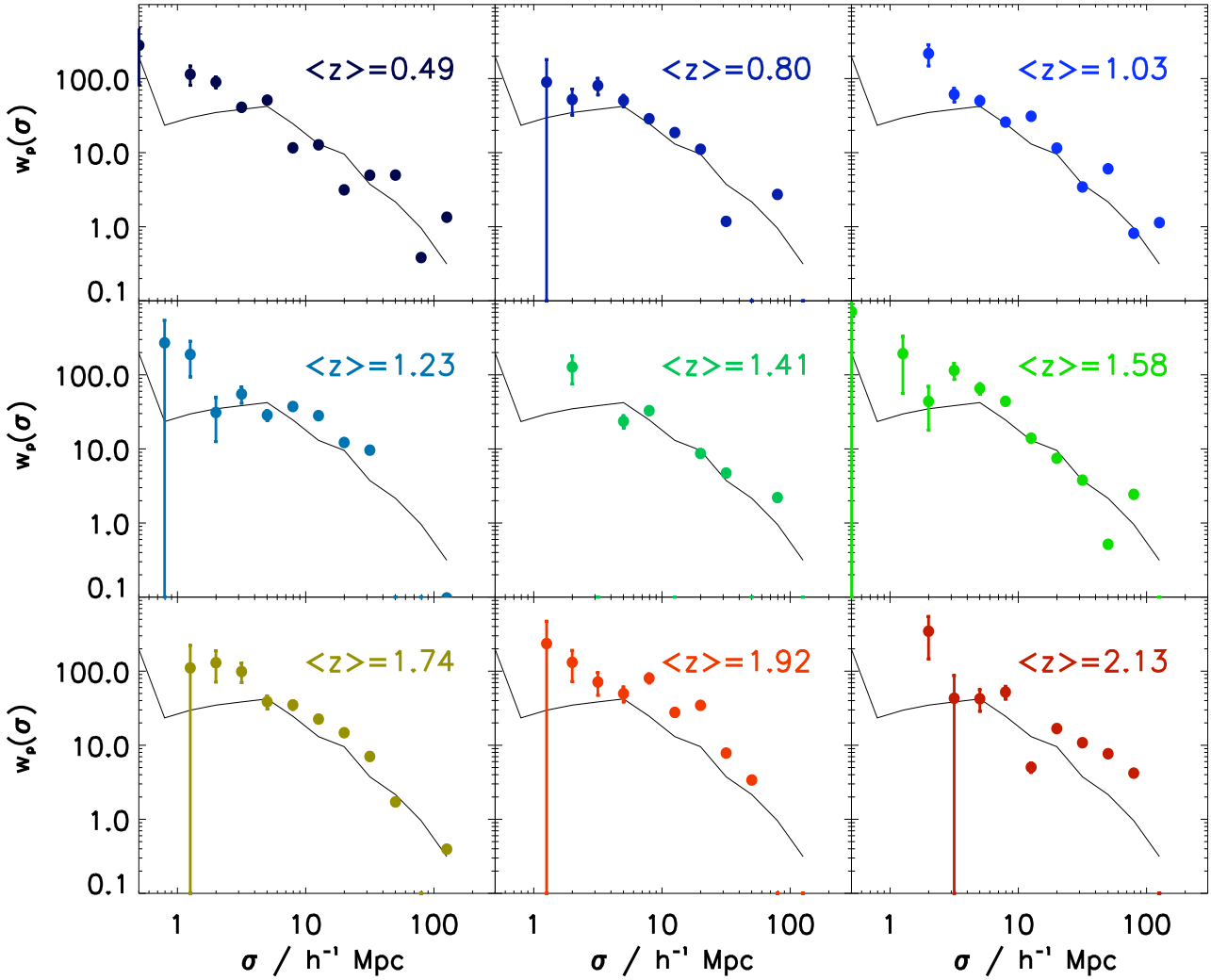

Figure 10. The SDSS DR5 Quasar projected 2PCF, $w_p(\sigma)$, and its evolution with redshift. The thin, solid line in each panel is $w_p(\sigma)$ for the full DR5Q 'Uniform' sample, over $0.30 < z < 2.2$.

"high-peaks" models etc.
cf. White, Martini, Cohn models (0711.4109)!!!!

6.3 Link to other $z \sim 2$ Galaxies...

EROs, SMG, DRGs, BzKs, ULIRGS, LBGs, LRGs etc.
Halo masses????
ULIRGS as progenitors of quasars, with the ULIRG phase being a (merger) starburst, which then triggers quasar activity, Hopkins et al. (2007); Lidz et al. (2006, e.g.).

Comparison to obscured AGN clustering measurements....

7 CONCLUSIONS

We have used the SDSS Quasar Survey to calculate the 2-point correlation function over $\sim 5000\mathrm{deg}^2$ of the sky, covering a redshift range of $0 < z \leq 2.9$, thus representing a measurement over the largest volume of the Universe ever sampled (if using our 'PRIMARY' catalogue data).

We find that

-
-
-

In Shen et al. (2008) we shall continue our investigations into the SDSS Quasar population and study the clustering properties of DR5 quasars as a function of luminosity, virial mass, colour and radio-loudness.

