Systematic errors in weak lensing: application to SDSS galaxy–galaxy weak lensing

Rachel Mandelbaum,1∗ Christopher M. Hirata,1 Uroš Seljak,1,2 Jacek Guzik,3,4 Nikhil Padmanabhan,1 Cullen Blake,5 Michael R. Blanton,6 Robert Lupton7 and Jonathan Brinkmann8

1Department of Physics, Jadwin Hall, Princeton University, Princeton, NJ 08544, USA
2International Centre for Theoretical Physics, Strada Costiera 11, 34014 Trieste, Italy
3Astronomical Observatory, Jagiellonian University, Orla 171, 30-244 Kraków, Poland
4Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA
5Harvard-Smithsonian Centre for Astrophysics, MS-10, 60 Garden Street, Cambridge, MA 02138, USA
6Centre for Cosmology and Particle Physics, Department of Physics, New York University, 4 Washington Pl, New York, NY 10003, USA
7Princeton University Observatory, Princeton, NJ 08544, USA
8Apache Point Observatory, 2001 Apache Point Road, Sunspot, NM 88349-0059, USA

Accepted 2005 June 2. Received 2005 April 17; in original form 2005 January 12

ABSTRACT
Weak lensing is emerging as a powerful observational tool to constrain cosmological models, but is at present limited by an incomplete understanding of many sources of systematic error. Many of these errors are multiplicative and depend on the population of background galaxies. We show how the commonly cited geometric test, which is rather insensitive to cosmology, can be used as a ratio test of systematics in the lensing signal at the 1 per cent level. We apply this test to the galaxy–galaxy lensing analysis of the Sloan Digital Sky Survey (SDSS), which at present is the sample with the highest weak lensing signal-to-noise ratio and has the additional advantage of spectroscopic redshifts for lenses. This allows one to perform meaningful geometric tests of systematics for different subsamples of galaxies at different mean redshifts, such as brighter galaxies, fainter galaxies and high-redshift luminous red galaxies, both with and without photometric redshift estimates. We use overlapping objects between SDSS and the DEEP2 and 2df-Sloan LRG and Quasar (2SLAQ) spectroscopic surveys to establish accurate calibration of photometric redshifts and to determine the redshift distributions for SDSS. We use these redshift results to compute the projected surface density contrast ΔΣ around 259 609 spectroscopic galaxies in the SDSS; by measuring ΔΣ with different source samples we establish consistency of the results at the 10 per cent level (1σ). We also use the ratio test to constrain shear calibration biases and other systematics in the SDSS survey data to determine the overall galaxy–galaxy weak lensing signal calibration uncertainty. We find no evidence of any inconsistency among many subsamples of the data.

Key words: gravitational lensing – galaxies: distances and redshifts – galaxies: halos.

1 INTRODUCTION
Recent years have seen tremendous progress in the detection of galaxy–galaxy weak lensing (Brainerd, Blandford & Smail 1996; Hudson et al. 1998; Fischer et al. 2000; McKay et al. 2001; Smith et al. 2001; Guzik & Seljak 2002; Hoekstra et al. 2003; Sheldon et al. 2004; Hoekstra, Yee & Gladders 2004; Seljak et al. 2005), the tangential shear distortion around galaxies due to their dark matter halos. Recent measurements of galaxy–galaxy weak lensing (Sheldon et al. 2004; Hoekstra et al. 2004; Seljak et al. 2005) demonstrate 20–30σ detections. In light of the increasing statistical precision with which this effect is measured, it is important to revisit common sources of systematic error, which are currently at the 10 per cent level, and therefore already dwarf the statistical error.

While galaxy–galaxy weak lensing is potentially a very powerful tool for studying the dark matter halo profiles around stacked foreground galaxies, it suffers from a large number of potential calibration biases. Because it involves measuring the projected surface...
density contrast,\\n(1)\\n\[ \Delta \Sigma(r) = \gamma_s(r) \Sigma_c, \]

averaged over stacked lens and source galaxies (where \( r \) is the transverse separation from the lens galaxy), calibration biases may be introduced via both the tangential shear and the redshifts used to compute \( \Sigma_c \).

Here we introduce some of the notations that are used in this paper to describe the weak-lensing signal. We compute the two-dimensional projected surface density contrast of stacked foreground galaxies as a function of transverse separation from those galaxies, where this separation is measured in comoving coordinates, using \( r = r_p (1 + z_l) \) (with subscript ‘l’ referring to the lens, ‘s’ to the source) and \( r_p = \theta_b D_A(z_l) \), the product of the observed angular lens–source separation on the sky and the angular diameter distance at the lens redshift. We can then measure the surface density contrast, which is related to the projected mass density \( \Sigma_\nu(r) \) and its average value inside radius \( r \), \( \Sigma(<r) \), as follows:

\[ \Delta \Sigma(r) = \Sigma(<r) - \Sigma(r) = \gamma_s(r) \Sigma_c. \]

The inverse critical surface density \( \Sigma_\nu^{-1} \) in comoving coordinates is defined by

\[ \Sigma_\nu^{-1} = \frac{4 \pi G}{c^2} \frac{D_b D_A (1 + z_l)^2}{D_A} \]

in terms of angular diameter distances, and has the following properties: for a given lens redshift, it is zero for \( z_s < z_l \), then increases rapidly above \( z_s > z_l \) until it flattens out as \( z_s \gg z_l \); the asymptotic value for \( z_s \gg z_l \) increases with \( z_l \).

For a full discussion of the errors that can be introduced in the shear computation, see Hirata et al. (2004, hereinafter H04). As shown there, several types of error can be introduced when computing the shear, including biases due to point spread function (PSF) correction, noise rectification bias and selection biases. These biases are considered in detail for the linear PSF correction method used in that paper. In this paper, we use a different PSF correction scheme, ‘re-Gaussianization’, which was introduced and tested in Hirata & Seljak (2003), and include a discussion of the effects of that choice on systematic error in the shear.

Because systematic errors in the shear computation have been well studied, and statistical errors can be decreased to a fairly low level when data from large surveys such as the SDSS are used, systematic errors in the redshift distribution (and consequently, in \( \Sigma_\nu \)) are of much greater importance than they were previously. For the purposes of this paper, we assume that lens redshifts are known to high precision via spectroscopy, so our concern is the source redshift distribution; as shown in Kleinheinrich et al. (2004), if the lens redshifts are also unknown, g-g weak lensing is not nearly as powerful a tool. Several common methods of determining the source redshift distribution are inadequate for precision cosmology due to previously unquantified systematic uncertainties that they introduce via their effects on \( \Sigma_\nu^{-1} \); this paper includes a study of the biases that may be introduced using these methods, and a comparison against a well-determined reference distribution.

In addition to biases introduced due to the redshift distribution, we also discuss systematics that can be introduced while computing the signal. These include biases due to intrinsic alignments, selection effects and several other effects. One new effect not noted before is a problem with the determination of the sky flux near bright objects by the SDSS PHOTO pipeline that leads to a problem in detection of sources within about 90 arcsec of bright objects (Section 6.3.7). This problem affects any galaxy–galaxy or cluster–galaxy weak lensing analysis using SDSS data.

Many of the sources of systematic errors discussed above are common to both weak lensing autocorrelation analysis and galaxy–weak lensing correlations (galaxy–galaxy lensing). Weak lensing autocorrelation analysis at present is limited by the statistical precision (see summaries in Refregier 2003b and Hoekstra 2003), and it is difficult to test for the presence of systematics within each data set. However, the size of these data sets is rapidly increasing, and in the near future systematic errors are likely to dominate the statistical errors. Understanding of weak lensing systematics is essential if one is to exploit the full potential of upcoming and planned surveys such as the Canada–France–Hawaii Telescope (CFHT) legacy survey (http://www.cfht.hawaii.edu/Science/CFHTLS, Mellier 2001), Pan-Starrs (http://pan-starrs.ifa.hawaii.edu/, Kaiser et al. 2002; Kaiser 2004), LSST (http://www.lsst.org/lsst_home.html, Tyson 2002) and SNAP (http://snap.lbl.gov/, Rhodes et al. 2004; Massey et al. 2004; Refregier et al. 2004), some of which may reach statistical precision at the 0.1 per cent level.

Many of our systematic tests are done using the following method. Several authors (Jain & Taylor 2003; Bernstein & Jain 2004) have proposed geometric tests of dark energy using the fact that \( \Delta \Sigma \) is an invariant of the projected lens mass distribution, and therefore must be the same when measured with two different source samples at different redshifts. The use of this fact to test the dark energy density and equation of state requires control of systematics at the 0.1 per cent level. Since systematics in weak lensing are currently only constrained at the 10 per cent level, we turn this test around to use \( \Delta \Sigma \) measured with reference samples to check for systematic error, knowing that cosmology plays a negligible role in the comparison. If the source samples being compared vary in quantities affecting both the shear computation and the redshift distribution (e.g. if one sample is more distant, with lower SNR shape measurement and less well-known redshift distribution), then the systematics in both quantities may be different as well, and we can only use this method to test the overall calibration of the signal rather than the shear calibration, redshift distribution and other effects separately. Fortunately, the SDSS now covers a large enough area that there is significant statistical power for such tests.

In Section 2, we describe the data acquisition, selection criteria and processing. The common redshift distributions used for weak lensing analyses are discussed in Section 3, including specifically how these methods are implemented in this paper. Additional systematics issues introduced in the computation of the lensing signal are described in Section 4. Our implementation of the test for systematics is described in Section 5, and the results are given in Section 6. Finally, the implications of these tests are discussed in Section 7.

Here we note the cosmological model and units used in this paper. All computations assume a flat \( \Lambda \)CDM universe with \( \Omega_m = 0.3 \) and \( \Omega_\Lambda = 0.7 \). Distances quoted for transverse lens–source separation are comoving (rather than physical) \( h^{-1} \) kpc, where \( H_0 = 100 \) \( h \) \( \mathrm{km \ s}^{-1} \) \( \mathrm{Mpc}^{-1} \). Likewise, \( \Delta \Sigma \) is computed using the expression for \( \Sigma_\nu^{-1} \) in comoving coordinates (equation 3). In the units used, \( H_0 \) scales out of everything, so our results are independent of this quantity. All confidence intervals in the text and tables are 95 per cent (2\( \sigma \)) confidence level (CL) unless explicitly noted otherwise.

