

Luminosity- and redshift-dependent quasar clustering

Cristiano Porciani¹[★] and Peder Norberg^{1,2}

¹*Institute for Astronomy, ETH Zürich, CH-8093 Zürich, Switzerland*

²*SUPA, †Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ*

Accepted 2006 July 13. Received 2006 July 6; in original form 2006 February 28

ABSTRACT

We present detailed clustering measurements for a flux-limited sample of $\sim 14\,000$ quasars extracted from the 2dF QSO Redshift Survey in the redshift range $0.8 < z < 2.1$. After splitting the sample into three redshift bins and each of them into six luminosity intervals, we estimate the quasar-projected auto- and cross-correlation functions at a given redshift for separations $3 \lesssim r/h^{-1} \text{ Mpc} \lesssim 20$. Fitting the data with a biased cold dark matter model and using a frequentist analysis (the F -test), we find that models with luminosity-dependent clustering are statistically favoured at the 95 per cent confidence level for $z > 1.3$. On the other hand, a number of tests based on information theory and Bayesian statistics show only marginal evidence for luminosity-dependent clustering. Anyway, the quality of the data is not good enough to accurately quantify how quasar biasing depends on luminosity. We critically discuss the limitations of our data set and show that a much larger sample is needed to rule out current models for luminosity segregation. Studying the evolution of the clustering amplitude with redshift, we detect an increase of the quasar correlation length with lookback time at the 99.3 per cent confidence level. Adopting the concordance cosmological model, we discuss the evolution of quasar biasing with cosmic epoch and show that quasars are typically hosted by dark matter haloes with mass $\sim 10^{13} M_{\odot}$.

Key words: galaxies: active – quasars: general – cosmology: observations – cosmology: theory – large-scale structure of Universe.

1 INTRODUCTION

It is widely believed that quasars are powered by accretion on to supermassive black holes. However, a detailed understanding of the physical processes leading to quasar activity (and their connection with galaxy formation) is still lacking.

Simple semi-analytic models associate quasars with galaxy major mergers and assume a tight relation between their instantaneous luminosity and the mass of the central black hole, M_{bh} (Kauffmann & Haehnelt 2000; Volonteri, Haardt & Madau 2003; Wyithe & Loeb 2003). The fraction of gas accreted on to the black hole during each merger is chosen to match the observed relation between the velocity dispersion of the bulge and M_{bh} (Ferrarese & Merritt 2000). This ends up producing a correlation between the quasar luminosity and the mass of the host dark matter halo. Since the clustering properties of dark matter haloes strongly depend on their mass, the quasar-clustering amplitude is thus expected to sensibly depend on luminosity.

Recent numerical simulations of galaxy mergers including black hole accretion and feedback have cast some doubts on this picture (Springel, Di Matteo & Hernquist 2005). These numerical experi-

ments suggest that a given black hole produces quasar activity with a wide range of luminosities (Hopkins et al. 2005). During its active phase, the black hole is most likely observed as a relatively low-luminosity quasar with a small Eddington ratio. For a short period of time, however, its emission reaches its peak value (close to the Eddington luminosity) which is indeed proportional to the mass of the powering black hole. Based on these models, Lidz et al. (2006) conclude that quasar clustering should depend only weakly on luminosity.

From the observational point of view, only recently quasar samples have grown big enough (in terms of number of objects) to attempt the study of the clustering amplitude as a function of luminosity. By analysing the galaxy–quasar cross-correlation at $1.8 \lesssim z \lesssim 3.5$, Adelberger & Steidel (2005) found no evidence for luminosity-dependent clustering. They used 79 quasars spanning 4.4 orders of magnitude in absolute luminosity which have been divided into two luminosity bins. Larger samples are obviously required to confirm this result. No indication of luminosity-dependent clustering is also found from the angular clustering of quasars with photometric redshifts in the Sloan Digital Sky Survey (Myers et al. 2006). Croom et al. (2002, 2005) studied the redshift-space clustering amplitude of 2dF quasars as a function of their apparent magnitude. Even though these authors initially found weak evidence for brighter quasi-stellar objects (QSOs) being more strongly clustered, their most recent analysis does not show such a trend. These two

[★]E-mail: porciani@phys.ethz.ch

†The Scottish Universities Physics Alliance.

studies, however, do not address the issue of luminosity-dependent clustering. In fact, they consider magnitude-limited samples within a broad redshift range ($0.3 < z < 2.9$ corresponding to ~ 8 Gyr in the currently favoured cosmological model) and totally ignore any changes in the clustering signal as a function of cosmic epoch. Moreover, their redshift-space analysis complicates the interpretation of the clustering amplitudes, as the effect of non-linear peculiar velocities could also depend on luminosity and redshift. For these reasons, their null result does not necessarily imply that quasar clustering is independent of luminosity.

In this paper, which is a follow-up analysis of Porciani, Magliocchetti & Norberg (2004, PMN04 hereafter), we study the clustering properties of $\sim 14\,000$ quasars extracted from the complete 2dF QSO Redshift Survey (2QZ; Croom et al. 2004). Our goal is to accurately measure the real-space clustering amplitude of 2dF quasars as a function of redshift and absolute luminosity. To do this, we first split our quasar sample into three redshift bins and each subsample into six luminosity intervals. We then compute, in each redshift range, the associated projected auto- and cross-correlation functions, which, by construction, are not affected by redshift-space distortions.

Using the largest quasar sample presently available, our results reveal a statistically significant evolution of the clustering length with redshift (as found already in PMN04) but only weak evidence for luminosity-dependent quasar clustering.

The layout of the paper is as follows. In Section 2, we describe our quasar sample and how we split it both in redshift and in luminosity. In Section 3, we measure the projected auto- and cross-correlation functions of the different subsamples. The issue of luminosity-dependent clustering is addressed in Section 4. Using a Monte Carlo Markov Chain, we estimate the quasar-clustering amplitude as a function of redshift and luminosity. A number of robust tests are then used to evaluate the statistical significance of the measured luminosity dependence in a given redshift range. A critical discussion of the limitations of our data and of possible future strategies is also presented here. In Section 5, we focus on the pure redshift dependence of the quasar-clustering amplitude and we provide fitting functions for the evolution of the quasar correlation length and bias with cosmic epoch. All our results are summarized in Section 6.

Throughout this paper, we assume a ‘concordance’ cosmological model with mass density parameter $\Omega_0 = 0.3$ (with a baryonic contribution $\Omega_b = 0.049$), vacuum energy density parameter $\Omega_\Lambda = 0.7$ and present-day value of the Hubble constant $H_0 = 100 h \text{ km s}^{-1} \text{ Mpc}$ with $h = 0.7$. We also adopt a cold dark matter (CDM) power spectrum with primordial spectral index $n = 1$ and with normalization fixed by σ_8 , the rms linear density fluctuation within a sphere with radius of $8 h^{-1} \text{ Mpc}$.

2 QUASAR SAMPLE DEFINITION

The 2QZ includes 23 338 quasars which span a wide redshift range ($0.3 \lesssim z \lesssim 2.9$) and are spread over 721.6 deg^2 on the sky (see Croom et al. 2004). In order to minimize systematic effects, we restrict our analysis to regions with spectroscopic (photometric) completeness larger than 70 (90) per cent, which limits the redshift range to $0.5 < z < 2.1$. Only quasars brighter than $M_{\text{bj}} - 5 \log_{10} h = -21.7$ are considered, which ensures the exclusion of quasars for which the contribution from the host galaxy may have led to a misidentification of the source.