Looking further afield, even with the dramatic increase in data that surveys such as the 2QZ and SDSS have provided, the desire to increase dynamic range continues. For instance, due to the steepness of the faint-end of the quasar luminosity-function, low-luminosity quasars should be relatively plentiful, as long as one can identify these objects. This will be a strong challenge for the next-generation of quasar redshift surveys but one that will lead again to another significant increase in our understanding of quasars,

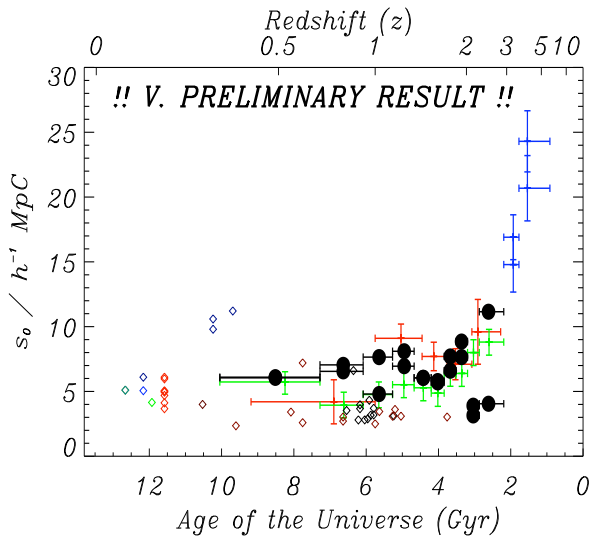

Figure 12. Evolution of the redshift-space Correlation Length, s_0. Black Points, this work; Green Points, Croom et al. (2005); Red points, Myers et al. (2006); Blue points, Shen et al. (2007). With diamonds from various other surveys incl. 2dfGRS, SDSS MAIN, SDSS LRG, 2SLAQ LRG, VVDS, DEEP2.

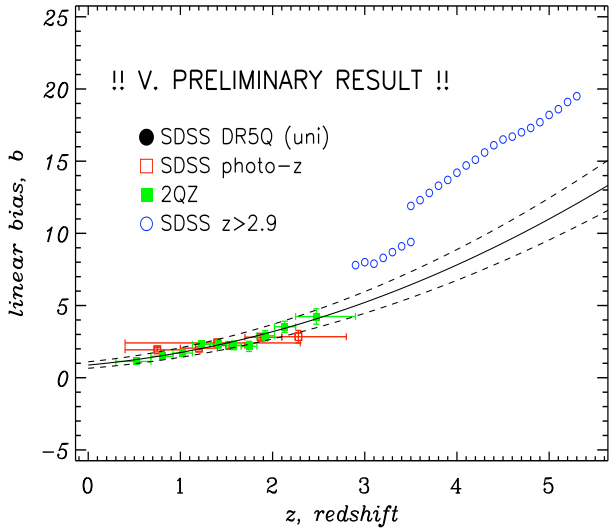

Figure 13. Evolution of the linear bias of quasars, b_Q. with redshift. Black Points, this work (to be plotted!!); Green Points, Croom et al. (2005); Red points, Myers et al. (2006); Blue points, Shen et al. (2007). Thin solid line from empirical model, $b_{Q(z)} = (0.53 \pm 0.19) + (0.289 \pm 0.035)(1+z)^2$ from Croom et al. (2005).

supermassive black holes, galaxy formation and evolution and the properties of the Universe.

ACKNOWLEDGMENTS

This work was partially supported by National Science Foundation grant AST-0607634 (N.P.R. and D.P.S.). Funding for the creation and distribution of the SDSS Archive has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the U.S. Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS Web site is http://www.sdss.org/. The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The Participating Institutions are The University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, the Korean Scientist Group, Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

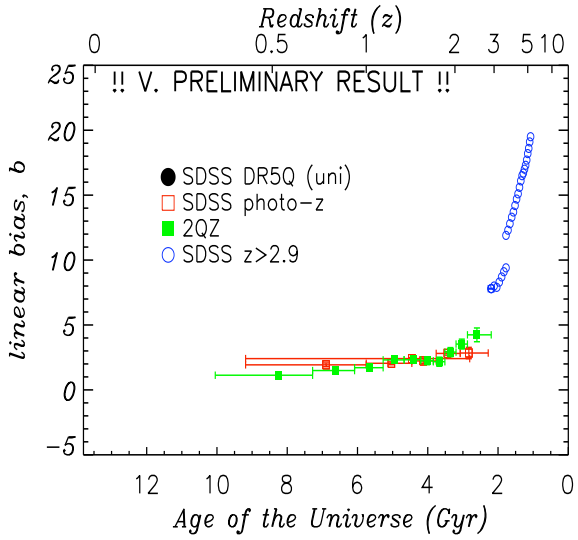

Figure 14. Evolution of the linear bias of quasars, b_Q. with look-back time. Black Points, this work (to be plotted!!); Green Points, Croom et al. (2005); Red points, Myers et al. (2006); Blue points, Shen et al. (2007).

APPENDIX A: SDSS TECHNICAL DETAILS

A1 The Catalogue Archive Server

The SDSS database can be interrogated through the Catalog Archive Server[1] (CAS) using standard Structured Query Language (SQL) queries. As such, we present all directly relevant SQL code in Appendix A3, so that an interested reader can re-create our samples in order to perform independent analysis.