2 TECHNICAL APPARATUS

In this section, we describe the data used for our computation of the lensing signal. The source of this data is the Sloan Digital Sky Survey (SDSS), an ongoing survey that will eventually image...
approximately one quarter of the sky (10 000 deg$^2$). Imaging data are taken in drift-scan mode in five filters, $ugriz$, centred at 355, 469, 617, 748 and 893 nm (Fukugita et al. 1996) using a wide-field CCD (Gunn et al. 1998). After the computation of an astrometric solution (Pier et al. 2003), the imaging data are processed by a sequence of pipelines, collectively called PHOTO, that estimate the PSF and sky brightness, identify objects and measure their properties. The software pipeline and photometric quality assessment is described in Ivezci et al. (2004). Bright galaxies and other interesting objects are selected for spectroscopy according to specific criteria (Eisenstein et al. 2001; Richards et al. 2002; Strauss et al. 2002). The SDSS has four major data releases: the Early Data Release (EDR) (Stoughton et al. 2002), DR1 (Abazajian et al. 2003), DR2 (Abazajian et al. 2004) and DR3 (Abazajian et al. 2005). While we use imaging data more up-to-date than DR3, we are limited in area by the spectroscopic coverage available to us because spectroscopy lags significantly behind photometry.

2.1 Lens catalogue

The lens (foreground) galaxies used for this study are included in the SDSS main galaxy spectroscopic sample (Strauss et al. 2002), as part of the NYU Value-Added Galaxy catalogue (VAGC, Blanton et al. 2004), though the version of the VAGC used here includes more area than the public one described in Blanton et al. (2004). The VAGC is used because of its consistent overall calibration (Schlegel et al., in preparation). The sample, after redshift and magnitude cuts to be described below, includes 259 609 galaxies (decreased from 314 906 after exclusion of the southern Galactic region). We only use lenses at redshift $z > 0.02$ because of the computational expense of computing pairs out to $2 \, h^{-1}$ Mpc for lenses at lower redshifts, and because the lower redshift galaxies have low $\Sigma_c^{-1}$, and therefore contribute little weight. Furthermore, for this study, we only use galaxies with $r$-band Petrosian absolute magnitude $-23 \leq M_r \leq -17$, divided into six magnitude bins, each 1-mag wide. The spectra used were processed by a separate pipeline at Princeton (Schlegel et al., in preparation). The fluxes were extinction corrected using dust maps from Schlegel, Finkbeiner & Davis (1998), then $k$-corrected to $z = 0.1$ using KCORRECT v1.11 with values given directly in the VAGC.

This sample is approximately flux limited to Petrosian apparent magnitude $r = 17.77$; all absolute magnitudes for the lens sample used in this paper are Petrosian $r$-band magnitudes. Redshift evolution of luminosity consistent with Blanton et al. (2003a) was used, so that the absolute magnitude used for all cuts was

$$M_r \text{(used)} = M_r \text{(measured)} + 1.6 \, (z - 0.1). \tag{4}$$

The effect of this shift is to include higher luminosity lenses with $M_r \text{(measured)} < -23$ in the brightest luminosity bin because their $(z)$ is greater than 0.1, and to include fainter lenses in the faintest bin, because their $(z)$ is approximately 0.03.

Fig. 1 shows the lens redshift distribution. The $M_r$ limits, mean redshifts and widths of the distribution for the six luminosity bins used here are shown in Table 1, as is the mean effective redshift (taking into account the weights used for the computation of signal) and mean effective relative to $L_*$ (with $M_* = -20.44$ as in Blanton et al. 2003a) since they are more relevant for the lensing signal. Note that $z_{\text{eff}} < (z)$ because the larger number of pairs for lower redshift lenses (due to the larger angular scale associated with the fixed transverse comoving scale) overcomes the fact that $\Sigma_c^{-1}$ for a given source redshift increases with lens redshift.

Figure 1. The redshift distribution for lens galaxies, shown overall and for the six luminosity bins. As shown, brighter samples peak at higher redshifts. The cutoff at $z = 0.02$ was imposed artistically on the sample. Solid dots are used to indicate the weighted mean redshift of each lens sample.

### Table 1

| Sample, $M_r$ range | $N_{\text{gal}}$ | $\langle z \rangle$ | $\sigma (z)$ | $z_{\text{eff}}$ | $L_{\text{eff}}/L_*$ |
|---------------------|------------------|---------------------|-------------|-----------------|------------------|
| L1, $[-18, -17]$    | 6 524            | 0.032               | 0.011       | 0.032           | 0.080            |
| L2, $[-19, -18]$    | 19 192           | 0.048               | 0.015       | 0.047           | 0.20             |
| L3, $[-20, -19]$    | 58 848           | 0.074               | 0.021       | 0.071           | 0.49             |
| L4, $[-21, -20]$    | 104 752          | 0.11                | 0.03        | 0.10            | 1.2              |
| L5, $[-22, -21]$    | 63 794           | 0.16                | 0.05        | 0.14            | 2.5              |
| L6, $[-23, -22]$    | 6 499            | 0.22                | 0.06        | 0.19            | 5.6              |

One important fact about the lens samples L1–L6 is that they are flux limited, not volume limited. As a result, since $\Delta \Sigma$ is averaged over lens–source pairs with a weight proportional to $\Sigma_c^{-2} \, (z_l, z_s)$, the mean effective redshift and luminosity may vary depending on the redshift distribution of the source sample. When comparing $\Delta \Sigma$ for a given lens sample but different source samples, we explicitly computed the mean effective redshift $z_{\text{eff}}$ and luminosity $L_{\text{eff}}$ of each lens sample with each source sample (representative values are given in Table 1). Variations in $1 + z_{\text{eff}}$ and $L_{\text{eff}}$ for the same lens sample but different source sample were found to be quite small, a maximum of 2 per cent; because this variation is so small, and because we lose significantly in statistics by going to a volume-limited sample, we choose to keep the full flux-limited sample and, when necessary, apply corrections to the computed $\Delta \Sigma$ when comparing between different source subsamples. Corrections will be described further in Section 4.9.

2.2 Source catalogue

2.2.1 Constructing the catalogue

The source sample consists of galaxies selected from the SDSS photometric catalogue (York et al. 2000; Hogg et al. 2001; Smith...
et al. 2002; Stoughton et al. 2002; Abazajian et al. 2003; Pier et al. 2003). The catalogue contains information about the images from the SDSS camera (Gunn et al. 1998) processed at Princeton by the PHOTO software (Lupton et al. 2001; Finkbeiner et al. 2004), rerun 137. Note that for the source catalogue, we use the model magnitudes in all five bands rather than Petrosian magnitudes, because of their higher signal-to-noise ratio for fainter galaxies.

The source catalogue used for this work is not the same as that used in H04. Here we describe the catalogue used, emphasizing differences from the previous one. The most trivial difference is the size of the data set; here we use imaging data acquired from 1998 September 19 (run 94) through 2004 June 15 (run 4682), whereas the H04 catalogue did not include imaging data acquired after 2003 March 10 (run 3712); but there are also significant changes in the pipeline. There are several steps involved in the development of the catalogue: (i) object selection based on PHOTO outputs, (ii) shape measurement and cuts on the shape measurement, (iii) other cuts and (4) organization.

We begin by describing basic object selection starting from the PHOTO outputs. Star/galaxy separation was accomplished using the PHOTO pipeline output OBIC_TYPE, and the cut on resolution factor described below should further reduce stellar contamination. We defer discussion of possible stellar contamination to Section 4.3. Unlike the catalogue used for H04, this catalogue includes deblended child galaxies. Because the deblender has been significantly improved for DR2, the phenomenon noted for EDR and DR1 that the deblender sometimes ‘shreds’ large galaxies rarely occurs (according to Abazajian et al. 2004, inspection of several hundred deblends indicates that they are correct roughly 95 per cent of the time). While shape measurement was performed for all galaxies brighter than magnitude 22 in r-band and 21.6 in i-band (no requirements on detection in g-band), these cuts were applied using model magnitudes before the extinction correction. Several other cuts on the PHOTO flags were performed: the galaxy must have been detected in unbinned images in the r- and i-bands; also, several flags indicating problems in shape measurement or problems with the image (e.g., interpolated pixels) must not have been set.

Next, we describe the shape measurement determination. The PSF correction algorithm used for this work was the ‘re-Gaussianization’ scheme described and tested in Hirata & Seljak (2003). Recent SDSS lensing works, including Sheldon et al. (2004) and H04, have used the linear scheme described there, but as shown in Hirata & Seljak (2003), re-Gaussianization is much more successful at avoiding various shear calibration problems, reducing them to the several per cent level (rather than ~10 per cent) even for poorly resolved galaxies. Unlike the linear scheme, which involves correcting the measured adaptive moments of the image by factors involving the adaptive moments of the PSF, re-Gaussianization involves fitting the PSF shape to a Gaussian, and using the deviations of the PSF from Gaussianity in the PSF correction. The re-Gaussianization method was implemented by reading the atlas images and the PSF maps from PHOTO (Stoughton et al. 2002), since it is impossible to implement using the object catalogues alone.

1 OBIC_TYPE classifies objects as ‘galaxies’ if the flux estimated from the linear combination of de Vaucouleurs and exponential profiles (composite model magnitude, or cmodel magnitude) fit to the object exceeds the flux estimated from the best-fitting PSF by at least 0.24 mag. This works because the profile fit will pick up more of the light from an extended object than the PSF fit. At faint magnitudes (r > 21) this separation scheme mistakes some stars as galaxies; see Section 4.3.

Re-Gaussianization is a perturbative PSF correction scheme based on the observation that if the PSF P and the pre-seeing galaxy image f are Gaussians, and have covariance matrices M_P and M_f, then the observed image I = P \otimes f (here \otimes represents two-dimensional convolution) is a Gaussian of covariance M_P + M_f. A simple PSF correction scheme is thus to find the covariance matrices M_P and M_f of the PSF and M_f of the observed galaxy image, and estimate

\[ M_f = M_f - M_P. \]  

(5)

In practice, galaxy shapes are not perfectly Gaussian, but one can fix this by finding A, \epsilon_f and M_f that minimizes

\[ \int |I(x) - A e^{-i(x-\epsilon_f)M_f^{-1}(x-\epsilon_f)^2}|^2 dx. \]  

(6)

The covariance matrix M_f so obtained is known as the ‘adaptive’ covariance matrix, and its trace T_f is known as the ‘adaptive’ trace. In principle, one can evaluate equation (5), and then estimate the galaxy ellipticity

\[ (\epsilon_{f,x}^i, \epsilon_{f,y}^i) = \left( \frac{M_{f,x}^{i,i} - M_{f,y}^{i,i}}{T_f}, \frac{2M_{f,y}^{i,i}}{T_f} \right). \]  

(7)

In practice, equation (7) does not work very well for real PSFs and galaxies – both PSF and galaxy tend to have sharper central peaks and wider tails than Gaussians with the adaptive covariance – and thus two corrections are made. The first correction, due to Bernstein & Jarvis (2002), accounts for the non-Gaussianity of the galaxy. If the PSF is circular, equation (7) reduces to

\[ \epsilon_f = \frac{\epsilon_f}{R^2_f}. \]  

(8)

where \[ R_f = 1 - T_f/P_t \] is the resolution factor. Bernstein & Jarvis (2002) worked to first order in T/P for a non-Gaussian galaxy, and found that equation (8) still applies, but with \[ R_f \] replaced by the non-Gaussian resolution factor

\[ R_f(NG) = 1 - T_f/P_t + 1/a_{t,i} \]  

(9)

where \[ a_t = R_{t,i}^2 \] is the dimensionless kurtoisis of the galaxy, defined to be zero for a Gaussian (see Bernstein & Jarvis 2002 for a precise definition). This equation can be generalized to an elliptical PSF by requiring SL(2, R) shear invariance, and is found to work well for Gaussian PSFs in ‘toy’ simulations (Hirata & Seljak 2003).