In order to make a physically motivated analysis and isolate evolutionary effects, we subdivide our sample into three redshift bins. As in PMN04, we require a similar number of quasars to lie in each

Table 1. Main properties of our data sets. The subscripts min, max and med, respectively, denote the minimum, maximum and median values of a variable.

z_{min}	z_{max}	Data	N_{QSO}	z_{med}	M_{max}	M_{med}	M_{min}
$M_{\text{bj}} - 5 \log_{10} h$							
0.8	1.3	F25	1232	0.93	-21.7	-22.4	-22.7
0.8	1.3	F50	2464	0.99	-21.7	-22.7	-23.2
0.8	1.3	F75	3696	1.04	-21.7	-22.9	-23.7
0.8	1.3	B75	3698	1.12	-22.7	-23.4	-25.3
0.8	1.3	B50	2466	1.14	-23.1	-23.6	-25.3
0.8	1.3	B25	1234	1.15	-23.6	-24.0	-25.3
1.3	1.7	F25	1285	1.44	-22.8	-23.3	-23.5
1.3	1.7	F50	2571	1.48	-22.8	-23.5	-23.9
1.3	1.7	F75	3857	1.49	-22.8	-23.6	-24.3
1.3	1.7	B75	3858	1.53	-23.4	-24.0	-26.0
1.3	1.7	B50	2572	1.53	-23.8	-24.3	-26.0
1.3	1.7	B25	1286	1.55	-24.3	-24.7	-26.0
1.7	2.1	F25	1081	1.83	-23.4	-23.8	-24.0
1.7	2.1	F50	2163	1.86	-23.4	-24.0	-24.3
1.7	2.1	F75	3244	1.87	-23.4	-24.1	-24.9
1.7	2.1	B75	3245	1.91	-23.9	-24.5	-26.4
1.7	2.1	B50	2163	1.91	-24.3	-24.8	-26.4
1.7	2.1	B25	1082	1.92	-24.8	-25.2	-26.4

redshift bin and that each subsample covers a not too different interval of cosmic time. To better satisfy these criteria (see PMN04 for further details), we impose an additional redshift cut so as to keep only quasars within $0.8 < z < 2.1$. With this selection, we end up with nearly 14 000 2QZ quasars that we split into the redshift intervals: $0.8 < z < 1.3$, $1.3 < z < 1.7$ and $1.7 < z < 2.1$ (containing each between ~ 4300 and ~ 4900 quasars).

To study the dependence of the quasar-clustering amplitude on luminosity, we further divide each subsample into six complementary sets based on quasar absolute luminosity: (F25, B75), (F50, B50) and (F75, B25), where F_x and B_y correspond to the x per cent faintest quasars and the y per cent brightest quasars (in terms of their absolute magnitude). Table 1 lists the main properties of each sample, including the number of quasars (Column 4), the median redshift (Column 5), the absolute magnitude range (Columns 6 and 8) and the sample median absolute magnitude (Column 7).

3 QUASARS CLUSTERING ANALYSIS

3.1 Estimating the correlation functions

The simplest statistic which can be used to quantify clustering in the observed quasar distribution is the 2-point correlation function in redshift space, $\xi^q(r_\perp, \pi)$. To measure this quantity, we first build a catalogue of unclustered points which has the same angular and radial selection function as the data. The radial selection function is obtained by heavily smoothing the observed quasar comoving distance distribution, $\mathcal{N}(r)$. The quasar autocorrelation function is then estimated by comparing the probability distribution of quasar and random pairs on a two-dimensional grid of separations (r_\perp, π) . We use both the Landy–Szalay estimator (Landy & Szalay 1993) and the Hamilton estimator (Hamilton 1993):

$$\xi_{\text{LS}}^q = \frac{\text{DD} - 2\text{DR} + \text{RR}}{\text{RR}}, \quad \xi_{\text{H}}^q = \frac{\text{DD} \cdot \text{RR}}{(\text{DR})^2} - 1 \quad (1)$$

where DD, DR and RR are the suitably normalized numbers of weighted data–data, data–random and random–random pairs in each

bin.¹ As expected, the two estimators give comparable answers within the errors. In what follows we only present results obtained with the Landy–Szalay estimator. Similarly, we use the appropriately symmetrized Landy–Szalay estimator to compute the quasar cross-correlation function

$$\xi_{LS}^q = \frac{D_i D_j - (D_i R_j + D_j R_i) + R_i R_j}{R_i R_j}, \quad (2)$$

where $D_i D_j$, $D_i R_j$, $D_j R_i$ and $R_i R_j$ are the suitably normalized numbers of weighted data and random pairs from samples i and j .

To avoid redshift-space distortions, and determine the quasar-clustering amplitude in real space, one can then use the ‘projected correlation function’ which is obtained by integrating $\xi^q(r_\perp, \pi)$ in the π direction:

$$\frac{\Xi^q(r_\perp)}{r_\perp} = \frac{2}{r_\perp} \int_0^\infty \xi^q(r_\perp, \pi) d\pi. \quad (3)$$

With the current quasar sample, we find that a reliable measure of $\xi^q(r_\perp, \pi)$ is only achievable on scales $\pi \lesssim 50 h^{-1}$ Mpc. In order to avoid the measured signal to be dominated by noise, we limit the integration to an upper limit, π_{\max} . As explained in PMN04, $\pi_{\max} = 45 h^{-1}$ Mpc is an adequate choice, and we have checked that our results are not sensitive to the precise value adopted.

As in PMN04, we use a blockwise bootstrap resampling technique to estimate the uncertainties of our clustering measurements. For each redshift interval, we divide each of the two 2QZ areas into 28 equal-volume regions, containing, within a factor of 2, the same number of quasars. We build hundreds of bootstrap samples (see Section 3.3), each of them composed by 56 subsamples (28 for each 2QZ region) randomly drawn (allowing repetitions) from the set described above. We measure the projected auto- and cross-correlation functions (the minimum size of a subvolume is nearly $180 h^{-1}$ Mpc for the low-redshift sample, $250 h^{-1}$ Mpc for the medium-redshift sample and $220 h^{-1}$ Mpc for the high-redshift; since we measure correlations on scales smaller than $20 h^{-1}$ Mpc, our error estimates should hardly be affected by edge effects and/or missing large-scale structure²) for all the bootstrap samples. For each r_\perp , we identify the rms variation of Ξ^q over the bootstrap samples with the 1σ error for the projected auto- and cross-correlation function. Note that our bootstrap uncertainties account for both sampling and estimation errors.

Our method for estimating errors relies on the fact that our data set is statistically representative of the quasar distribution in the Universe.³ However, this cannot be true for bins of spatial separations which contain just a few quasar pairs. Therefore, in what follows, we ignore clustering results obtained with less than 40 quasar pairs. Depending on the sample, this corresponds to $r_\perp < 2 - 6 h^{-1}$ Mpc.

Our results are presented in Fig. 1 (filled squares with errorbars). All the correlations in a given redshift bin have similar amplitudes and are rapidly decreasing with r_\perp . As expected, errorbars are smaller for the largest quasar samples.

¹ Note that, in this case, there is no need to use the standard J_3 (minimum variance) weighting scheme since the mean density of quasars, n_q , is so low that $1 + 4\pi J_3 n_q \simeq 1$ for any reasonable quasar-clustering amplitude.

² We have also checked that doubling the size of the subvolumes does not significantly alter the diagonal errors.

³ Our blockwise bootstrap errors are in good agreement with those obtained using mock galaxy catalogues extracted from high-resolution numerical simulations. Discrepancies between the two types of errors are of the order of 20 per cent or less on the scales used in this paper.