When querying the CAS, you have a choice to query either the `best` database or the `target` database for a given Data Release (in our case, DR5). The difference between the `best` and `target` is that the former database contains information on all the photometric and spectroscopic objects obtained using the latest versions (and i.e. the "best") of the

[1] http://cas.sdss.org

data reduction and analysis pipelines, (Section 3, Abazajian et al. 2004). The `target` database however, contains the information on objects at the time when the targeting algorithm pipelines were run, which, in general, pre-dates the current `best` data by a significant amount. Thus, an objects properties, such as magnitude or colour, can be different between target allocation and the most recent data processing. This of course can lead to objects changing their status to either be included or excluded from a given selection. More details regarding the CAS, `best` and `target` are given in the relevant SDSS Data Release papers (Stoughton et al. 2002; Abazajian et al. 2004; Adelman-McCarthy et al. 2007).

Thus, in order to create a statistical data sample, or indeed to mimic it for a comparative 'random' sample, we need to know the properties of our chosen objects *at the time of targetting*, i.e. which objects were selected as quasar candidates. The DR5Q catalog gives information from both the `best` and `target` databases, however, we only use information from `target` unless explicitly stated otherwise.

A2 SDSS Survey Geometry

For our purposes, we define three functions, which have dependence on angular position in the sky only.

- Coverage Completeness, $f_c(\theta)$. The coverage completeness is simply the ratio of the number of quasar targets that are assigned a fibre to the overall number of quasar candidates in a given area, expressed as a percentage. The natural area for our purposes, shall be a "sector" (as described below).
- Spectroscopic Completeness, $f_s(\theta)$. This is the ratio of the number of high-quality spectra obtained, where there is sufficient confidence that the objects true redshift has been measured.
- Quasar Fraction, $f_q(\theta)$. This is simply the fraction of the number of high-quality spectra that do indeed turn out to be high-redshift quasars (as opposed to e.g. stars in our Galaxy).

An 'overall completeness', $f_{\rm O}$, is then easily defined as $f_{\rm O} = f_q \times f_s \times f_c$. We justify the assumption that these functions depend only on the angular position by referring to the checks made in Vanden Berk et al. (2005), who state that, when dealing with a well-defined statistical sample, the SDSS Quasar Survey is $\sim 95\%$ complete.

As such, we run various SQL queries on the SDSS CAS. The first query (Appendix A3) simply asks the CAS to return all the objects in the Photometric database that were targetted as being "primary" candidate quasars. When run on DR5, this returns 203 185 unique objects, from the PhotoObjAll table.

We next calculate which 'Primary PhotoObjAll' objects (POAs) fall within the spectroscopic survey plate boundaries. We use the parameter file, `maindr5spectro.par` found on the SDSS website[2], which contains the plate number, Modified Julian Date (MJD), and plate centre (in J2000 Right Ascension and Declination). 1 278 plate details consisting of the SDSS DR5 are given in this file. Note, we do not use any of the "Extra", "Special", or "ExtraSpecial" plates for our analysis and warn that the PlateX table found in the CAS does not explicitly make these plate distinctions. We find there are 145,524 POA objects that fall within 1.49 degrees of a given DR5 plate centre, noting that since plates overlap due to the tiling scheme, an object can be in more than one plate.

Of these 145,524 objects (which recall were all photometrically labelled as 'primary quasars'), we would next like to know, how many were *(a)* designated as spectroscopic ("tilable") targets by the process of 'Tiling' and *(b)* how many of these Targets were allocated fibres. Note here that the difference in the SDSS between "tiles" and plates is that a tile is a 1.49 degree radius circle on the sky determined by tiling, and which contains the locations of up to 592 tilable targets and other science targets (the other 48 fibres are assigned to calibration targets). For each tile, one, or more, physical aluminum plates will be created. The plates will have holes drilled in them for fibres to be plugged, in order to observe the tiled targets. Thus, one tile can have several Plates, with the plates having common centres.

The design of the SDSS survey with regards to the tiling procedure is described in detail by Blanton et al. (2003)[3] and its goal is to maximise the probability of a target being fibred. Furthermore, because of the large-scale structure in the quasar/galaxy distribution that is known to exist, an optimal Tiling procedure will cause individual Tiles to overlap with each other. As described in Blanton et al. (2003); Tegmark et al. (2004); Blanton et al. (2005); Percival et al. (2007), a "sector" is defined as a set of sphercial polygons (i.e. tile overlap regions) that could have only been observed by a unique combination of tiles and survey "chunks". A 'chunk' is a unit of SDSS imaging data and is a part of an SDSS 'stripe', which is a 2.5 degree wide cylindrical segment aligned at a great circle between the survey poles. Thus a slight complication arises in that plates, sectors and plate centres deal with α and δ coordinates, whereas chunks are defined by the survey coordinates η and λ (with the coordinate transformations given in Section 3.2.2 of Stoughton et al. 2002). Again, the SDSS website, `www.sdss.org` and the relevant data release papers have complete details. These sectors are thus the appropriate regions on which to define the completeness of our sample and survey.

Using the RegionID field in the TARGET table (which gives the sector identification number if set, zero otherwise) we match the positions (R.A.'s and Decs) of objects in TARGET to those that are in PhotoObjAll and the DR5Q.

From here we figure out sector completenesses and the angular mask of the SDSS DR5 Survey. We also note that the imaging 'chunks' also play a role in defining the geometry of the SDSS survey since a chunk is 2.5° wide, whereas a tile is 3.0° wide and thus there will be regions that have been tiled but which have no imaging data and thus no spectroscopic targets to contribute to our sample. Note, we do not include the effect of that targets less than 100 arcsec from the tile center are excluded.