The second correction required is for the non-Gaussianity of the PSF. This correction begins by finding the Gaussian G(x) that best fits the PSF P(x) according to the unweighted least-squares method, i.e. minimizing \[ \int |G - P|^2 dx. \] It then renormalizes G to integrate to unity – a condition not always satisfied by the best-fitting Gaussian even though \[ \int P(x)dx = 1 \] and finds the residual \[ \epsilon(x) = P(x) - G(x). \] Next the pre-seeing image of the galaxy is approximated by a Gaussian f_0(x) whose covariance matrix M_{f_0} is obtained by subtracting the adaptive covariance matrices M_f - M_P, and a ‘re-Gaussianized’ image I' = I - \epsilon \otimes f_0 is constructed, which is supposed to approximate what would have been observed had the PSF been Gaussian. The Bernstein & Jarvis (2002) prescription is then applied to I' using R_{t,i}^2(NG).

The re-Gaussianization scheme is exact to first order in PSF non-Gaussianity, however higher order approaches have been proposed based on expansions of the galaxy and PSF in orthogonal functions (Bernstein & Jarvis 2002; Refregier 2003a). Suggestions include

© 2005 RAS, MNRAS 361, 1287–1322
direct fitting of the convolved galaxy to the data (Bernstein & Jarvis 2002) or deconvolution regularized by a cutoff in the orthogonal function expansion (Refregier & Bacon 2003). These methods have not yet come into general use, but are likely to become more widely used in the future due to the demanding calibration requirements of cosmic shear surveys. However, we note that galaxy–galaxy lensing is a promising ‘testing ground’ for these methods since the same systematics tests that we use to test redshift distributions in Section 5 could also be used to study the relative calibrations of the various PSF correction methods. Such tests could be done independent of redshift distribution information by using the same sets of sources for a comparison of the shear $\gamma_1$ computed from ellipticities determined by each PSF correction method.

Only galaxies passing certain cuts on the shape measurement were included in the catalogue. The shape measurements used were the average of those in the $r$ and $i$ bands; the $u$, $g$, and $z$ bands were not used because their lower signal-to-noise ratio did not justify the large computational expense of performing the re-Gaussianization. To eliminate galaxies that may cause large noise-rectification bias, to avoid the untrustworthy results of PSF correction when the galaxy is unresolved and to help with star–galaxy separation, we only include galaxies with resolution factor $R_2 > 1/3$; $R_2$ here is defined as

$$R_2 = 1 - \frac{T(r)}{T(i)} \tag{10}$$

where $T(r)$ is the trace of the moment matrix for the PSF and $T(i)$ is that for the re-Gaussianized galaxy image. Note that this cut is only applied in the bands used for shape measurement, and since we only attempt shape measurement in $r$ and $i$, it is possible that $R_2 < 1/3$ in the other bands (indeed, the object may not even be visible in $u$, $g$, or $z$). We require that the galaxy pass this cut in both bands, since if we only require it in one band, then we can create a selection bias by preferentially using the shape measurement that has $R_2 > 1/3$ over that with $R_2 < 1/3$.

Once shape measurement was complete, several other types of cuts were applied. To ensure a relatively uniform source sample across the survey area and to avoid regions near the Galactic plane, only regions in which the extinction was lower than 0.2 mag in $r$-band were used. The extinction was determined using dust maps from Schlegel et al. (1998). To ensure a uniform sample, we also require $r < 21.8$ (extinction corrected).

The shape error estimates are in principle used for three purposes: weighting, determination of the shear responsivity and determination of the error bars on final quantities such as $\Delta \Sigma$. We obtain the errors due to Poisson fluctuations in the sky and CCD dark current via the simple formula (appropriate for Gaussians; Bernstein & Jarvis 2002)

$$\sigma_{\text{sky + dark}} = \frac{\sigma^{(i)}}{R_2 F} \sqrt{4 \pi n_i}, \tag{11}$$

where $n_i$ is the sky and dark current brightness in photons per pixel, $\sigma^{(i)} = \det \mathbf{M}^{(i)}$ is the size of the galaxy in pixels and $F$ is the flux. This equation is crude, but for the purposes of weighting there is no need for high accuracy, and the error bars on our results are computed via analytic, random catalogue and bootstrap methods that depend on the actual dispersion of the ellipticities, including shape noise, rather than $\sigma_\varepsilon$ itself. (The responsivity determination is addressed in Section 2.2.2.) There is also a contribution to the shape measurement uncertainty due to Poisson noise from the galaxy itself; this is given by (again for Gaussians)

$$\sigma_{\text{gal}} = R_2^{-1} \sqrt{\frac{64}{27N_e}}, \tag{12}$$

where $N_e$ is the number of photoelectrons from the galaxy. This is only significant for galaxies bright compared to the sky; the typical sky brightness is $21$ mag arcsec$^{-2}$ in $r$ band and $20.3$ mag arcsec$^{-2}$ in $i$, and even a poorly resolved ($R_2 = 1/3$) galaxy usually has a full width at half maximum of $\sim 1.7$ arcsec after seeing, so sky brightness dominates for $r \geq 20$ and $i \geq 19.3$. Since the shape noise is dominant over measurement noise for the brighter objects, we have not included the galaxy noise (equation 12) in our weighting, nor has it been included in the adaptive moment errors from PHOTO.

Some additional cuts were designed to eliminate regions with faulty data. These cuts eliminated less than 1 per cent of the data total. First, the mean ellipticities ($\varepsilon_1$) and ($\varepsilon_2$), and their rms deviations, were computed on a run/camcol basis. Those few run/camcols that had $|\langle \varepsilon_1 \rangle | > 0.05$ (for either ellipticity component, in either band) were excluded from the analysis; visual inspection of several of those runs showed severe PSF anisotropy [despite a reasonable PSF full width at half-maximum (FWHM)] for which our correction scheme was unable to account. Furthermore, based on the distribution in rms ellipticities, those run/camcols with values less than 0.38 or greater than 0.52 were excluded (the mean value was 0.45, higher than the expected shape noise since all galaxies, even those that had significant measurement error, were included). Those with rms ellipticities above the acceptable range typically were imaged in particularly poor seeing, which led to greater noise in the PSF-corrected ellipticities. Within runs that were accepted, galaxies with total ellipticity $\varepsilon^2 = \varepsilon_1^2 + \varepsilon_2^2 > 4$ were rejected. Finally, those galaxies in a small region (\(\sim 4\) deg$^2$) that had faulty astrometry were eliminated.

Once these cuts were applied, a few final steps were necessary to make the catalogue useful. In the case of multiple observations of the same galaxy, only one observation was used, that which was taken in better seeing. The shape measurements in the two bands were combined, weighting by the $(S/N)^2$ of the detection in each band.

Unlike in H04, photometric redshifts were assigned to all extinction-corrected $r < 21$ galaxies using a template-based program kphoto v3.2 (Blanton et al. 2003c). The performance of these photometric redshifts will be described in Section 6.1. Approximately 4 per cent of galaxies with $r < 21$ had failed photometric redshift determination, and so were not used for any analysis requiring the use of photometric redshifts.

For reference, we include here some plots showing information about the source sample (these plots also show information about the high-redshift Luminous Red Galaxy, or LRG sample, a subsample of the source catalogue, that will be described in more detail in Section 3.4). The magnitude distribution of sources in the catalogue is shown in Fig. 2. The rms ellipticity as a function of magnitude is shown in Fig. 3. The plot of the average $R_2$ as a function of magnitude, and of the overall distribution of $R_2$ values, is in Fig. 4. Some general information about the catalogue is included in Table 2. Note that for all tests performed in this paper, we divide the sources into three samples: $r < 21$, $r > 21$ and high-redshift LRGs.

### 2.2.2 Shear calibration bias

Here, we list the sources of shear calibration bias that were described in detail in H04, and estimate their magnitudes for the source catalogue used here; refer to that paper for more detail about estimated shear calibration uncertainty.

There are five major sources of shear calibration bias, as listed in H04. First, we consider the PSF dilution correction, the correction to the measured galaxy image to account for the blurring due to convolution with the PSF. Unlike the linear PSF correction method...
Figure 2. The $r$-model magnitude distribution for the source catalogue, shown for the full catalogue (solid) and high-redshift LRGs (dotted line). The turnover at faint magnitudes occurs because we lose many sources at the faint end due to our selection criteria for the shape measurement (e.g. the cut on $R_2$).

Figure 3. The rms ellipticity (for a single ellipticity component, averaged over both bands) as a function of magnitude, shown for all sources, and for the high-redshift LRG sample. We show the results both with and without measurement noise (including noise from the galaxy itself), as labelled. The increase in the result with noise at faint magnitudes shows the effects of noise in the shape measurement, but it seems that the increase in the result without noise at faint magnitudes may indicate a real trend in the rms ellipticities of the galaxies with magnitude.

Figure 4. The top plot shows the average $R_2$ value for the source catalogue as a function of magnitude for all sources (solid) and high-redshift LRGs (dashed). As expected, $\langle R_2 \rangle$ is higher for brighter galaxies, indicating that their shapes are better resolved. The bottom plot shows the distribution of $R_2$ values. The plot is shown for the $r$-band $R_2$; results are nearly identical for the $i$-band $R_2$.