3.2 Modelling the data

Given the relatively large errorbars, it is hard to spot any luminosity dependence of quasar clustering by simply looking at the correlations in Fig. 1. For this reason, we use a reference model to fit the data. This is obtained by assuming that each sample is characterized by a linear bias parameter which does not depend on spatial separation within the range of scales we analyse. In other words, we assume that the quasar 2-point auto- and cross-correlations scale as

$$\xi_{ij}^q(r) = b_i b_j \xi(r), \quad (4)$$

where i and j are the labels of two quasar subsamples within the same redshift bin, r denotes the comoving separation between quasar pairs and b_i is the bias parameter of the i th subsample with respect to the mass autocorrelation function, $\xi(r)$, computed as in Peacock & Dodds (1996; nearly undistinguishable results are found using the method by Smith et al. 2003) assuming a value for σ_8 .⁴ Within the framework of halo models (see PMN04, for a direct application to quasars), the assumption that the amplitude of the cross-correlation function between haloes of different masses scales as the geometric mean of the correlations of the individual haloes is very natural in the two-halo regime. We have checked against numerical simulations that, for dark matter haloes, the assumption holds over the redshift range and scales considered here, i.e. $3 \lesssim r \lesssim 20 h^{-1}$ Mpc. The corresponding projected correlation functions are obtained through this simple integral relation:

$$\Xi_{ij}^q(r_\perp) = 2 b_i b_j \int_{r_\perp}^\infty \frac{r \xi(r)}{(r^2 - r_\perp^2)^{1/2}} dr. \quad (5)$$

3.3 Fitting the data with correlated errors

We use a minimum least-squares method (corresponding to a maximum likelihood method in the case of Gaussian errors) to determine the bias parameters that best describe the clustering data. In each redshift range, we sample the six-dimensional parameter space (one bias parameter per luminosity range) using a Markov Chain Monte Carlo method. A principal component analysis (see e.g. Porciani & Giavalisco 2002) is used here to deal with correlated errorbars. The principal components of the errors are computed by diagonalizing the covariance matrix obtained by resampling the data with the bootstrap method. The objective function (the usual χ^2 statistic) is then obtained by only considering the most significant principal components (i.e. those contributing the largest fraction of the total variance).⁵ Unfortunately, there is no unique objective way of deciding how many principal components should be considered for a given data set. Considering too few components (and thus discarding information contained in the data), one obtains too good values for the objective function (i.e. $\chi_{\min}^2/\text{d.o.f.} \ll 1$, where d.o.f.

⁴ Note that samples with different luminosities have slightly different median redshifts. However, the variation of the mass correlation function among them is of the order of a few per cent, much smaller than the uncertainty in the quasar-clustering amplitude. This implies that we can safely use the same $\xi(r)$ (evaluated at the median redshift of the redshift bin) for all the luminosity subsamples. We have tested that this does not influence our results.

⁵ The covariance matrix obtained from bootstrap resampling is only an estimate of the true one and contains an intrinsic uncertainty. The errors in its component propagate in the calculation of its eigenvalues and eigenvectors. It is therefore recommended to consider only the principal components corresponding to the largest eigenvectors in the fitting procedure (see section 4.2 in Porciani & Giavalisco 2002, for further details).

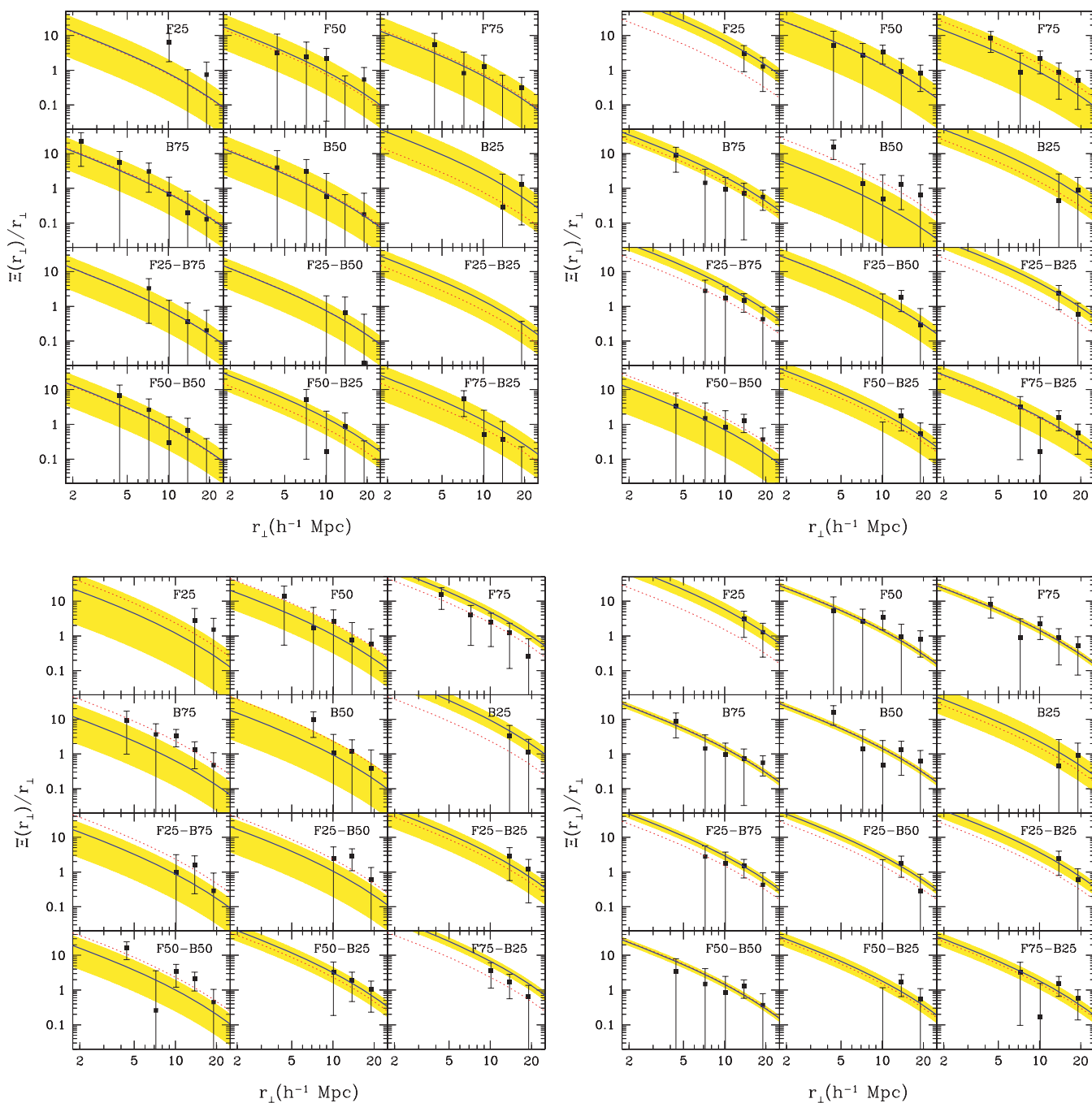


Figure 1. Top left-hand panel: the projected autocorrelation function (top two rows) and cross-correlation function (bottom two rows) for the low-redshift quasar samples ($0.8 < z < 1.3$). Only measurements obtained with more than 40 quasar pairs are plotted (filled squares with errorbars). Bins with less pairs are omitted and not considered for model fitting. For each luminosity subsample, the shaded band show the central 68 per cent range of best-fitting correlation functions (assuming that each luminosity interval is associated with a different bias parameter). The mean value of the marginalized bias distribution is represented with a solid line. For reference, we also plot the best-fitting correlation function for the entire low-redshift sample derived in PNM04 (dotted line). Top right-hand and bottom left-hand panels: same as the top left-hand panel, but for the medium- ($1.3 < z < 1.7$) and high-redshift ($1.7 < z < 2.1$) quasar samples, respectively. Bottom right-hand panel: same as the top right-hand panel but using a three-parameter model where only the B25 and F25 subsamples have different bias parameters from all the others.

indicates the number of degrees of freedom) that correspond to unrealistically large uncertainties for the fitted parameters. On the other hand, considering too many components (and thus, most likely, introducing noise), one often obtains bad fits (i.e. $\chi^2_{\min}/\text{d.o.f.} \gg 1$) that correspond to unrealistically small errors for the bias param-

eters. In factor analysis, a number of empirical methods (e.g. Kaiser criterion, scree test) are often employed to select the number of significant components. The robustness of these techniques for model fitting is, however, rather weak. We thus decided to select the number of components to account for a fixed fraction (95 per cent)

of the bootstrap variance. We have checked that this compression guarantees a detailed reconstruction of the model-data residuals and simultaneously avoid the χ^2 to be dominated by deviations along the components corresponding to tiny eigenvalues. Moreover, the reduced χ^2 of the best-fitting models obtained this way is of the order of unity in all cases, as expected theoretically. In what follows, we thus use the symbol χ_{95}^2 to denote the χ^2 function computed by considering the first principal components that, in total, account for 95 per cent of the variance. This corresponds to 19 principal components (out of 42 data points) for the high-redshift sample, 21 (out of 45) for the median-redshift sample and 22 (out of 47) for the low-redshift sample.