[2] http://www.sdss.org/dr5/coverage/index.html

[3] see also http://www.sdss.org/dr6/algorithms/tiling.html

A3 SQL Queries

All SQL queries were used on the SDSS TARGET DR5 database using the CAS Server. In all the queries, using `SELECT *` will usually return many more record columns than is necessary for a clustering analysis (especially with the `PhotoObjAll` table). "--" is the comment symbol in SQL.

A3.1 Generation of "PhotoObjAll"

```
SELECT *
FROM TARGDR5..PhotoObjAll
-- The entire photometric database
WHERE ( (PrimTarget & 0x0000001f) > 0)
-- Using the bitwise operator & to select targets
-- with QSO, HIZ or FIRST selections.
```

A3.2 "Primary"

```
SELECT *
FROM DR5QuasarCatalog
-- The DR5 Quasar catalog, a.k.a. DR5Q
WHERE ( (PrimTarget & 0x0000001f) > 0 )
```

A3.3 "Uniform" (adapted from Richards et al. (2006))

```
SELECT *
FROM
WHERE region2table r,
-- The region2table table contains...
    tilinggeometry g,
-- The tilinggeometry table contains...
    g.targetversion > 'v3_1_0'
-- As noted in Richards et al. (2006), although
-- maybe not ''proper'' SQL, this seems to work
-- well.
etc.
```

APPENDIX B: JACKKNIFE ERRORS

TO DO:
Calculate Covariance Matrix.
Leading diag. terms vs. non-leading diag. terms
Also have to think carefully about errors on the redshift evolution subsamples.
Poisson vs. Jackknife comparison
COMMENTS??

$$C_{ij} = C(\theta_i, \theta_j) = \sum_{L=1}^{N} \sqrt{\frac{RR_L(\theta_i)}{RR(\theta_i)}} \left[\omega_L(\theta_i) - \omega(\theta_i)\right] . \sqrt{\frac{RR_L(\theta_j)}{RR(\theta_j)}} \left[\omega_L(\theta_j) - \omega(\theta_j)\right] \quad \text{(B1)}$$

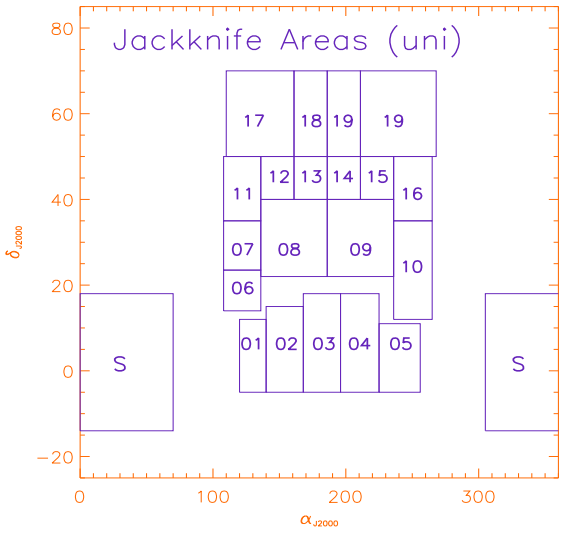

Figure B1.

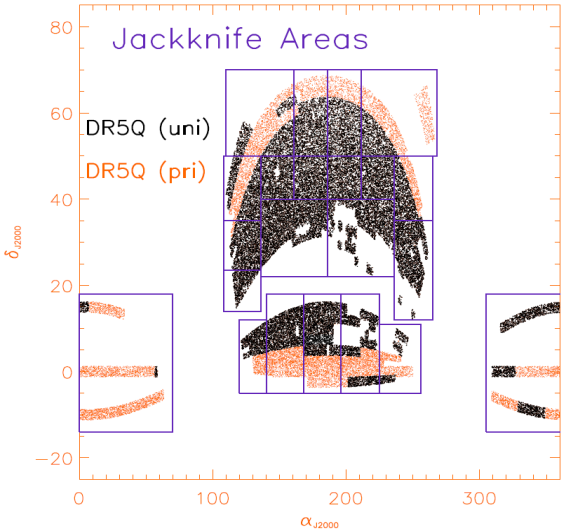

Figure B2.

APPENDIX C: SYSTEMATICS IN THE SDSS QUASAR 2PCF.

Here we produce evidence showing how various changes to our data and random methodology described in Sections 2 and 3 affect our main results. We generally report on $\xi(s)$ as this is the most robust of our 2PCF measurements.

C1 Different Estimators for UNIFORM $\xi(s)$

Figure C1 shows the redshift-space 2PCF, $\xi(s)$ for the UNIFORM sample, using the different estimators of Davis & Peebles (1983), Hamilton (1992) and Landy & Szalay (1993). While the trend for the two estimators that use random-random (RR) pair counts are in extremely good agreement, the 'Standard' estimator, seems to have too much power on large, $s \geq 40 \quad h^{-1}$ Mpc. This feature is currently unexplained.