Table 2. Source catalogue properties. Note that only about 65 per cent of the sources are in regions with spectroscopic coverage, and hence only these are used in computing $\Delta \Sigma$. Also, because we require a successful measurement in both bands, only the number measured in both $r$ and $i$ is relevant.

| Sky coverage, $f_{\text{sky}}$ | 0.176 |
| Successful measurements | | |
| (all galaxies) | $r$ or $i$ | 39 436 326 |
| $r$ band | 35 789 302 |
| $i$ band | 35 344 893 |
| $r$ and $i$ | 31 697 869 |
| Successful measurements | | |
| ($r$ and $i$) | $r < 21$ | 18 709 472 |
| $21 \leq r < 21.8$ | 12 988 397 |
| Source density | | |
| (all galaxies) | $r$ or $i$ | 1.51 arcmin$^{-2}$ |
| $r$ and $i$ | 1.21 arcmin$^{-2}$ |
| Resolution factor $\langle R_2 \rangle$ | | |
| (mean ± std. deviation) | $r$ band | 0.61 ± 0.15 |
| $i$ band | 0.60 ± 0.16 |
| Mean magnitude | | |
| (extinction corrected) | $r$ | 20.68 |
| $i$ | 20.22 |
| High-redshift LRGs, $r$ and $i$ | | |
| successful measurements | 2 884 242 |
| source density | 0.11 arcmin$^{-2}$ |
| mean redshift | 0.55 |
| mean resolution factor | $\langle R_2(r) \rangle$ | 0.58 |
| $\langle R_2(i) \rangle$ | 0.55 |
| mean magnitude | $\langle r \rangle$ | 20.93 |
| $\langle i \rangle$ | 20.03 |

used in H04, which has significant shear calibration uncertainty due to this effect for the less well-resolved galaxies, the re-Gaussianization method only has a few per cent shear calibration uncertainty (Hirata & Seljak 2003) even for the lower limit.
\( R_2 = 1/3 \) studied in that paper. For this paper, a plot of the PSF dilution correction as a function of \( R_2 \), for both exponential and de Vaucouleurs profiles, for various values of source ellipticity, is shown in Fig. 5.

We model the calibration error due to the PSF dilution correction as due to a sum of contributions from exponential profile and de Vaucouleurs profile galaxies,

\[
\frac{\delta \gamma}{\gamma} = f_{\exp} \left( \frac{\delta \gamma}{\gamma} \right)_{\exp} + f_{\text{deV}} \left( \frac{\delta \gamma}{\gamma} \right)_{\text{deV}},
\]

where \( f_{\exp} \) and \( f_{\text{deV}} \) are the weighted fractions of the two types of galaxies with \( f_{\exp} + f_{\text{deV}} = 1 \). A lower bound on \( \delta \gamma/\gamma \) can be obtained from Fig. 5 by noting that \( \delta \gamma/\gamma \geq -0.014 \) for the exponential profile and \( \delta \gamma/\gamma \geq -0.035 \) for the de Vaucouleurs profile, hence

\[
\frac{\delta \gamma}{\gamma} \geq -0.014 f_{\exp} - 0.035 f_{\text{deV}}.
\]

For the upper bound, we repeat this calculation, except that we use the \( \delta \gamma/\gamma \) corresponding to \( R_2 = 1/3 \) and ellipticity equal to \( \sqrt{2} \sigma_{\text{rms}} \) (since this is the rms total ellipticity). This calculation is conservative since most galaxies are at \( R_2 > 1/3 \) where \( \delta \gamma/\gamma \) is less (at fixed \( e \)). (We find that \( R_2 \) and \( e^2 \) are almost completely uncorrelated after the noise \( \sigma_e^2 \) is subtracted from \( e^2 \) for each of our three samples, and for both the de Vaucouleurs and exponential sources within each of these samples.) The results are shown in Table 3; these should be interpreted as \( 2\sigma \) bounds, since it is likely that some cancellation between positive and negative dilution occurs.

Next, we consider errors in PSF reconstruction, which can arise if the PSF ellipticity or trace is misestimated by the Photo PSF pipeline. This error was considered by H04, which showed that the PSF trace is accurately reconstructed to within \( |\delta T_{\text{psf}}/T_{\text{psf}}| < 0.03 \). We can use equation (20) of H04 to estimate the resulting calibration error; we find weighted averages \( \langle R_{21}^{-1} \rangle = 1.71, 1.81 \) and 1.90 for the \( r < 21 \), \( r > 21 \) and LRG samples\(^2\), respectively, resulting in calibration errors of \( \pm 2.1 \) (\( r < 21 \)), \( \pm 2.4 \) (\( r > 21 \)) and \( \pm 2.5 \) per cent (LRG).

\(^2\) The \( \langle R_{21}^{-1} \rangle \) values are the same for the \( r \) and \( i \) bands to within \( \pm 0.01 \) for the \( r < 21 \) and \( r > 21 \) samples. For the LRG sample, we find \( \langle R_{21}^{-1} \rangle = 1.83 \) in \( r \) and 1.90 in \( i \); we quote the \( i \) value here because the signal-to-noise ratio for the LRGs is typically greater in \( i \) and hence this measurement is weighted more heavily. This is certainly conservative as the error increases with increasing \( \langle R_{21}^{-1} \rangle \).

Table 3. Shear calibration biases and other parameters for the various source samples, at the 2\( \sigma \) level.

| Sample | \( r < 21 \) | \( r > 21 \) | LRG |
|--------|-------------|-------------|-----|
| \( f_{\exp} \) | 0.59 | 0.58 | 0.33 |
| \( \sigma_{\text{rms}}(\exp) \) | 0.39 | 0.41 | 0.41 |
| \( \sigma_{\text{rms}}(\text{deV}) \) | 0.38 | 0.42 | 0.37 |

Calibration bias (per cent)

- PSF dilution: \([-2.2, +2.9] \] \([-2.2, +4.0] \] \([-2.8, +3.9] \]
- PSF reconstruction: \( \pm 2.1 \) (\( \pm 2.4 \)) \( \pm 2.5 \)
- Selection bias: \( [0, 5.7] \) \( [0, 10.3] \) \( [0, 11.1] \)
- Shear responsivity error: \( [0, 1.7] \) \( [0, 1.7] \) \( [0, 1.7] \)
- Noise rectification: \( [-1, 0] \) \( [-3.8, 0] \) \( [-1.2, 0] \)
- Total \( 2\sigma \) \( \delta \gamma/\gamma \) (per cent): \( [-5, +12] \) \( [-8, +18] \) \( [-6, +19] \)

We also must be concerned about shear selection bias, the preferential selection of galaxies at low or high ellipticity. Considering that Fig. 3 shows clear evidence for evolution of \( \sigma_{\text{rms}} \) with magnitude, we do not estimate selection bias using the method from Section 3.2.3 of H04, which assumes no evolution of rms ellipticity with magnitude. An alternative method of estimating the shear selection bias utilizes a simple model of the selection criteria. This method is not dependent on assumptions about the evolution of the ellipticity distributions. The main selection criteria that can be influenced by shear is the resolution factor cut, which favours highly elongated galaxies, since these have a larger trace \( T^{(i)} \) after PSF convolution than a circular galaxy with the same area. We model this by noting that for a Gaussian galaxy and PSF, the resolution factor obeys

\[
R_2 = 1 - \frac{T^{(P)}}{T^{(P)} + 2\sigma^{(f)^2} \sqrt{1 - e^{(f)^2}}},
\]

where \( \sigma^2 = \sqrt{\text{det}(M^{(f)})} \) and \( e^{(f)} \) is the ellipticity of the pre-seeing galaxy image. A gravitational shear along the \( x \)-axis leaves \( \sigma^{(f)^2} \) fixed but changes \( e^{(f)} \) according to

\[
\Delta e^{(f)} = 2 |1 - e^{(f)^2}| \frac{e^{(f)}}{1 - e^{(f)^2}},
\]

where \( \gamma \) is the amount of the shear. Therefore, the change in resolution factor is

\[
\Delta R_2 = \frac{\partial R_2}{\partial e^{(f)}} \bigg|_{\gamma}, \quad \Delta e^{(f)} = 2 e^{(f)} R_2 (1 - R_2) \gamma \frac{e^{(f)}}{1 - e^{(f)^2}}.
\]
Now the effect of the $R_2 > 1/3$ cutoff on the mean ellipticity can be estimated by averaging the ellipticities of the galaxies that are accepted into the catalogue because of the shear (the integrand is negative when galaxies are removed or their $e^{(f)}_+ < 0$)

$$\Delta(e^{(f)}_+) = \int_0^{1} d\phi^{(f)}_+ \frac{dn}{d\phi^{(f)}_+} \left| \frac{dR_2}{d\phi^{(f)}_+} \right|_{R_2=0.33} \int_0^{\pi} \frac{d\phi^{(f)}_-}{\pi} e^{(f)}_+ \Delta R_2 \approx 2R_{2,\text{min}}(1-R_{2,\text{min}}) \sigma_{\text{rms}}^2 n(R_{2,\text{min}}),$$

where $\phi^{(f)}_+$ is the position angle\(^3\) and $\frac{dn}{d\phi^{(f)}_+} dR_2$ is the joint ellipticity-resolution factor distribution. In the last line we have approximated $e^{(f)}_+$ and $R_2$ as independent, which we have found to be very nearly true, and noted that the mean value of $e^{(f)}_+^2$ is $2\sigma_{\text{rms}}^2$ because the ellipticity has two components. The shear calibration error due to selection at the $R_2$ cut, assuming that all galaxies are weighted equally, is

$$\frac{\delta y}{y} = \frac{R_{2,\text{min}}(1-R_{2,\text{min}})}{\mathcal{R}} \sigma_{\text{rms}}^2 n(R_{2,\text{min}}),$$

where $\mathcal{R}$ is the shear responsivity, which is discussed further in Subsection 2.3. Note that $n(R_{2,\text{min}})$ cannot be estimated from Fig. 4 because that plot shows the distribution of $R_2$ values in the $r$-band. For this calculation, the relevant quantity is the distribution of $R_2$ values formed by choosing (for each source) the lower of the two $R_2$ values, since that number is what determines whether the object is included in our catalogue. Furthermore, we must use the distribution of $R_2$ values weighted by the weights used in our lensing analysis. For a typical value of $\mathcal{R}$ for each sample from Section 2.3, our cut value $R_{2,\text{min}} = 1/3$, and weighted values $n(R_{2,\text{min}}) = 1.6, 2.4$ and 2.8 for the $r < 21$, $r > 21$ and LRG samples, respectively, we obtain a selection bias estimate $dy/y = 0.057, 0.103$ and 0.111 for $r < 21$, $r > 21$ and LRGs.

However, the $R_2$ cut is not the only one that will cause shear selection bias; the detection requirement that $S/N = v > 5$ will lead to selection bias in the opposite direction, though the magnitude of the effect is not as great (see Appendix C, which shows the calculation of the estimate as $dy/y = -0.036, -0.066$ and -0.037 for the three samples, respectively). Consequently, we consider the $2\sigma$ estimate of shear selection bias to be as low as zero and as high as the values estimated only taking into account the $R_2$ selection.