Since the bootstrap technique is very time consuming, a careful choice of the number of resamplings is required. For this reason, we performed a number of Monte Carlo simulations checking for the stability of eigenvalues, eigenvectors and χ_{95}^2 estimates. In practice, we first bootstrapped our data and computed the corresponding covariance matrix \mathbf{C}_{ij} . We then built a large number of realizations of a Gaussian process with covariance \mathbf{C}_{ij} . Finally, we estimated the covariance matrix of the Gaussian process from a finite number of realizations, N , and studied its dependence on N . We found that a few hundred bootstrap resamplings are needed for a robust estimation of the χ_{95}^2 function. To overcome the Gaussian hypothesis, we also studied the convergence of the covariance matrix obtained by directly bootstrapping our data. We found that estimates of χ_{95}^2 converge (i.e. present negligible scatter) when nearly 250 bootstrap resamplings are used. This has been obtained by using 500 resamplings of the median redshift sample. To be on the safe side, we thus decided to use at least 350 bootstrap resamplings for each redshift bin. We note that this implies over 12 000 correlation function estimates, with each containing at least 50 000 random points.

As an additional test of the robustness of our results, we checked, a posteriori, how much our best-fitting models depend on the number of adopted components, N_{pc} . This is discussed in detail in Section 4 where we present our results.

4 LUMINOSITY-DEPENDENT CLUSTERING

In this section, we present the results obtained by fitting the auto- and cross-correlation functions presented in Fig. 1 with the model given in equation (4). Our aim is to quantify how the bias parameter of optically bright quasars depends on redshift and luminosity.

The best-fitting functions obtained with the Markov Chain Monte Carlo method are overplotted to the data in Fig. 1 (shaded regions). Note that, for all redshift bins, they accurately describe the quasar auto- and cross-correlations in the whole range of separations under analysis. In the left-hand panel of Fig. 2, we plot σ_8^q (i.e. $b \sigma_8$) as a function of absolute magnitude. As expected, the bias parameter of all quasars lying in a given redshift bin (measured by PMN04 and represented with an open symbol) lies in between the values found here. Taken at face value, the data show some evidence for luminosity segregation; however, samples at different redshifts show different trends. The brightest quasars in the high-redshift bin seem to be more strongly clustered than the others. In the medium-redshift bin, the bias seems to follow a U shape: the faintest and the brightest quasars in the set are typically more strongly clustered than M_{bj}^* quasars. On the other hand, the low-redshift bin does not show any particular trend: the bias keeps constant with luminosity.

Note, however, that correlations between errorbars in Fig. 2 might create spurious trends with luminosity. Moreover, all the different subsamples in a given redshift bin have consistent bias parameters at the 2σ level. Given these concerns, we want to use a robust statistical test to investigate whether our data show any sign of luminosity-dependent clustering.

4.1 Does quasar clustering depend on luminosity?

If there is no luminosity segregation, all our data should be described by one single bias parameter (i.e. all the b_i should assume the same identical value). If, instead, some luminosity bins

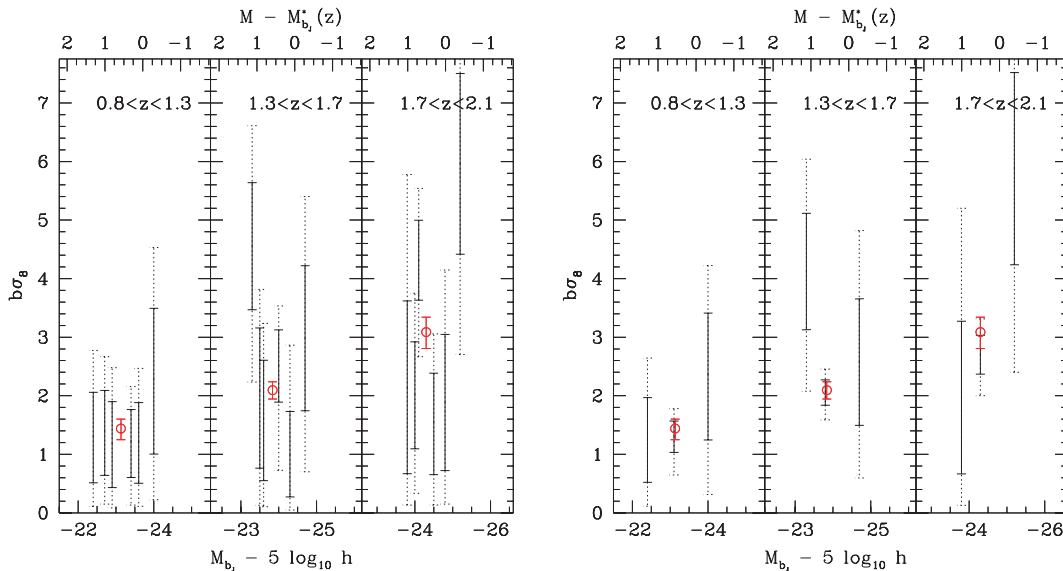


Figure 2. Left-hand panel: σ_8^q , i.e. $b \sigma_8$, as function of the median absolute magnitude for the low (left), medium (centre) and high (right) redshift bins. For each sample, the errorbars indicate the central 68.3 (solid) and 95.4 (dotted) credibility intervals of the marginalized bias probability distribution. On the upper axis, we indicate the magnitude difference with respect to $M_{\text{bj}}^*(z)$, where the later is given by equation (13) of Croom et al. (2004). The open symbol shows the result obtained by fitting the correlation function of all the quasars lying in this redshift bin (see PMN04). Right-hand panel: same as the left-hand panel but for a three-parameter fit (see the text).

show a statistically significant deviation from the overall clustering amplitude, then a description with a number of different bias parameters should be preferred. Therefore, in this section we address the issue of luminosity-dependent clustering by answering the following question: how many bias parameters are required to adequately describe the quasar auto- and cross-correlations in a given redshift bin? In other words, we want to understand how many different bias parameters can be reliably measured in a statistically significant way. This is a classical problem of model selection where we want to find the proper tradeoff between goodness of fit (in the χ^2 sense) and complexity (in terms of number of free parameters). We use five different methods: the F -test, the Akaike information criterion (AIC), the Bayes factors (BFs), the Bayesian information criterion (BIC) and the deviance information criterion (DIC; see the Appendix for a brief review). We consider three different models. In the first one, there is no clustering segregation with luminosity and all the luminosity bins at a given redshift are associated with the same bias parameter. In the second model, we use three bias parameters (one each for the brightest, B25, and faintest samples, F25, and one for the remaining quasars: hereafter those samples are referred to as B, M and F, for bright, medium and faint, respectively). In the third model, each luminosity bin has its own bias parameter, as already described in the previous sections. In Fig. 2, we show that the results of the three- and six-parameters models are qualitatively and quantitatively similar.

Before comparing the different models, it is worth noting that, based on the χ^2 statistic, a single parameter fit gives an acceptable description of the data for all redshift bins. Of course, using additional parameters reduces the minimum χ^2 of the best-fitting models. We want now to understand whether this χ^2_{\min} reduction is statistically significant.

Performing the F -test and using the 95 per cent confidence level as a threshold to prefer a model with respect to another, we find that the low-redshift sample is best fitted by a one-parameter model, while the medium- and high-redshift samples are best described by a six-parameter model. Similarly, all the Bayesian and information theory based tests mentioned above clearly indicate that the low-redshift sample is best described by a one-parameter model. On the other hand, for the medium- and high-redshift samples, the situation is more confused. Because of the different penalties for model complexity, depending on the adopted test, either the six- or the one-parameter model is the preferred one. However, with the exception of the six-parameter fit with the Akaike criterion, no model can be rejected with high confidence (see Table 2 for a summary of the results). Therefore, we conclude that, while there is no evidence of clustering segregation with luminosity in the low-redshift sample, we find marginal evidence for it in the medium- and high-redshift samples.

Table 2. Number of parameters of the preferred models according to various model-selection criteria. When the evidence for the best model is not strong, we list all the acceptable models (in order of decreasing evidence). We use a 95 per cent confidence level for the F -test, AIC and DIC while we discard models with substantial evidence according to the Kass–Raftery criterion (see the Appendix) for BF and BIC.