Region Name	RA min	RA max	Dec min	Dec max	No. of Quasars
N01	120.	140.	-5.	12.	29 445
N02	140.	168.	-5.	15.	28 445
N03	168.	196.	-5.	18.	27 904
N04	196.	225.	-5.	18.	
N05	225.	256.	-5.	11.	
N06	108.	136.	14.	23.5	
N07	108.	136.	23.5	35.	
N08	136.	186.	22.	40.	
N09	186.	236.	22.	40.	
N10	236.	265.	12.	35.	
N10	108.	136.	35.	50.	
N11	136.	161.	40.	50.	
N12	161.	186.	40.	50.	
N13	186.	211.	40.	50.	
N14	211.	236.	40.	50.	
N15	236.	265.	35.	50.	
N16	110.	161.	50.	70.	
N17	161.	186.	50.	70.	
N18	186.	211.	50.	70.	
N19	211.	268.	50.	70.	
S	0.	70.	-14	18	
S	305.	360.	-14	18	

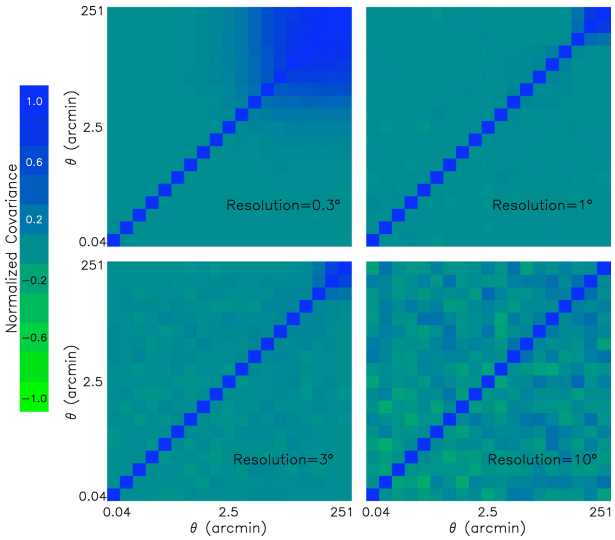

Figure B3. The normalised Covariance Matrix, (a.k.a. the Regression Matrix) for $\xi(s)$ from jackknife error anaylsis on 20 subsamples of the UNIFORM DR5Q. **[PLACE HOLDER FROM MYERS et al. (2007, ApJ, 658, 85)]**

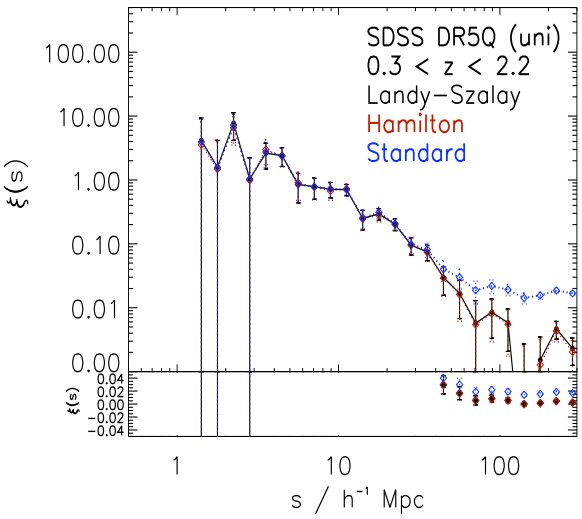

Figure C1. The SDSS DR5 Quasar $\xi(s)$ for the UNIFORM sample using the 'Standard', Hamilton and Landy-Szalay estimators.

[NOTE TO CO-AUTHORS: ANY SUGGESTIONS HERE AT ALL??!]

C2 PRIMARY vs. UNIFORM $\xi(s)$

Figure C2 shows the difference in the redshift-space correlation function, $\xi(s)$, for the PRIMARY sample, versus that of the UNIFORM sample. Again we see excellent agreement of the two samples at small $s \leq 20 \;\; h^{-1}$ Mpc scales, but the PRIMARY sample exhibits a higher clustering strength at large-scales, $s \geq 40 \;\; h^{-1}$ Mpc. Althought this trend was seen in an unpublished calculation on the SDSS DR3 Quasar sample, we currently do not believe it to be a real affect but it remains unexplained at this point.

C3 UNIFORM $z < 2.2$ vs. $z < 2.9$ $\xi(s)$

Figure C3 shows the redshift-space 2-point correlation function $\xi(s)$ for the UNIFORM sample with the high-redshift cut-off being changed from $z \leq 2.2$ to $z \leq 2.9$.

C4 UNIFORM NGC vs. SGC $\xi(s)$

Figure C4 shows the redshift-space 2-point correlation function $\xi(s)$ for the UNIFORM sample, split into quasars from

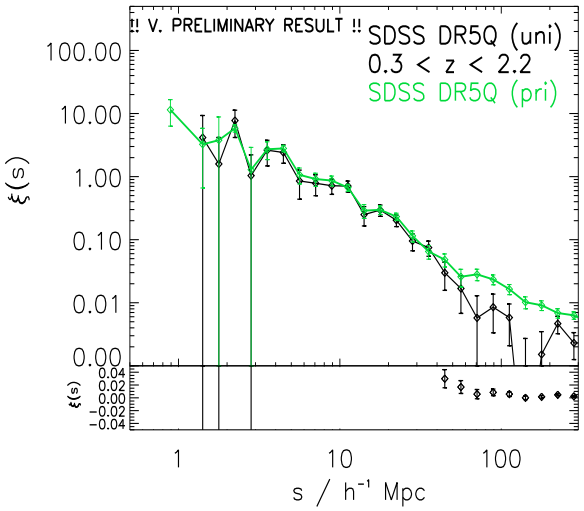

Figure C2. The SDSS DR5 Quasar $\xi(s)$ the PRIMARY and UNIFORM samples. The lower panel shows the behaviour of the Uniform $\xi(s)$ near zero on a linear scale