Shear responsivity error is an error in the shear responsivity via a systematic uncertainty in $\sigma_{\text{rms}}$. We use a value of $\sigma_{\text{rms}}$ (mag) estimated from Fig. 3, with different results used for the full source catalogue and for the high-redshift LRG sample. (Note that in light of our shear selection bias results, we may consider that the increase of $\sigma_{\text{rms}}$ with magnitude is in part due to shear selection bias, since the average $R_2$ is lower at fainter magnitudes, and therefore the selection bias should be more severe there. Even if this is true, it is still correct to use the value of $\sigma_{\text{rms}}$ (mag) when computing the shear responsivity.) We estimate the systematic error in $\sigma_{\text{rms}}$ assuming that the uncertainty in $\sigma_{\text{rms}}$ is its primary source of uncertainty. To estimate uncertainty in $\sigma_{\text{rms}}$, we looked at the southern galactic survey area, for which there are many repeat observations of the same area (as many as 27 for some areas) that can be used to get empirical values of $\sigma_{\text{rms}}$ that can be compared against the theoretical value derived from equations (11) and (12). We find that $\sigma_{\text{rms}}$ is overestimated by about 0.01 ($2\sigma$ confidence interval $[0.00, 0.02]$), yielding an estimate of shear calibration bias of $[0, 0.017]$ according to equation (25) in H04.

\(^3\) Defined by $e^{(f)}_+ + i w^{(f)}_+ = e^{(f)}_+ e^{i2\phi^{(f)}_+}$.

The final major source of shear calibration bias is noise-rectification bias, whereby the noise in the image leads to a bias in the ellipticity due to the non-linearity of the PSF correction process. As shown in H04, equations (26) and (27) and Appendix C, the noise-rectification bias can be estimated as

$$\frac{\delta y}{y} \approx K_N v^{-2} = 4(1 - 3R_2^{-1} + R_2^{-2} + 2\epsilon_{\text{rms}}^2) v^{-2},$$

where $v$ is the signal-to-noise ratio of the detection averaged over bands $v^{-2} = \frac{2}{v_i^2 + v_f^2}$.

For high-$R_2$ galaxies, $K_N \approx -2.7$, decreasing to $-3.7$ for $R_2 = 2/3$ and then increasing to 5.3 at $R_{2,\text{min}} = 1/3$ (and rising rapidly at lower $R_2$, as high as 21 for $R_2 = 1/4$). For each source sample, we compute the weighted average value of $K_N v^{-2}$, to find noise-rectification bias of $-0.005 (r < 21), -0.019 (r > 21)$ and $-0.006$ (LRG). To estimate the $2\sigma$ error, we consider the allowed range of the noise-rectification bias to be equal to the magnitude of the error estimated above; results are shown in Table 3.

Those five effects are the major sources of shear calibration bias; there are also several minor sources, at the 0.1 per cent level. These include camera shear (for which we correct using the astrometric solutions, as described in Section 2.2.1), errors due to pixelization and atmospheric refraction effects. We do not attempt to estimate values for these subdominant sources of error. The total shear calibration bias (at the $2\sigma$ level) with the five main sources of error taken into account is shown at the bottom of Table 3 for the three source samples individually. These estimates are conservative in that they do not assume any distribution for these errors, allowing the actual values to add, rather than adding them in quadrature (which assumes some possible cancellation).

### 2.3 Shear estimator

The weighting used for this work differs from that of H04 in two ways. First, rather than the uniform weighting used in that paper, we weight by measurement error

$$w_k = \frac{1}{\sigma_{\text{rms}}^2 + \sigma_e^2},$$

where $\sigma_{\text{rms}}$, the rms shape noise in one ellipticity component, was determined as a function of $r$ model magnitude from Fig. 3 for the full source catalogue and LRGs separately, and $\sigma_e$ is the error per component on the ellipticity from equation (11). This weight is then multiplied by $\Sigma_i e_i^2$, downweighting lower redshift lenses and lens–source pairs with small redshift separation relative to those at large separation. Consequently, the weight used for a given pair is $w_{1,5} = w_i \Sigma_i^{-2}(z_i, z_s)$. The shear responsivity $\mathcal{R}$ appropriate for this weighting scheme is then computed using equations (5-33) and (5-35) from Bernstein & Jarvis (2002), with the average value for our source samples being 0.86 for the $r < 21$ sources, 0.83 for the $r > 21$ sources and 0.85 for LRGs.

Using these weights, the shear estimator is then

$$\Delta \Sigma = \frac{\sum_{1,5} w_{1,5} e_i \Sigma_i^{-1}}{2\mathcal{R} \sum_{1,5} w_{1,5}} = \frac{\sum_{1,5} w_i \Sigma_i^{-1}(c_i, z_s)}{2\mathcal{R} \sum_{1,5} w_{1,5}}.$$
2.4 Error determination

Several methods of determining the errors on $\Delta \Sigma$ were used, each with its own advantages and shortcomings. We describe them here, and in Section 6.2 we compare the results in order to determine on which to rely.

2.4.1 Analytic computation

Analytic expressions for $\Delta \Sigma$ for a given weight function may be derived from equation (5-27) in Bernstein & Jarvis (2002). This method is the least computationally expensive method of deriving errors, but suffers from several shortcomings. First, it gives incorrect results in the presence of spurious shear power in the source catalogue. Secondly, it does not allow an easy way to include errors on the boost factors, which may be significant. Finally, it does not account for correlation of radial bins, which can be significant at large radius, where the average lens–source separation is larger than the average lens–lens separation, so a given source contributes to the measurement in several radial bins.

2.4.2 Random catalogues

A more computationally expensive way of determining the errors is using random lens catalogues. In the absence of systematic shear, the average signal around random points should be zero, and the rms deviation around the mean gives some measure of the noise of the signal, and therefore of the errors when using the real lens catalogue. However, as described in H04, this method is only valid on small distance scales, less than about 500 $h^{-1}$ kpc for the faint lens samples and around 1 $h^{-1}$ Mpc for the brighter samples. Furthermore, this method also cannot take into account the errors on the boost factors, though it can account for the correlation of radial bins. To get a reasonably smooth measure of the errors, a large number of random catalogues must be used; for this work we used 24.

2.4.3 Bootstrap resampling

As described in H04, bootstrap resampling is a useful method of determining the errors. For this purpose, the lens catalogue is divided up into 200 subregions, and the signal is computed for each subregion. The bootstrap-resampled data sets are generated by combining the signal from 200 subregions with replacement. Then, a large number of resamplings (for this work, 2500) can be used to determine the average signal and its error. This method has several advantages over the other two; for example, it naturally incorporates errors in the boost factor, and the correlation of radial bins. However, because of the finite size of the subregions, the errors are once again not trustworthy at large lens–source separation. Also, as described in H04, the noise in the covariance matrix means that the $\chi^2$ values for fits performed using the covariance matrix do not follow a $\chi^2$ distribution.

3 REDSHIFT DISTRIBUTIONS

In this section, we first describe the many commonly used methods of source redshift determination, including potential errors. Then, we describe the ‘reference’ redshift distributions that we use for calibration.

3.1 Photometric redshifts

Surveys that collect photometric information in several passbands allow for the determination of photometric redshifts, which use the galaxy colours to extract an approximate redshift, typically based on the construction of galaxy spectral energy distribution (SED) templates that are evolved with redshift. In principle, these photometric redshifts may be used directly in the computation of $\Sigma_0^{-1}$. To be accurate, this computation should take into account the photometric redshift error distribution, which may be highly non-Gaussian, particularly in certain redshift ranges and areas of colour space.

We studied photometric redshifts computed by two independent groups for the SDSS, and found that they performed statistically nearly identically (i.e. photometric redshifts for particular galaxies were not necessarily the same, but the mean bias and scatter were nearly the same as a function of magnitude and of photometric redshift). One set of photometric redshifts was computed using the program KPHEROZ as mentioned in Section 2.2. The other set was computed by the SDSS photometric redshift working group (as discussed in Csabai et al. 2003, based on work in Csabai et al. 2000 and Budavári et al. 2000), for DR1 data only. Because of the lack of photometric redshifts for the full DR3 sample, instead of using the nearest-neighbour search method from H04 to get photometric redshifts for the full sample, we demonstrate results in this paper using redshifts from KPHEROZ for the full sample.

We only use photometric redshifts for sources at $r < 21$, because of the large scatter at fainter magnitudes that will be demonstrated in Section 6.1.2 The mean redshift of the source sample at $r < 21$ is ~0.35, with a fairly large width. This fact raises several concerns for use with the lens sample, which extends out towards the peak of this redshift distribution at the bright end. Lenses at higher redshift are weighted more highly because of their higher $\Sigma_0^{-1}$, so that if many of the lens–source pairs are at small redshift separation, and there is a bias on the photometric redshifts, then even with the $\Sigma_0^{-2}$ weighting which downweights nearby pairs, we still end up with a large bias in $\Delta \Sigma$. Consequently, the photometric redshift error distribution, which can be difficult to determine, is very important. Previous studies used the spectroscopic sample, which is brighter, with much lower photometric errors, supplemented by the CNOC2 spectroscopic redshifts down to $r = 21$ (Csabai et al. 2003), or used stacked images from the southern SDSS survey due to their lower photometric errors (for KPHEROZ; M. Blanton, private communication). Consequently, the applicability of the photometric redshift errors from these studies to a full catalogue based on fainter single images (that will have additional photometric redshift error due to noise) is questionable.

Fortunately, as will be shown in Section 3.3, we now have the capability of studying the errors in photometric redshifts directly for a representative subsection of our source sample rather than for atypical brighter or noiseless samples, and the application of these error distributions in the computation of the lensing signal can help eliminate the bias in $\Delta \Sigma$ due to photometric redshift error.

3.2 COMBO-17 distribution

Another commonly used option for the source redshifts is the use of a probability distribution rather than individual redshifts. This method has the disadvantage of leading to a larger boost factor due to the inability to weed out physically associated pairs. However, for the $r > 21$ galaxies, distributions are a better option than photometric redshifts due to photometric noise; consequently, we do not consider photometric redshifts for $r > 21$. 