Criterion	$0.8 < z < 1.3$	$1.3 < z < 1.7$	$1.7 < z < 2.1$
F -test	1	6	6
AIC	1–3	1–3	1–3
BF	1–3–6	6–3–1	6–3–1
BIC	1	1–3–6	1–3–6
DIC	1–3–6	6–3–1	6–3–1

Following a suggestion of the referee, we have also tried to combine data from the different redshift bins by using a simple parametrization of the luminosity- and redshift-dependent bias. For simplicity, we adopted a linear relation with quasar luminosity $b(M_{\text{bj}}, z) = b_0(z) + b_1 10^{-0.4[M_{\text{bj}} - M_{\text{bj}}^*(z)]}$. In this case, the 68.3 per cent confidence interval for the luminosity-dependent parameter is $b_1 = 0.3^{+0.6}_{-0.7}$. This confirms that the evidence for clustering segregation with luminosity is marginal in our sample. Similar conclusions are drawn adopting quadratic or cubic relations for the luminosity dependence of the bias parameter. In all cases, the best-fitting model favours a larger bias parameter for the brightest quasars in our sample. However, the statistical significance of the result is low.

4.2 Limitations of the PCA method

A number of systematic effects can alter the model-fitting procedure (and thus the statistical significance of the results). In particular, for strongly correlated data, the choice of the number of principal components used in the fit plays a delicate role. We fixed N_{pca} assuming that our bootstrap errors are accurate estimates of the real uncertainties and using the χ^2 test. Had we decided to include a few more components to account for 99 per cent of the variance (which is equivalent to consider seven additional components for each redshift sample and still gives acceptable values of the reduced χ^2), the evidence for clustering segregation would have been stronger. In Fig. 3, we compare the solutions for the three-parameter models obtained by minimizing χ^2_{95} (left-hand panel) and χ^2_{99} (right-hand panel). What is shown here are the (marginal) joint probability distributions of pairs of bias parameters (for the faint, medium and bright samples). Clustering segregation with luminosity is suggested by the data whenever the contours of the joint distributions lie away from the diagonal line. For both χ^2_{95} and χ^2_{99} , this never happens for the low-redshift sample (the bottom left set of curves) and the one-parameter solution has to be preferred. At best, clustering segregation with luminosity is detected at the $\sim 2\sigma$ level (for χ^2_{95}) and at $\sim 3\sigma$ level (for χ^2_{99}) in the high- and medium-redshift samples. Note that, even though a 3σ detection is still consistent with pure statistical fluctuations, when χ^2_{99} is used, all the tests for model selection strongly prefer either the three- or the six-parameter solution with respect to the one-parameter fit. In other words, the results presented in the previous section have to be considered conservative. If one decides to use the full covariance matrix, one will infer the presence of a statistically significant clustering segregation with luminosity in the medium- and high-redshift samples. Hence, the uncertainty in the number of physical components is the main limitation of the principal component technique that we used to account for correlated errorbars. Bayesian techniques (e.g. Minka 2000) suggest that the components which contribute the last few per cent of the bootstrap variance are most likely dominated by noise. Therefore, we are confident that the results we presented in Section 4.1 are optimal and realistic. Larger data sets with smaller intrinsic uncertainties are thus needed to improve the significance of our results.

In order to cross-check our results with a method affected by different systematics, we have computed the marked correlation function of the quasars (e.g. Sheth, Connolly & Skibba 2006, and references therein) using their blue luminosity as the mark. Results for the medium redshift sample are shown in Fig. 4. Here, we define the marked correlation function (filled squares) as the ratio of the projected quasar correlation function weighted by the quasar luminosity (numerator) and the usual projected quasar correlation

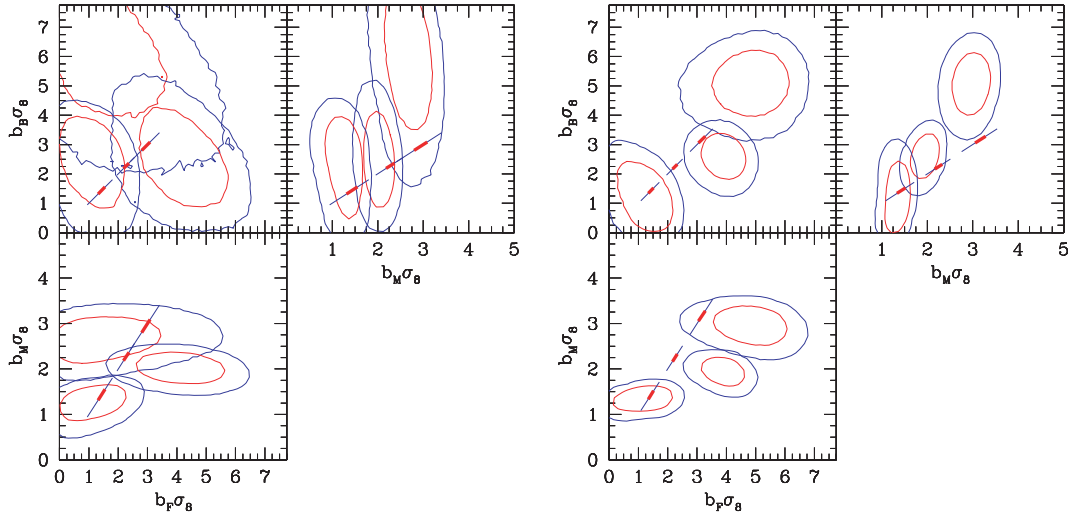


Figure 3. Left-hand panel: results obtained by fitting the data with a three-parameter model (see the text) where b_B , b_M and b_F are the bias parameters for the bright, medium and faint subsamples (at fixed redshift), respectively. The contours indicate the 68.3 and 95.4 per cent credibility intervals for the (marginalized) joint probability distribution of two bias parameters. On the diagonal, we indicate the corresponding confidence levels for the one-parameter fits. The results for the low-, medium- and high-redshift samples are plotted from left to right: $b\sigma_8$ clearly increases with redshift. Right-hand panel: same as left-hand panel but using more principal components of the errors so as to account for 99 per cent of the variance (as opposed to 95 per cent on the left-hand panel). Note that errorbars shrink and b_F is shifted by almost 1σ with respect to the left-hand panel.

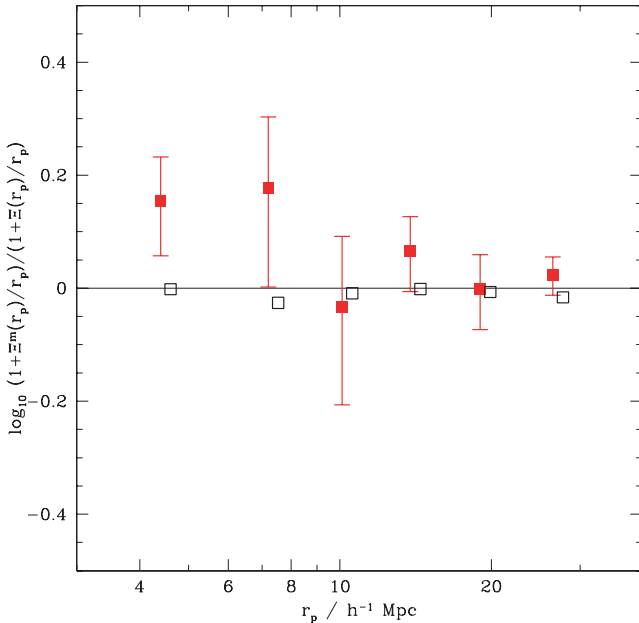


Figure 4. Marked correlation function of the medium redshift sample obtained using the blue luminosity of the quasars as the mark. The empty squares (that have been slightly displaced to the right) show the mean obtained by averaging over 100 realizations of the same measurement after randomizing the marks. Errorbars show the standard deviation around this mean. If one includes bootstrap errors into the total error budget (not shown here), the errors on the marked correlation function typically increase by 20–30 per cent.

function (denominator). By construction, this ratio is not affected by redshift-space distortions. The open symbols correspond to the mean marked correlation function obtained by averaging over one hundred realizations with randomized marks. To assess the significance of the marked statistics, one usually compares the signal with

the standard deviation around the mean of the randomized realizations. From Fig. 4, we see that, on average, the closest quasar pairs have higher luminosities, thus suggesting the presence, on those scales, of some clustering segregation. However, given the size of the errorbars, the statistical significance of this trend is rather small. We note that the errors used here only reflect the uncertainty arising from the distribution of the marks and do not include any uncertainty in the actual estimate of the correlation function. Considering the corresponding bootstrap errors (which, for instance, allow for sample variance), the uncertainties on the marked correlation function typically increase by 20–30 per cent. This makes the detection of clustering segregation even more uncertain. In summary, the analysis of the marked statistics reaches exactly the same conclusions as our main work: even though there is some evidence for luminosity-dependent clustering in our high-redshift data, our sample is too small to provide a robust detection. Similar results are obtained for the other redshift bins.