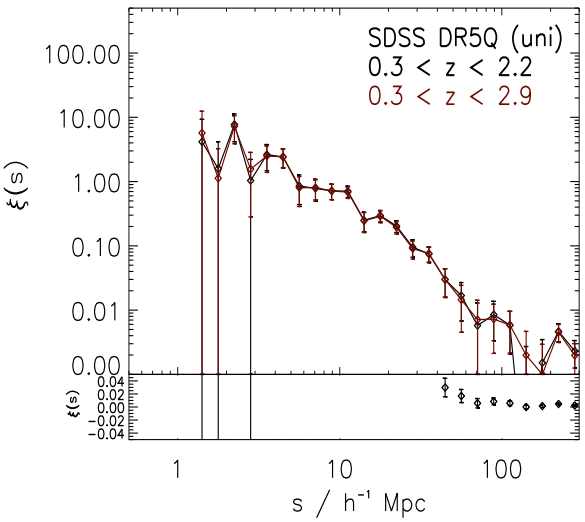

Figure C3. The SDSS DR5 Quasar $\xi(s)$ the PRIMARY and UNIFORM samples. The lower panel shows the behaviour of the Uniform $0.3 < z < \xi(s)$ near zero on a linear scale

the North Galactic Cap (NGC) and the South Galactic Cap (SGC). Note the data is heavily dominated by the NGC in the UNIFORM sample.

C5 PRIMARY and "Bad Fields" $\xi(s)$

Figure C5 shows the redshift-space 2-point correlation function $\xi(s)$ for the PRIMARY sample, including solid (green) and excluding dashed (red) lines, the "Bad Fields" as defined by Shen et al. (2007).

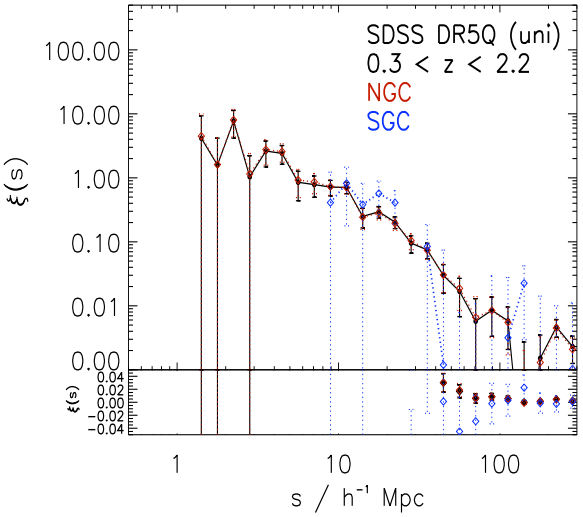

Figure C4. The SDSS DR5 Quasar $\xi(s)$ the UNIFORM sample with the sample split into the NCG and SCG.

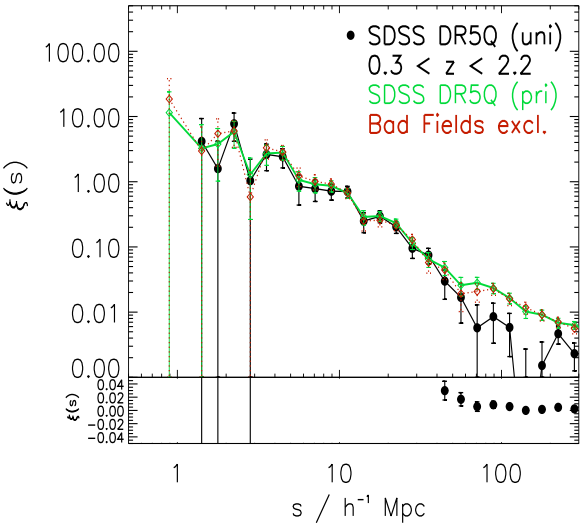

Figure C5. The SDSS DR5 Quasar $\xi(s)$ the UNIFORM and PRIMARY sample both including solid (green) line and excluding, dashed (red) line bad imaging fields

C6 PRIMARY from Single and Multiple Plates $\xi(s)$

Figure C6 shows the redshift-space 2-point correlation function $\xi(s)$ for the PRIMARY sample with data coming from Single (red) or Multiple (blue) plates.
[NOTE TO CO-AUTHORS: DOES THIS SHED ANY LIGHT ON THE ABOVE PRIMARY vs. UNIFORM ISSUES AT LARGE SCALES?]

C7 PRIMARY and "Empty" Sectors

Figure C7 shows the redshift-space 2-point correlation function $\xi(s)$ for the PRIMARY sample with "Empty Sectors" (those with no primary quasar targets at all), having 1, 0.5 or 0 completeness. There are 1983 "Empty" sectors contributing 93 deg^2 to the DR5Q.