© 2005 RAS, MNRAS 361, 1287–1322

Downloaded from https://academic.oup.com/mnras/article-abstract/361/4/1287/1117325
don 27 July 2018
The redshift distribution used for this paper is derived from data from the Classifying Objects by Medium-Band Observations, or COMBO-17 survey. COMBO-17 includes photometry in 17 passbands spanning the wavelength range from 350 to 930 nm, yielding far more information for the determination of photometric redshifts than the SDSS. The area of the survey used in the study that derived the redshift distribution covers 0.78 deg$^2$, spread over three disjoint regions (smaller than the full survey), so while the area is small, the concern about the derived redshift distribution $p(z|r)$ being unduly influenced by large-scale structure (LSS) is somewhat lessened. Wolf et al. (2003) includes the luminosity functions upon which these distributions were based. Since the distributions are for all photometric galaxies rather than those passing our lensing cuts, we may expect that they lie at slightly higher redshift than our catalogue on average; we show results of tests of this hypothesis in Section 6.4.2. A plot of this distribution (averaged over $r$) is shown in comparison with other distributions in Fig. 7.

Note that the COMBO-17 photometric redshifts have been used both directly for cosmic shear studies in the COMBO-17 survey itself (Brown et al. 2003; Heymans et al. 2004), and to estimate redshift distributions for other cosmic shear investigations (Heymans et al. 2005).

3.3 DEEP2

Another way to determine the true redshift distribution for our sources is to find another survey that is flux-limited and complete to a desired flux. As shown in Ishak & Hirata (2005) in the context of cosmic shear surveys, even just 100 spectroscopic redshifts may be sufficient to make this determination. Fortunately, the DEEP2 survey (Davis et al. 2003; Madgwick et al. 2003; Davis, Gerke & Newman 2004; Coil et al. 2004a) provides results that are useful for this purpose, with spectroscopic completeness well beyond $r = 21.8$, the limits of our source catalogue. The DEEP2 survey will eventually include spectroscopy of ~60,000 galaxies in four fields totalling 3.5 deg$^2$. While the targeting in three fields involves the use of photometric information to select galaxies with $z > 0.7$, the targeting in the extended Groth strip (EGS) does not attempt to place such restrictions, and because it overlaps with the SDSS, it may be used to study redshift distributions and photometric errors in SDSS data. Observations are complete in pointing one of field 1 (EGS), centred at Dec. +52° 12′ and RA 14°15′.7, which has area approximately 0.15 deg$^2$ (roughly 1/4 the area in the full EGS). The Groth strip is situated >50′ from the nearest of the three COMBO-17 fields, so the redshift distribution obtained from it can be considered statistically independent from the COMBO-17 results. The detectors on the CFHT (used for imaging for DEEP2 target selection) saturate at $R_{\text{sat}}$ ≈ 17.6 at the bright end, so no galaxies brighter than that limit have spectra, but fortunately those galaxies constitute a very small fraction (~2 per cent) of the source sample. Also, about 1/3 of the galaxies in the Groth strip were not targeted at all, which reduces the number of potential matches against SDSS.

The target selection in the Groth strip did involve some colour and magnitude information (Faber et al., in preparation). At $R < 21.5$, all galaxies are selected uniformly; fortunately, the majority of the galaxies in our lensing catalogue fall into this category. At fainter $R$, galaxies are classified as low or high redshift via colour cuts; low redshift galaxies are downweighted significantly, which makes the selection fairly complicated. To verify that this selection has a negligible effect on the results in this paper, all redshift distributions computed using DEEP2 data were recomputed taking into account selection probability (i.e. weighting each redshift by $1/p$). The value of $\langle \Sigma_c^{-1} \rangle$ for various values of lens redshift and with the weighted redshift distributions were computed and compared to the values from the unweighted distributions. Even for lens redshift $\sim 0.2$, which is at the high end of the distribution, and should be more sensitive to the source redshift distributions than most other lenses, the fractional change in the value of $\langle \Sigma_c^{-1} \rangle$ was on the order of a few tenths of a per cent. Consequently, we consider the selection function to be of negligible importance for the distributions presented in this paper, and for the remainder of this work we use the unweighted results.

To use the DEEP2 redshifts to compute redshift distributions and photometric redshift error distributions for our catalogue, we first matched between the DEEP2 spectroscopic catalogue in the Groth strip and our lensing catalogue. This step ensures that all distributions that we derive will apply to lensing-selected galaxies, which are expected to be at lower average redshift than all galaxies at the same magnitude, due to our requirements on the shape measurement. In principle, we could have used redshift distributions as a function of magnitude presented in Coil et al. (2004b), but since those were derived for all galaxies, they are expected to be at slightly higher average redshift than the distributions for lensing-selected galaxies. Once matching was complete, there were 278 matches, 162 at $r < 21$ and 116 at $r > 21$. Our requirement that there be a high-quality redshift had eliminated 33 potential matches, giving a redshift determination success rate of 89 per cent for lensing-selected galaxies overall, or 91 per cent for $r < 21$ and 86 per cent for $r > 21$.

We must be concerned about the effects of redshift failures on our results. The lack of knowledge whether or not the failures lie in a particular region of redshift space (e.g. higher redshift on average) introduces an unknown systematic into our results. First, we note that as discussed in Coil et al. (2004a), redshifts $z > 1.45$ cannot be measured by the DEEP2 survey. However, for the magnitude ranges of interest in this paper, this limit is effectively of no importance. Secondly, we find that the fraction of matches with failed redshift determination varies somewhat with magnitude (6 per cent failed at 18 $< r < 19$, 8 per cent at 19 $< r < 20$, 11 per cent at 20 $< r < 21$ and 14 per cent at 21 $< r < 22$), implying an increase in failures at higher redshift; we may also have a problem if the majority of the failures lie at a particular part of the redshift distribution in a given magnitude range. We can place bounds on the effect of such a systematic as follows: we compute the change in $\langle \Sigma_c^{-1} \rangle$ that results from assuming that all the failed redshifts were at 0 and at $\infty$ (which yields the $z \gg z_1$ asymptotic value of $\Sigma_c^{-1}$). We can compute the fractional error

$$\delta \Sigma / \Sigma = -\delta \langle \Sigma_c^{-1} \rangle / \langle \Sigma_c^{-1} \rangle = -f_{\text{failed}} \left(1 - \frac{\langle \Sigma_c^{-1} \rangle_{\text{failed}}}{\langle \Sigma_c^{-1} \rangle_{\text{measured}}}\right).$$

(24)

where $f_{\text{failed}}$ is the fraction of redshift failures (0.11), $\langle \Sigma_c^{-1} \rangle_{\text{failed}}$ is the average value of $\Sigma_c^{-1}$ for those galaxies that had failed redshift determination and $\langle \Sigma_c^{-1} \rangle_{\text{measured}}$ is the average value of $\Sigma_c^{-1}$ for those galaxies that were used to compute the signal using the observed redshift distribution. If we assume all failed redshifts were 0, then $\langle \Sigma_c^{-1} \rangle_{\text{failed}} = 0$, and therefore we expect a bias of ~11 per cent (or ~14 per cent for $r > 21$ and ~9 per cent for $r < 21$), where our use of the redshift distribution from the redshift determination successes overestimated $\Sigma_c^{-1}$ (and therefore underestimated the signal) by that amount. If we assume all failed redshifts are $\gg z_1$, then the effect depends on the lens luminosity (where it is less important for lower luminosity and redshift lenses, for which all sources were essentially at $\infty$ anyway) and the source sample. The degree to which the signal was overestimated in this case is given, for each luminosity bin and
Table 4. The estimated overestimation of the lensing signal if all redshift failures are at very high redshift, as a function of lens bin and source sample.

| Lens sample | $\delta(\Delta \Sigma)/\Delta \Sigma$, per cent |
|-------------|--------------------------------------------|
|             | $r < 21$ | $r > 21$ |
| L1          | 0.1      | 0.4      |
| L2          | 0.2      | 0.6      |
| L3          | 0.5      | 0.9      |
| L4          | 0.9      | 1.4      |
| L5          | 1.6      | 1.9      |
| L6          | 2.7      | 2.5      |

source sample, in Table 4. As shown, the effect is at most 2.7 per cent overestimation for L6 with $r < 21$ sources, much less for the fainter bins. The reality is somewhere in between these two extremes, most likely towards the higher end of the range quoted since we expect more redshift failures at higher redshift, but cannot easily be estimated, so we use these values to define the 95 per cent confidence interval. (This estimate did not take into account the effect of changes in $\Sigma_{1}$ on the weighting, since the use of non-optimal weighting should increase the errors without inducing a bias.)

We compared the magnitude distribution among the matches to that in our source catalogue. If they are not comparable, redshift distributions determined using the DEEP2 sample may not be applicable to our source sample. Table 5 shows a comparison of the samples. The first column shows the fractions of galaxies in magnitude bins for our source catalogue overall, and the second column shows the fractions of galaxies in the Groth strip (with 95 per cent confidence intervals from the binomial distribution for the latter due to its small size). Since they agree within the error bars, we need not worry about the Groth strip being very different from the full source catalogue. The third column shows the fractions in magnitude bins for the full SDSS photometric sample (i.e. no lensing-related cuts), Groth strip only, restricted to the magnitude range for which the CFHT detector does not saturate. The fourth column shows the fractions of galaxies actually targeted as a function of magnitude (for $r < 21.8$); this column is statistically consistent with the third column, implying that the targeting is indeed independent of the magnitude. Finally, the fifth column shows the fractions of galaxies as a function of magnitude for the matches between the DEEP2 redshift catalogue and our lensing catalogue with successful redshift determination, which is consistent with column two, the fraction of lensing-selected galaxies as a function of magnitude in the Groth strip before matching against DEEP2, with the exception that the DEEP2 galaxies are by necessity missing the $r < 18$ sample, which is roughly 2 per cent of our catalogue. Therefore, the fraction of galaxies categorized as high-redshift LRGs was similar in the two samples (9 per cent in our background catalogue, versus 11.5 $^{+3.5}_{-1.5}$ per cent in the DEEP2 matches at the 95 per cent CL), implying that there is no colour-dependence of the targeting and redshift failures with regards to this particular subsample.

While using the spectroscopic redshift distributions may seem to be the ideal way of determining $\Sigma_{1}$, there are two caveats that make this solution less promising. First, the use of average redshift distributions without any use of photometric redshift errors means that many more physically associated lens–source pairs are included in the calculation. While the resulting dilution of signal may be corrected for using boost factors, boosting is another potential source of systematic error since we cannot correct for intrinsic shear, and it is desirable that the boost factors be as low as possible. Secondly, because these distributions were determined using a small portion of the sky, they may be unduly influenced by LSS, and not representative of the redshift distribution in the survey as a whole (in Section 6.1, we show how the effects of LSS increase the error bars determined purely using statistics). The ideal solution to the second problem would be to have spectroscopic redshifts determined for a random sample of galaxies selected from the entire SDSS survey area, which could be used to determine source redshift distributions or photometric redshift error distributions.