4.3 Discussion

When interpreting our results, we have to consider the limitations of both the data set and the method used to extract the bias parameters. First, the depth of the quasar sample is such that at a given redshift we can only probe a factor of 10 in luminosity. This implies that any luminosity segregation has to be observed over this relatively small luminosity range. By slicing the different redshift bins into six subsamples, our analysis tries to extract the maximum possible information from these limited data. However, to avoid to be dominated by measurement noise, we compare narrow luminosity intervals (at both the bright and the faint ends) with larger control samples. This naturally leads to the selection of partially overlapping luminosity ranges. For this reason, our samples, despite probing different absolute magnitude intervals, are not independent from each other. Note the striking difference from previous studies of galaxy clustering (see e.g. Norberg et al. 2001, 2002) where one can perform several

Table 3. Best-fitting constant-bias (Columns 4, 5 and 6) and power-law models for six complementary redshift bins. The goodness of each fit is measured by the quantity $\chi^2_{\min}/\text{d.o.f.}$ which gives the minimum value assumed by the χ^2 statistic divided by the number of degrees of freedom.

z_l	z_h	z_m	b	r_0 (h^{-1} Mpc)	$\chi^2_{\min}/\text{d.o.f.}$	r_0 (h^{-1} Mpc)	γ	$\chi^2_{\min}/\text{d.o.f.}$	$r_0^{(\gamma=1.8)}$ (h^{-1} Mpc)	$\chi^2_{\min}/\text{d.o.f.}$
0.80	1.06	0.933	$1.57^{+0.30}_{-0.37}$	$5.5^{+1.5}_{-1.8}$	3.47/3	$4.3^{+2.5}_{-4.2}$	$1.70^{+0.53}_{-0.67}$	2.02/2	$4.9^{+0.8}_{-1.1}$	2.05/3
1.06	1.30	1.185	$1.76^{+0.35}_{-0.43}$	$5.6^{+1.6}_{-1.9}$	1.07/3	$5.1^{+1.6}_{-4.3}$	$1.84^{+0.33}_{-0.68}$	0.87/2	$4.9^{+1.0}_{-1.2}$	0.89/3
1.30	1.51	1.410	$2.13^{+0.29}_{-0.33}$	$6.5^{+1.2}_{-1.3}$	1.05/3	$6.1^{+1.2}_{-2.1}$	$1.93^{+0.44}_{-0.42}$	1.44/2	$5.6^{+0.9}_{-0.9}$	1.54/3
1.51	1.70	1.602	$2.33^{+0.33}_{-0.39}$	$6.7^{+1.2}_{-1.5}$	5.33/3	$4.1^{+2.3}_{-3.9}$	$1.59^{+0.38}_{-0.55}$	4.92/2	$5.4^{+0.8}_{-0.9}$	5.17/3
1.70	1.89	1.796	$3.02^{+0.45}_{-0.53}$	$8.5^{+1.5}_{-1.9}$	0.53/3	$7.2^{+1.2}_{-2.7}$	$2.03^{+0.43}_{-0.47}$	1.19/2	$6.3^{+1.0}_{-1.3}$	1.48/3
1.89	2.10	1.987	$4.13^{+0.49}_{-0.55}$	$11.5^{+1.5}_{-1.7}$	2.76/3	$8.8^{+1.2}_{-4.9}$	$1.82^{+0.27}_{-0.49}$	2.86/2	$8.6^{+1.1}_{-1.2}$	2.87/3

independent measurements of the clustering amplitude as a function of luminosity and span a wide magnitude range.

Given these limitations and based on theoretical models of quasar clustering, should we have expected to detect some luminosity dependence in our data set? As discussed in the Introduction, models make very different predictions. From fig. 3 in Lidz et al. (2006), we infer that the models which assume a tight correlation between quasar luminosity and the mass of the host-haloes predict a difference $\Delta b = (b_{B25} - b_{F25}) \sigma_8 \sim 1$ for our high-redshift bin. On the other hand, models based on merger simulations predict $\Delta b \sim 0.2$. Similarly, the semi-analytic models by Kauffmann & Haehnelt (2002) predict $0.3 \lesssim \Delta b \lesssim 0.7$ depending on the characteristic quasar lifetime (this increases up to $\Delta b \sim 1$ at $z \simeq 1.5$). In all cases, the errorbars associated with our measurements are too large to lead to a statistically significant detection of luminosity-dependent clustering (in fact, our error on Δb is $\sigma_{\Delta b} \simeq 2.3$ and we measure $\Delta b \simeq 4.3$ which is only 1.4σ and 1.8σ away from the two reference models). Assuming that $\Delta b \simeq 1$, a sample which is nearly 50 times larger than ours (i.e. nearly all-sky) is needed to detect the luminosity dependence at the 3σ confidence level (assuming that the error scales with the sample size, n , as $n^{-1/2}$). Very likely a sample of this size will not be available in the next few years. Alternatively, one could try to beat the noise by reaching a fainter apparent magnitude limit. Fig. 3 in Lidz et al. (2006) shows that current models predict that, by reducing the median luminosity of our F25 sample by a factor of 10, one would get $0.6 \lesssim \Delta b \lesssim 2$. Note that a value of $\Delta b \sim 2$ could be ruled out at the 3σ confidence level by using a sample 12 times as large as ours. In conclusion, our analysis can only rule out (or detect) extreme luminosity-dependent biasing and future surveys are required to discriminate among current models.

5 REDSHIFT-DEPENDENT CLUSTERING

In this section, we focus on the redshift dependence of the quasar-clustering amplitude. This is done in real space, using the projected correlation function. Our study is thus complementary to the redshift-space analysis by Croom et al. (2005) and to the angular study by Myers et al. (2006) which might be affected by different systematics. In this section, we also discuss the effect of using different methods for determining the quasar correlation length.

To improve the analysis performed by PMN04 and better study the evolution with cosmic time, we split our quasar sample into six redshift bins. These are obtained by dividing into two equal-sized parts (in terms of quasar number) the redshift bins used in the previous sections and in PNM04. We first compute the projected autocorrelation functions of all the quasars which lie in a given redshift bin. We then use the method described in Section 3.2 to fit these

data with the model given in equation (4). The best-fitting models correspond to very good values of the χ^2 statistic, indicating that our models accurately describe the spatial dependence of the measured correlations. Finally, we estimate the correlation length, r_0 , by determining the scale at which the best-fitting model has $\xi^q(r) = 1$. For comparison, we also fit a power-law model $\xi^q(r) = (r/r_0)^\gamma$. We distinguish two cases where we either allow both r_0 and γ to vary or fix $\gamma = 1.8$. Our results are reported in Table 3 (together with the corresponding 68.3 per cent confidence levels) and plotted in Fig. 5. Note that the best-fitting values for the quasar correlation length derived from the power-law model are systematically lower than the CDM ones (the two estimates are anyway consistent within 1σ uncertainties). This is because, in a CDM model, the power-law index of the mass autocorrelation function is a function of the spatial separation. For instance, $\gamma_{\text{eff}} = d \ln \xi / d \ln r \simeq 1.3$ at $r = 5 h^{-1}$ Mpc, while $\gamma_{\text{eff}} \simeq 1.9$ at $r = 20 h^{-1}$ Mpc. Similarly, the projected correlation function, $\Xi(r_\perp)/r_\perp$, has a slope varying between 1.6 and 2.3 when $5 < r < 20 h^{-1}$ Mpc. Therefore, a CDM model has a much lower correlation on large scales with respect to a power-law model with the same r_0 (and $\gamma < 2$). It is then clear from equation (5) that a CDM model needs a higher overall normalization than a power-law model to fit the same projected correlation.