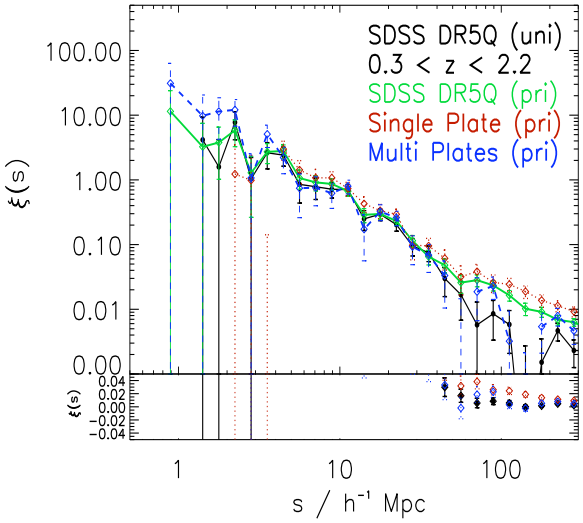

Figure C6. The PRIMARY sample $\xi(s)$ with data coming from Single (red) or Multiple (blue) plates.

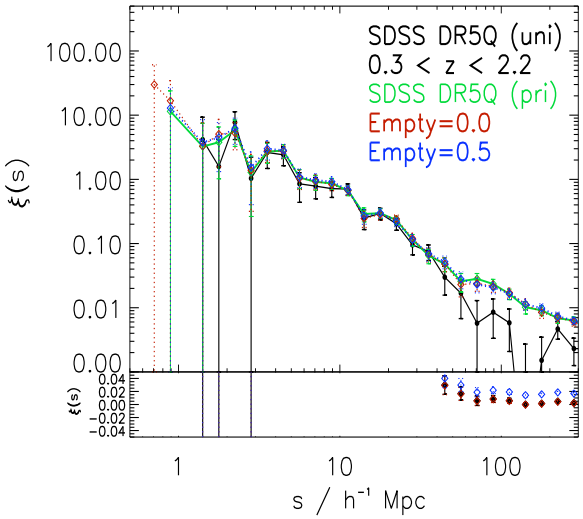

Figure C7. The SDSS DR5 Quasar $\xi(s)$ the UNIFORM sample with the sample split into the NCG and SCG.

C8 $w_p(\sigma)$, varying $\pi_{\max}$ limits

Figure C8 shows the projected correlation function $\xi(s)$ for the SDSS DR5Q Uniform samples, $0.30 < z < 2.2$, varying $\pi_{\max}$ from equation.

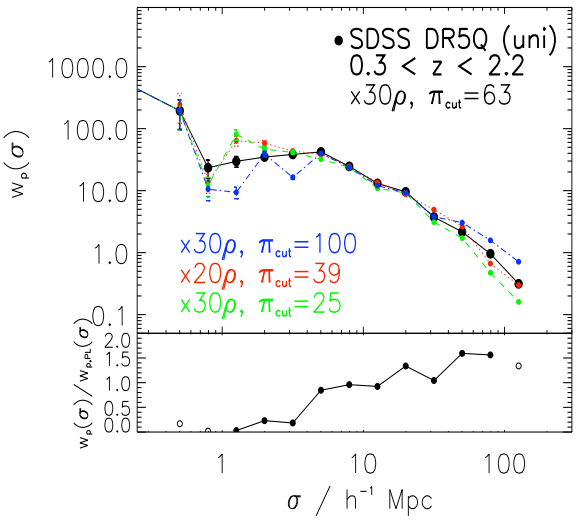

Figure C8.

REFERENCES

Abazajian K., et al., 2003, AJ, 126, 2081
Abazajian K., et al., 2004, AJ, 128, 502
Adelman-McCarthy J. K., et al., 2007, ApJS, 172, 634
Arp H., 1970, AJ, 75, 1
Baes M., Buyle P., Hau G. K. T., Dejonghe H., 2003, MNRAS, 341, L44
Becker R. H., White R. L., Helfand D. J., 1995, ApJ, 450, 559
Blanton M. R., et al., 2005, AJ, 129, 2562
Blanton M. R., Lin H., Lupton R. H., Maley F. M., Young N., Zehavi I., Loveday J., 2003, AJ, 125, 2276
Boyle B. J., Shanks T., Croom S. M., Smith R. J., Miller L., Loaring N., Heymans C., 2000, MNRAS, 317, 1014
Cannon R., et al., 2006, MNRAS, 372, 425
Croom S. M., et al., 2005, MNRAS, 356, 415
Croom S. M., et al., 2008, MNRAS, in prep., 415
Croom S. M., Shanks T., 1996, MNRAS, 281, 893
Croom S. M., Smith R. J., Boyle B. J., Shanks T., Miller L., Outram P. J., Loaring N. S., 2004, MNRAS, 349, 1397
da Ângela J., et al., 2008, MNRAS, 383, 565
da Ângela J., Outram P. J., Shanks T., Boyle B. J., Croom S. M., Loaring N. S., Miller L., Smith R. J., 2005, MNRAS, 360, 1040
Davis M., Peebles P. J. E., 1983, ApJ, 267, 465
Eisenstein D. J., et al., 2001, AJ, 122, 2267
Farrah D., et al., 2006, ApJ Lett., 641, L17
Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, AJ, 111, 1748
Gunn J. E., et al., 1998, AJ, 116, 3040
Gunn J. E., et al., 2006, AJ, 131, 2332
Guzzo L., et al., 2008, Nature, 451, 541
Hamilton A. J. S., 1992, ApJ Lett., 385, L5
Hawkins E., et al., 2003, MNRAS, 346, 78
Hawkins M. R. S., Reddish V. C., 1975, Nature, 257, 772
Hennawi J. F., et al., 2006, AJ, 131, 1
Hogg D. W., Finkbeiner D. P., Schlegel D. J., Gunn J. E., 2001, AJ, 122, 2129
Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Robertson B., Springel V., 2006, ApJS, 163, 1
Hopkins P. F., Lidz A., Hernquist L., Coil A. L., Myers A. D., Cox T. J., Spergel D. N., 2007, ApJ, 662, 110
Hoyle F., Outram P. J., Shanks T., Boyle B. J., Croom S. M., Smith R. J., 2002, MNRAS, 332, 311
Iovino A., Shaver P. A., 1988, ApJ Lett., 330, L13
Ivashchenko G., 2007, ArXiv:0710.3659
Ivezić Ž., et al., 2004, Astronomische Nachrichten, 325, 583
Kaiser N., 1987, MNRAS, 227, 1
Kerscher M., Szapudi I., Szalay A. S., 2000, ApJ Lett., 535, L13