Consequently, we also used these DEEP2 redshifts to determine photometric redshift error distributions, enabling us to use photometric redshift information to eliminate physically associated pairs, yet to also avoid any bias due to errors in the photometric redshifts. The determination of these error distributions was done by dividing the $r < 21$ matches into bins based on photometric redshift (in order to determine error distributions as a function of photometric redshift). Our computation of $\Sigma_{1}$ then takes into account the photometric redshift error distribution

$$
\Sigma_{1}^{-1}(z_{l}, z_{p}) = \int \frac{\sigma_{l}(z_{l}, z_{p})}{\eta_{l}(z_{l}, z_{p})} dz_{l}.
$$

The signal is then computed as usual, but instead of using $z_{l} = z_{p}$ (i.e. $\sigma_{l}(z_{l}, z_{p}) = \sigma_{l}(z_{l} - z_{p})$) we use the value computed by evaluating the integral in equation (25).

Results from our work with the DEEP2 data are shown in Section 6.1.

3.4 Reference distribution: LRGs

There is one subset of SDSS source galaxies for which excellent redshift information is known: Luminous Red Galaxies (LRGs) (Eisenstein et al. 2001). These galaxies have a well-known colour-redshift relation that allows photometric redshifts to be determined with excellent precision.

Table 5. The fraction of source galaxies in each magnitude bin for our source catalogue overall and in the Groth strip, all SDSS photometric galaxies in the Groth strip with $18 < r < 21.8$, all DEEP2 galaxies targeted with $r < 21.8$, and the matches between DEEP2 and our source catalogue. (95 per cent confidence intervals on all columns are from the binomial distribution; due to the constraint relating the sum of the numbers within each column, the error bars are anticorrelated.)

| Magnitude range | Lensing catalogue | Lensing catalogue, Groth strip | Photometric survey, Groth strip | DEEP2 targets | DEEP2 matches |
|-----------------|------------------|-------------------------------|-------------------------------|----------------|---------------|
| $r \leq 18$     | 0.022            | 0.020$^{+0.006}_{-0.010}$    | 0.000                         | 0.000          | 0.000         |
| $18 < r \leq 19$| 0.055            | 0.066$^{+0.020}_{-0.028}$    | 0.035$^{+0.012}_{-0.010}$     | 0.027$^{+0.017}_{-0.012}$ | 0.054$^{+0.033}_{-0.023}$ |
| $19 < r \leq 20$| 0.153            | 0.127$^{+0.013}_{-0.024}$    | 0.096$^{+0.018}_{-0.016}$     | 0.096$^{+0.023}_{-0.016}$ | 0.133$^{+0.046}_{-0.038}$ |
| $20 < r \leq 21$| 0.358            | 0.378$^{+0.044}_{-0.043}$    | 0.304$^{+0.027}_{-0.026}$     | 0.316$^{+0.041}_{-0.029}$ | 0.396$^{+0.058}_{-0.050}$ |
| $21 < r \leq 21.8$| 0.411           | 0.410$^{+0.045}_{-0.044}$    | 0.565$^{+0.028}_{-0.028}$     | 0.558$^{+0.042}_{-0.043}$ | 0.417$^{+0.060}_{-0.059}$ |
For the LRGs selected according to criteria shown below, the photometric redshifts perform significantly better than for the general source sample. Furthermore, as shown in Padmanabhan et al. (2004), redshift distributions for LRGs from the 2dF-Sloan LRG and Quasar (2SLAQ) survey can be used to construct very reliable redshift distributions for these galaxies. Note that for LRGs, the redshifts used were from a simple template code discussed in Padmanabhan et al. (2004) rather than from KPHOTOZ; however, for most regions of colour space occupied by LRGs, the results are highly correlated.

The selection criteria used for the LRGs are as follows:

(i) \( d_\perp \equiv (r - i) - 0.125(g - r) > 0.45 \)
(ii) \( c_1 \equiv 0.7(g - r) + 1.2(r - i - 0.18) > 1.6 \)
(iii) \( g - r < 2.5 \)
(iv) \( r - i < 1.5 \)
(v) \( 0.4 < z_p < 0.65 \)
(vi) \( r > 19. \)

These criteria were chosen as the combination of those from Eisenstein et al. (2001) and from Padmanabhan et al. (2004) that best suited the requirements of this work, namely very low levels of contamination from non-LRGs, from low-redshift LRGs and from stars, yet sufficient numbers that the LRG sample can be used as sources for weak lensing with reasonable signal-to-noise ratio.

Imposing these criteria on our source catalogue yields \( 2.884 \times 10^6 \) LRGs total, though only about 65 per cent are in regions with lenses. A plot of the LRG redshift distribution derived using 2SLAQ work (Padmanabhan et al. 2004) is in Fig. 6; for this distribution, only the 72 per cent of the LRG sample with photometric redshift greater than 0.45 were used because the inversion method gave unreliable results for \( 0.4 < z_p < 0.45 \). As shown, the redshift distribution peaks around \( z \sim 0.5-0.55 \), which illustrates another advantage of using LRGs as sources: when we use SDSS main spectroscopic sample galaxies as lenses, with redshifts mostly below \( z \sim 0.25 \), \( \Sigma_c^{-1} \) does not vary strongly with source redshift for sources at such high redshift, so an error in these distributions will make a very small difference in the results.

Note that approximately 57 per cent of the LRGs in the catalogue are fainter than \( i \) model magnitude of 20. Padmanabhan et al. (2004) only includes redshift distributions down to this limit, so our use of that inversion method for this sample is suspect. However, tests of photometric redshifts using the DEEP2 matches indicate that the mean bias is the same for the \( i > 20 \) LRGs as for the \( i < 20 \) LRGs (0.02), and the rms scatter is larger but not excessively so (0.07 versus 0.05), which suggests that the use of the inversion method in Padmanabhan et al. (2004) with the same error distributions for \( i < 20 \) and \( i > 20 \) is justified here.

In addition to computing the signal with the full sample using the photometric redshifts directly, and with the restricted sample using the redshift distribution, we also computed it with photometric redshifts on two subsamples, those with \( d_\perp > 0.5 \) and \( d_\perp > 0.55 \). Because these stricter cuts would help avoid contamination by lower redshift galaxies being scattered up to higher redshift, we try imposing them on our sample; if the signal does not differ when we do this, then we know that using the full sample is relatively safe.

Fig. 2 shows the \( r \) model magnitude distribution of sources classified as high-redshift LRGs, and Fig. 4 shows their \( R_c \) distribution, in comparison to the values for the overall source sample. As shown, the LRG sample is, on average, at lower \( R_c \) and fainter magnitudes than the other source samples, and thus we may be concerned that it will have more significant shear calibration bias, as evidenced by the confidence intervals in Table 3.

In principle, given the excellent redshift information for the high-redshift LRGs, one solution to our lack of reliable redshifts for all sources would be to do \( g \)-\( g \) lensing entirely with LRGs as sources. However, we then are faced with a problem of statistics, since LRGs are such a small fraction of the sources available (9 per cent), yielding large statistical errors on \( \Delta \Sigma \). This problem is the reason why our test for calibration bias is so useful; it allows us to use the excellent redshift information from LRGs to check the calibration of the lower redshift source samples, which contain many more galaxies, enough to obtain excellent statistical error bars.

### 4 OTHER SYSTEMATICS ISSUES

Here we consider a number of remaining systematics issues that arise when computing the lensing signal. Some are calibration uncertainties that scale with the signal, similar to the shear and redshift distribution calibration uncertainties discussed previously; others are systematics that do not scale with the signal, and have amplitudes that depend on the angular or physical lens–source separation.

#### 4.1 Random points test

The random points test requires computing signal \( \Delta \Sigma_{\text{rand}} (r) \) using random lens catalogues (i.e. sets of random positions generated with the angular mask of the spectroscopic survey area, using the same lens redshift and magnitude distributions as the real lens sample). In practice, these distributions are preserved by drawing the redshifts and magnitudes from the real sample, without replacement. The random points test is useful because a non-zero signal reveals the presence of spurious shear power (systematic shear) in the source catalogue, which would lead to an additive bias in the lensing signal. The random catalogues used for this work were generated using MANGLE (Hamilton & Tegmark 2004).

In the absence of systematic shear, we expect the random points test to show zero signal around random points. However, there is a slight smearing of images in the scan direction because the charge...
transfer is not continuous, instead occurring in very quick transfers from pixel to pixel, so that the PSF is convolved with a rectangle 1 pixel (0.4 arcsec) wide along the scan direction only. The result is that the PSF is not circularly symmetric. This effect may be exacerbated if the scan rate and charge transfer rate are not perfectly matched, leading to convolution on a scale larger than 1 pixel.\(^4\) While in principle PSF correction algorithms should correct for PSF asymmetry, in practice this is difficult to do perfectly. In addition, since the charge transfer efficiency may also depend on the amount of charge, the PSF asymmetry may be different for faint objects than for bright ones, but since \textsc{photo} fits the PSF using stars around \(r \approx 19\), code that uses the \textsc{photo} PSF may not be able to fully correct for this effect. We found that the re-Gaussianization scheme overcompensates for the effect.

We address the presence of this spurious signal by determining it to as high precision as possible using \(N\) random catalogues (where the number of random catalogues is limited by processor time), then subtracting it from the observed signal. This procedure results in the error bars on the signal rising by a factor of \(\sqrt{(N + 1)/N} \approx 1 + 1/2N\). Our choice of \(N = 24\) means that the error bars only increase by 2.1 per cent. Results for the random catalogue test are shown in Section 6.3.

However, random catalogue signal subtraction, while widely accepted in the literature, ignores the possibility that fluctuations in the number density and systematic shear may be correlated. We describe this problem as follows: A g-g weak lensing measurement entails computing the correlation \(\langle n_r \gamma \rangle\). The measured number density of lenses \(n\) can be decomposed into \(n = \bar{n} + \delta n\), an average density of sources on the sky plus fluctuations. The measured total shear \(\gamma\) can be decomposed into \(\gamma = \gamma_+ + \gamma_{\text{sys}}\), the true shear field plus systematic shear. When we compute the signal around random points, we obtain the quantity \(\langle \delta n \gamma \rangle\). Random catalogue subtraction thus corresponds to measuring

\[
\langle \delta n \gamma_+ \rangle - \langle \delta n \gamma_{\text{sys}} \rangle = \langle \delta n \gamma_+ \rangle = \langle \delta n \gamma_+ \rangle + \langle \delta n \gamma_{\text{sys}} \rangle.
\]

The first term on the right side, \(\langle \delta n \gamma_+ \rangle\), is the correlation that we hope to measure, but the second term is an additive systematic error that has not been discussed in previous works on g-g weak lensing. We cannot assume that \(\delta n\) and \(\gamma_{\text{sys}}\) are completely uncorrelated, because both quantities are slightly correlated in some way with the PSF. There are several ways in which such an effect could become significant. For example, since there is a gap between camcols on the SDSS camera, the same region must be rescanmed with some offset to fill in that gap on a different night, which may have very different seeing and other conditions, such that \(\gamma_{\text{sys}}\) fluctuates on small scales. If the number density of lenses also fluctuates on the same scale, then we could have some non-zero contribution from the \(\langle \delta n \gamma_{\text{sys}} \rangle\) term. For this work, we assume that this term is negligible, since the correlation should be small and has never been detected before, but this issue should be addressed more fully in future work to ensure that this assumption is reasonable.