The evolution of the quasar correlation length between $0.8 < z < 2.1$ can be approximately described by the linear relation⁶

$$r_0 = [7.3 + 5.2(z - 1.5)]h^{-1} \text{ Mpc} \quad (6)$$

(this becomes $r_0 = [6.1 + 3.6(z - 1.5)]h^{-1}$ Mpc for the values of r_0 determined with the power-law fit to $\Xi(r_\perp)/r_\perp$). For $0.8 < z < 1.6$, our results are in good agreement with Croom et al. (2005), but we find evidence for a stronger variation at high- z (see Fig. 5).⁷ This is more evident when we use the CDM model to fit the data. The models by Kauffmann & Haehnelt (2002) with a quasar lifetime of $\sim 10^7$ yr match our data at $z < 1.7$. However, as already pointed out in PMN04, an increase in the lifetime by nearly an

⁶ A nearly perfect interpolant between our measured points is $r_0 = [5 + (\frac{1+z}{2.4})^8]h^{-1}$ Mpc. This, however, ignores the statistical uncertainties in the measure of r_0 (i.e. has $\chi^2 \ll 1$).

⁷ Our estimates of the uncertainties for r_0 and γ are nearly a factor of 2 larger than in Croom et al. (2005). This is due to a number of facts. On one hand, we find that blockwise bootstrap errors on the redshift-space correlation function are typically 30 per cent smaller than on the projected correlation function. On the other hand, Croom et al. (2005) assume statistically independent Poisson errorbars for the redshift-space correlation function at separations smaller than $50 h^{-1}$ Mpc. In particular, neglecting correlations between points at different spatial separations results into smaller errors for the fitted parameters.

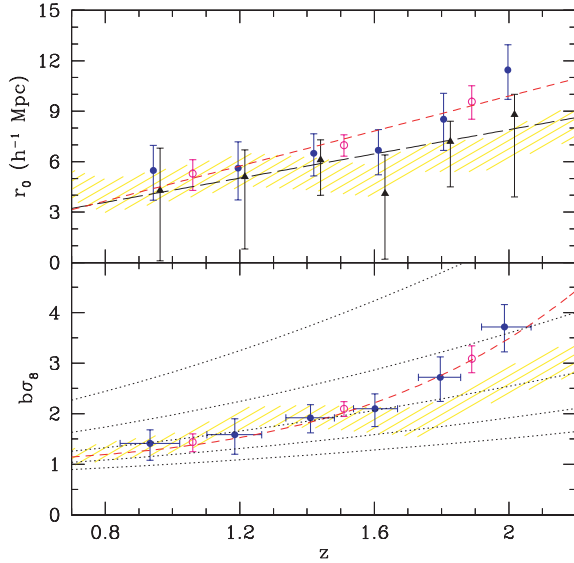


Figure 5. Top panel: redshift evolution of the quasar correlation length. Filled circles with errorbars indicate the best-fitting values obtained using a biased CDM model. The short dashed line is the analytical fit to the data given in equation (6). Filled triangles with errorbars refer to the best-fitting values obtained with a power-law model (where both γ and r_0 are free to vary). These points have been slightly displaced to the right to improve readability. The long dashed line is the linear fit to the data given in the main text. Empty symbols are analogous to their filled counterparts but are obtained using larger redshift intervals as in PMN04. The shaded area marks the 68.3 confidence levels obtained by Croom et al. (2005). Bottom panel: same as the top panel but for the quasar bias parameter. Horizontal errorbars mark the 16th and the 68th percentiles of the redshift distribution in a bin. The dashed line is the fitting function given in equation (7). The dotted lines indicate the bias evolution for dark matter haloes with masses 10^{12} , $10^{12.5}$, 10^{13} , $10^{13.5}$ and $10^{14} M_{\odot}$ (from bottom to top).

order of magnitude is required to reproduce the biased CDM fits at $z \sim 2$.

Given that the mass autocorrelation function rapidly increases with cosmic time, the bias parameter of our quasars has to increase with z . The bottom panel of Fig. 5 shows the evolution of b . Note that, within the range of cosmological models allowed by observations, the quasar bias parameter scales linearly with σ_8 . For this reason, we decided to plot the product $\sigma_8^q = b \sigma_8$ which is independent of the assumed value for the amplitude of the linear power spectrum. Similarly, to the correlation length, σ_8^q seems to increase rapidly for $z > 1.6$ and the bias evolution is well approximated by the relation

$$b\sigma_8 = 1 + \left(\frac{1+z}{2.5}\right)^5. \quad (7)$$

In CDM models, the clustering amplitude of massive dark matter haloes mainly depends on their mass. It is then interesting to see what halo masses correspond to the bias parameters of the 2QZ quasars. In the bottom panel of Fig. 5, we plot the evolution of σ_8^q for a number halo masses which is obtained using the model by Sheth & Tormen (1999).⁸ The observed bias evolution is well reproduced by assuming that quasars are associated with haloes of

⁸ A concordance cosmology with $\sigma_8 = 0.9$ is assumed here but results depend only slightly on σ_8 .

mass $\sim 10^{13} M_{\odot}$. This is fully consistent with the more detailed analysis performed in PMN04.

6 SUMMARY

We have used a flux limited sample of $\sim 14\,000$ 2QZ quasars brighter than $M_{\text{bj}} - 5 \log_{10} h = -21.7$ to study the quasar-clustering properties in the redshift range $0.8 < z < 2.1$. Our main results are summarized as follows.

(i) Splitting the sample in three redshift intervals each divided into six luminosity ranges and combining information from the corresponding auto- and cross-correlation functions, we find some evidence for clustering segregation with luminosity. For redshifts $z > 1.3$, a frequentist model selection technique (the F -test) prefers a multiparameter fit to the data at the 95 per cent confidence level.

(ii) A number of statistical tests based on information theory and Bayesian techniques show weak evidence for luminosity-dependent clustering at high redshift ($z > 1.3$) and no evidence at low redshift ($z < 1.3$). These results somewhat depend on the number of principal components used in the fitting procedure. Accounting for a larger fraction of the bootstrap variance increases the significance of the detection of clustering segregation with luminosity.

(iii) Larger data sets, possibly with a deeper coverage, are needed to discriminate among current models of quasar formation and to pinpoint the detailed quasar-clustering trends as a function of luminosity at a given redshift.

(iv) Splitting the sample into six complementary redshift bins, we find strong evidence for an increase of the clustering amplitude with lookback time. We detect pure quasar-clustering evolution between $z_{\text{eff}} = 0.93$ and 1.99 at the 2.7σ confidence level. A linear fit for the evolution of r_0 with redshift is given in equation (6).

(v) Accounting for the evolution of the mass density in the concordance (Λ CDM) model, we find that the high-redshift quasars ($z_{\text{eff}} = 1.99$) are ~ 2.6 times more biased than their low-redshift counterparts ($z_{\text{eff}} = 0.93$). Evolution in b is detected at the 4.3σ confidence level.

(vi) The clustering amplitude of optically selected quasars suggests that they are hosted by haloes with mass $M \sim 10^{13} M_{\odot}$ (see also PMN04).

ACKNOWLEDGMENTS

PN acknowledges support from the Zwicky Prize fellowship programme at ETH-Zürich and the PPARC PDRA fellowship held at the IfA. We thank Martin White, Andrey Kravtsov and Andrew Zentner for useful discussions. The 2QZ was compiled by the 2QZ survey team from observations made with the 2dF on the Anglo-Australian Telescope.