Kundic T., 1997, ApJ, 482, 631
La Franca F., Andreani P., Cristiani S., 1998, ApJ, 497, 529
Landy S. D., Szalay A. S., 1993, ApJ, 412, 64
Lidz A., Hopkins P. F., Cox T. J., Hernquist L., Robertson B., 2006, ApJ, 641, 41
Lupton R., Gunn J. E., Ivezić Z., Knapp G. R., Kent S., 2001, in Harnden Jr. F. R., Primini F. A., Payne H. E., eds, Astronomical Data Analysis Software and Systems X Vol. 238 of Astronomical Society of the Pacific Conference Series, The SDSS Imaging Pipelines. p. 269
Magliocchetti M., Maddox S. J., Lahav O., Wall J. V., 1999, MNRAS, 306, 943
Martini P., Weinberg D. H., 2001, ApJ, 547, 12
Moscardini L., Coles P., Lucchin F., Matarrese S., 1998, MNRAS, 299, 95
Mountrichas G., Shanks T., 2007, MNRAS, 380, 113
Myers A. D., Brunner R. J., Nichol R. C., Richards G. T., Schneider D. P., Bahcall N. A., 2007, ApJ, 658, 85
Myers A. D., Brunner R. J., Richards G. T., Nichol R. C., Schneider D. P., Bahcall N. A., 2007, ApJ, 658, 99
Myers A. D., et al., 2006, ApJ, 638, 622
Myers A. D., Outram P. J., Shanks T., Boyle B. J., Croom S. M., Loaring N. S., Miller L., Smith R. J., 2005, MNRAS, 359, 741
Osmer P. S., 1981, ApJ, 247, 762
Overzier R. A., Röttgering H. J. A., Rengelink R. B., Wilman R. J., 2003, Astron. & Astrophys., 405, 53
Padmanabhan N., White M., Norberg P., Porciani C., 2008, ArXiv:0802.2105
Peacock J. A., 1999, Cosmological Physics. Cosmological Physics, by John A. Peacock, pp. 704. ISBN 052141072X. Cambridge, UK: Cambridge University Press, January 1999.
Peacock J. A., et al., 2001, Nature, 410, 169
Peebles P. J. E., 1980, The Large-Scale Structure of the Universe. Princeton University Press.
Percival W. J., et al., 2007, ApJ, 657, 645
Phillipps S., Fong R., Fall R. S. E. S. M., MacGillivray H. T., 1978, MNRAS, 182, 673
Pier J. R., Munn J. A., Hindsley R. B., Hennessy G. S., Kent S. M., Lupton R. H., Ivezić Ž., 2003, AJ, 125, 1559
Porciani C., Magliocchetti M., Norberg P., 2004, MNRAS, 355, 1010
Porciani C., Norberg P., 2006, MNRAS, 371, 1824
Richards G. T., et al., 2002, AJ, 123, 2945
Richards G. T., et al., 2006, AJ, 131, 2766
Ross N. P., et al., 2007, MNRAS, 381, 573
Sánchez A. G., Baugh C. M., Percival W. J., Peacock J. A., Padilla N. D., Cole S., Frenk C. S., Norberg P., 2006, MNRAS, 366, 189
Schneider D. P., et al., 2007, AJ, 134, 102
Scranton R., et al., 2002, ApJ, 579, 48
Scranton R., et al., 2005, ApJ, 633, 589
Serber W., Bahcall N., Ménard B., Richards G., 2006, ApJ, 643, 68
Shanks T., Fong R., Green M. R., Clowes R. G., Savage A., 1983, MNRAS, 203, 181
Shen Y., et al., 2007, AJ, 133, 2222
Smith J. A., et al., 2002, AJ, 123, 2121
Spergel D. N., et al., 2007, ApJS, 170, 377
Stoughton C., et al., 2002, AJ, 123, 485
Strauss M. A., et al., 2002, AJ, 124, 1810
Tegmark M., et al., 2004, ApJ, 606, 702
Tucker D. L., et al., 2006, Astronomische Nachrichten, 327, 821
Vanden Berk D. E., et al., 2005, AJ, 129, 2047
Wyithe J. S. B., Loeb A., 2005, ApJ, 621, 95
Wyithe J. S. B., Padmanabhan T., 2006, MNRAS, 366, 1029
Yahata K., et al., 2005, PASJ, 57, 529
York D. G., et al., 2000, AJ, 120, 1579
Zehavi I., Blanton M. R., Frieman J. A., Weinberg D. H., Waddell P., Yanny B., York D. G., 2002, ApJ, 571, 172