REFERENCES

- Adelberger K. L., Steidel C. C., 2005, *ApJ*, 627, L1
 Akaike H., 1973, in Petrox B. N., Caski F., eds, *Second Int. Symp. on Information Theory: Information Theory and Extension of the Maximum Likelihood Principle*. Akademiai Kiado, Budapest, p. 267
 Burnham K. P., Anderson D. R., 1998, *Model Selection and Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, New York
 Croom S. M., Boyle B. J., Loaring N. S., Miller L., Outram P. J., Shanks T., Smith R. J., 2002, *MNRAS*, 335, 459

- Croom S. M., Smith R. J., Boyle B. J., Shanks T., Miller L., Outram P. J., Loaring N. S., 2004, MNRAS, 349, 1397
- Croom S. M. et al., 2005, MNRAS, 356, 415
- Ferrarese L., Merritt D., 2000, ApJ, 539, L9
- Hamilton A. J. S., 1993, ApJ, 417, 19
- Han C., Carling B. P., 2000, MCMC Methods for Computing Bayes Factors: A Comparative Review, Working Paper, Division of Biostatistics, School of Public Health, University of Minnesota
- Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Robertson B., Springel V., 2005, ApJ, 630, 716
- Jeffreys H., 1961, Theory of Probability, 3rd edn. Oxford Univ. Press, Oxford
- Kass A. E., Raftery R. E., 1994, Technical Report No. 254, Dept. of Statistics, Univ. of Washington, Technical Report no. 571, Dept. of Statistics, Carnegie-Mellon University
- Kauffmann G., Haehnelt M., 2000, MNRAS, 311, 576
- Kauffmann G., Haehnelt M., 2002, MNRAS, 332, 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
- Minka T. P., 2000, Automatic Choice of Dimensionality for PCA, M.I.T. Media Laboratory Perceptual Computing Section Technical Report No. 514
- Myers A. D. et al., 2006, ApJ, 638, 622
- Norberg P. et al., 2001, MNRAS, 328, 64
- Norberg P. et al., 2002, MNRAS, 332, 827
- Peacock J. A., Dodds S. J., 1996, MNRAS, 280, L19
- Porciani C., Giavalisco M., 2002, ApJ, 565, 24
- Porciani C., Magliocchetti M., Norberg P., 2004, MNRAS, 355, 1010 (PMN04)
- Schwarz G., 1978, Ann. Stat., 6, 461
- Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
- Sheth R. K., Connolly A. J., Skibba R., 2006, MNRAS, submitted (astro-ph/0511773)
- Smith R. E. et al., 2003, MNRAS, 341, 1311
- Spiegelhalter D. J., Best N. G., Carlin B. P., van der Linde A., 2002, J. R. Stat. Soc. B, 64, 583
- Springel V., Di Matteo T., Hernquist L., 2005, MNRAS, 361, 776
- Volonteri M., Haardt F., Madau P., 2003, ApJ, 582, 559
- Wyithe J. S. B., Loeb A., 2003, ApJ, 595, 614

APPENDIX A: MODEL SELECTION CRITERIA

The F-test. If the uncertainties of the measurements are known to a good precision, model selection can be performed using a simple χ^2 test. However, we cannot prove that our bootstrap errors accurately reproduce the true scatter in the data. An alternate test that does not require knowledge of the true standard deviation (up to a constant scaling factor) is to form the ratio

$$\mathcal{F} = \frac{\chi_{n-k}^2 - \chi_n^2}{k} \frac{N - n}{\chi_n^2}, \quad (\text{A1})$$

where N is the number of independent data values used in the fit (in our case $N = N_{\text{pca}}$). The numerator in this equation is the difference between the χ^2 calculated for n parameters and that for $n - k$ parameters. Being a ratio of two χ^2 -distributed variables, \mathcal{F} follows a F -distribution with $(k, N - n)$ degrees of freedom. We can then estimate how significant is the addition of k parameters by integrating this distribution from 0 to \mathcal{F} .

The Akaike information criterion. The question to find which model best approximates a given data set can be addressed in terms of information theory: the best model minimizes the loss of information. The AIC evaluates models using the Kullback–Leibler information (Akaike 1973). In terms of the number of independent

data points, n , and of model parameters, p ,⁹

$$\text{AIC} = \chi_{\min}^2 + 2p + \frac{2p(p+1)}{n-p-1}, \quad (\text{A2})$$

where the second-order term on the right-hand side is needed when the size of the data set does not exceed the number of model parameters by a large factor ($\gtrsim 40$). The model with the minimum AIC has to be considered the best. Note that the AIC penalizes for the addition of parameters and thus selects a model that fits well but has a minimum number of parameters. The relative probability that a model is the correct solution is given by the Akaike weights

$$w_i = \frac{e^{-\Delta \text{AIC}_i/2}}{\sum_{k=1}^N e^{-\Delta \text{AIC}_k/2}}, \quad (\text{A3})$$

where the indices i and k run over the different models and $\Delta \text{AIC} = \text{AIC}_i - \min_k(\text{AIC}_k)$. In general, models receiving AIC within 2 of the ‘best’ deserve consideration, those within 3–7 have considerably less support and those above 7 are basically rejected by the test (Burnham & Anderson 1998).

The Bayes factor. BFs are the dominant method for Bayesian model testing (see Kass & Raftery 1994, for an extensive review). The BF, B_{12} , is the ratio between the marginalized likelihoods (i.e. between the probabilities of the data given the models) for two different models (1 and 2) and provides a scale of evidence in favour of a model versus another. The model with the highest marginalized likelihood is the best. Following an early suggestion by Jeffreys (1961), Kass & Raftery (1994) proposed the following ‘rule of thumb’ for interpreting the BFs: $1 < B_{12} < 3$ provides weak evidence (barely worth mentioning) for Model 1, $3 < B_{12} < 12$ provides substantial evidence for Model 1, $12 < B_{12} < 150$ provides strong evidence for Model 1, $B_{12} > 150$ provides decisive evidence for Model 1. Several numerical approaches have been proposed to compute BFs based on MCMC sampling but most of these methods are subject to numerical or stability problems (see e.g. Han & Carling 2000, for a review). As a compromise between accuracy and coding complexity, we estimate the BFs using the harmonic mean estimator (see e.g. Kass & Raftery 1994). We repeat the calculation using three different chains and we use the standard deviation among them as an estimate of the uncertainty of our marginalized likelihoods.

The Bayesian information criterion. Unfortunately, while BFs are rather intuitive, as a matter of fact they are often quite difficult to calculate. This makes simpler (but approximate) estimates of the BFs of great interest. Schwarz (1978) derived the BIC as a large sample approximation to twice the logarithm of the BF.

$$\text{BIC}_1 - \text{BIC}_2 \simeq -2 \ln(B_{12}), \quad (\text{A4})$$

with

$$\text{BIC} = \chi_{\min}^2 + p \ln n. \quad (\text{A5})$$

Similarly, to the AIC, the preferred model is the one with the lowest value of the criterion. The Kass–Raftery criterion for model selection is also applied to the BIC.

The Deviance Information Criterion. Spiegelhalter et al. (2002) have recently proposed a Bayesian generalization of the AIC: the DIC. It is based on the posterior distribution of the log-likelihood

⁹ In our case, p is the number of model parameters plus one to account for the estimation of the χ^2 function.

or deviance ($D = -2 \log \mathcal{L}$ which, in our case, coincides with χ^2 function):

$$\text{DIC} = \langle D \rangle + p_{\text{D}}, \quad (\text{A6})$$

where $\langle D \rangle$ is the posterior expectation of the deviance while the effective number of parameters, p_{D} , is defined as the difference between the posterior mean of the deviance and the deviance evaluated at the posterior mean of the parameters. $p_{\text{D}} = \langle D \rangle - D(\langle \theta \rangle)$. The DIC can be rewritten as

$$\text{DIC} = D(\langle \theta \rangle) + 2p_{\text{D}} \quad (\text{A7})$$

which makes the analogy with the first-order AIC explicit. A good model corresponds to a low DIC and model-selection criteria developed for the AIC appear to work well also for the DIC (Spiegelhalter et al. 2002). Note, however, that, since $p_{\text{D}} \leq p$, the DIC tends to be less conservative than the AIC in terms of model complexity. The main attraction of using this measure is that it is trivial to compute when performing MCMC on the models.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.