### 4.2 45° Test

Another useful test of systematics in the lensing signal is the 45° test, which requires computing the lensing signal with the coordinate system rotated by 45°. By inversion symmetry, the 45° rotated signal \(\gamma_{45}\) should vanish, with weak lensing only contributing to the tangential shear, \(\gamma_+\). The presence of such a signal could indicate a variety of shear systematic errors, since they generally contribute both to \(\gamma_+\) and \(\gamma_{45}\). We computed the 45° rotated signal for all three source samples, so that together these tests were done for all sources in the catalogue (we assume that the choice of redshift distribution does not matter for this test, as \(\gamma_{45}\) would be non-zero due to shear systematics, so we only use one choice of redshift determination method for each of those samples). Results for this test are in Section 6.3.

### 4.3 Star–galaxy Separation

Star–galaxy separation is an important issue for weak lensing, since the inclusion of stars in the source sample would dilute the signal. Thus, a balance must be struck, ensuring the purity of the galaxy sample used as sources, while avoiding being overly conservative and eliminating too many galaxies in the course of doing this separation (which would lead to poor statistics). Star–galaxy separation for this catalogue was accomplished via two cuts: first, the requirement that the \textsc{photo} flag OBJC\_TYPE be equal to 3, or galaxy; secondly, the requirement that \(R_2 > 1/3\) (i.e. the object must be 50 per cent larger than the PSF). The type determination for DR1 and DR2 is described in Lupton et al. (2001); in brief, it utilizes the linear combination of fits of galaxy profiles to two models (exponential and de Vaucouleurs), then allows the ratio of the flux in a fit to a PSF shape to that in the linear combination of galaxy models to determine the type, via the requirement that \(m_{\text{PSF}} - m_{\text{model}} \geq 0.24\) for galaxies. As shown in Fig. 3 in that paper, this procedure ensures a relatively pure galaxy sample even close to \(r = 22\), with stellar contamination fraction (determined using \textsc{Hubble Space Telescope}, or \textsc{HST} data in the Groth strip) being negligible at \(r\) brighter than 21, and 7 per cent for \(21 < r < 22\). It is also clear from that figure that OBJC\_TYPE is much more likely to default to calling galaxies to stars rather than vice versa, with galaxy contamination of \(\sim 33\) per cent in a sample of ‘stars’ for \(21 < r < 22\). While there are probabilistic methods that are more accurate at the faint end (\(r > 21\)), such as that used in Scranton et al. (2002) and Sheldon et al. (2004), OBJC\_TYPE’s conservative tendency should not cause significant stellar contamination, though it does reduce the density of sources in poor seeing at faint magnitudes.

In order to check that our shape measurement cuts reduce stellar contamination due to failures in OBJC\_TYPE, we used publicly available catalogues from the Great Observatories Origins Deep Survey (GOODS), carried out via the \textsc{HST} (Giavalisco et al. 2004). Characterizations of stars versus galaxies are much more accurate to fainter magnitudes in space-based surveys such as this because the PSF is much smaller. We found 577 objects at \(r < 21.8\) (of which 60 per cent are at \(r > 21\)) that were matches between the full SDSS photometric catalogue and the GOODS north field, centred at Dec. +62°15’ and RA 12°37’\(\pm\)7. We found that the ‘galaxy’ classification is incorrect about 1.5 per cent of the time for \(r < 21\) and 7 per cent of the time for \(21 < r < 21.8\) objects, and the ‘star’ classification is incorrect a larger fraction of the time (3 per cent at \(r < 21\) and 40 per cent at \(21 < r < 21.8\)), confirming the results from Lupton et al. (2001) that OBJC\_TYPE tends to default to calling small, faint objects stars.

However, when we restrict to the subset of 146 objects that passed all shape measurement and other criteria to be included in our source catalogue, of which 73 are at \(r < 21\) and 73 at \(r > 21\), we find that only 2 (1.4 per cent) of those included are stars, or 0 per cent contamination in the \(r < 21\) sample and 2.7 per cent contamination in the \(r > 21\) sample. We see that the resolution factor and other

\(^4\)We thank James Gunn for pointing out this effect.
cuts reduce stellar contamination by a factor of 3 from the result using OBJC_TYPE alone. Using the binomial distribution to get 95 per cent confidence intervals on our stellar contamination estimates yields [0, 0.040] \( r < 21 \) and [0.003, 0.096] \( r > 21 \). Technically, we should take into account that the GOODS north field is at \( 1/\sin b \sim 1.2 \), but the average value for the full lens sample is \( 1/\sin b \sim 1.4 \), so we might expect slightly higher stellar contamination in the full catalogue than that computed for the GOODS field. However, because the dependence of stellar contamination on \( 1/\sin b \) is difficult to model, we do not attempt any correction.

To check the signal for contamination of our source catalogue by stars, we computed the signal at low versus at high galactic latitude (cutting at \( \sin b < 0.7 \)) to compare the results. Note that while finding a lower signal at low galactic latitude may indicate a problem with star/galaxy separation, it may also indicate the presence of other systems. In particular, since the extinction is greater at low galactic latitudes, galaxies near the faint end at a given magnitude were actually lower signal-to-noise measurements at low galactic latitude than they were for high galactic latitude, so in principle there could be a shear systematic causing a difference between these two samples as well. The results of this test will be shown in Section 6.3.

### 4.4 Seeing dependence of calibration

Because our ability to correct the galaxy image for effects due to the PSF depends on the relative size of the galaxy and the PSF, we must consider the possibility of a seeing-dependent shear systematic. Consequently, for each galaxy, we consider the size of the PSF used, and split our sample into ‘good seeing’ (PSF size less than the median value, 1.25 pixel in the \( r \) band) and ‘bad seeing’ (PSF size greater than the median value). The signal was then computed using these two source samples, and compared. Results for this test will be shown in Section 6.3.

### 4.5 \( R_2 \) dependence of calibration

Because some calibration biases may be more prominent at lower \( R_2 \), it is important to check for \( R_2 \)-dependence of the calibration. There are several effects that could lead to apparent \( R_2 \) dependence of the calibration: noise rectification bias, selection biases and biases due to PSF correction, which would be particularly important for less-well resolved galaxies. We computed signal using sources with \( R_2 > 0.55 \) in each band to check for bias; results of this test are presented in Section 6.3.

### 4.6 Systematic differences between bands

Like many other studies, we used shape measurements averaged over two bands, the \( r \) and \( i \) bands. While the shape measurements between the two bands may legitimately differ for individual objects, due to, for example, spectral differences in emissions from the disc versus from the bulge of spiral galaxies, we also checked to ensure that the signal computed with the shape measurements from each band individually gives the same \( \Delta \Sigma \). The results of this test will be shown in Section 6.3.

### 4.7 Boosts

As discussed in H04, the lensing signal at small transverse separations is diluted by the inclusion of sources that are physically associated with the lens (i.e. are in the same group or cluster), and therefore are not really lensed. To correct for this effect, the signal for a given luminosity bin is boosted according to the weighted number of galaxies per unit area relative to the number from random catalogues. The signal is multiplied by a factor

\[
B(r) = \frac{n(r)}{n_{\text{rand}}(r)},
\]

where \( n(r) \) is the weighted number of galaxies per unit area when the signal is computed, and \( n_{\text{rand}}(r) \) is the same computed with random lens catalogues. The number from random catalogues takes into account the decrease in the number per area with radius due to survey edge effects, so the boost can accurately account for the dilution of signal by physically associated pairs. Consequently, \( B(r) \sim 1 - \xi_b \) is the lens–source correlation function.

The boosts add two sources of statistical error and two sources of systematic error. In the rest of this section we consider each of these.

#### 4.7.1 Statistical errors

The statistical error arises because the boost factor is the ratio of two noisy quantities, \( n(r) \) and \( n_{\text{rand}}(r) \). The noise in \( n_{\text{rand}}(r) \) can be minimized by computing signal from a large number of random catalogues, where 24 are used for this work. However, for subsamples with a small number of lenses, \( n_{\text{rand}}(r) \) is still slightly noisy at small separations due to the small size of the radial bins; this noise is taken into account in the bootstrap by multiplying the boost from each data set and radial bin by a Gaussian random number of mean 1 and standard deviation equal to the fractional error in \( n_{\text{rand}}(r) \) as computed from the random catalogues. The noise in \( n(r) \) is taken into account naturally by the bootstrap, since each bootstrap-resampled data set will have a slightly different \( n(r) \) used for the boost. Hence, the statistical error due to the boosts is simple to take into account.

#### 4.7.2 Systematic error: non-uniformity of boost factor

One potential systematic error arises because both \( \Delta \Sigma \) and the boost factor vary strongly with luminosity at the bright end of the lens sample, so the luminosity bins that are 1 mag wide may be too wide to properly compute the signal in the innermost radial bins, \( r < 50 \, h^{-1} \) kpc, where the boost is most important. By averaging over a large range in luminosity, with \( \Delta \Sigma(r) \) and \( B(r) \) varying with luminosity, we may run into a situation where the product of two averages \( \Delta \Sigma(r)B(r) \) separately averaged over luminosity, then multiplied) differs significantly from what we really want, the average of products \( B(r)\Delta \Sigma(r) \) averaged over luminosity).

One rudimentary method of detecting the effects of using wide luminosity bins on the boost factor is to split the brightest luminosity bin in half, compute the signal separately for each half (boosting each one individually), then average the signal from each half. The resulting signal can be compared against the signal computed using the full luminosity bin. Results for this test will be presented in Section 6.3.

#### 4.7.3 Systematic error: magnification bias

Another boost-related source of error is magnification bias, since the number of galaxies per area around real lenses may not be expected to be the same as that around the random points. There are three competing effects: first, that due to the magnification \( \mu = 1 + 2\kappa \) (in the weak lensing limit), where \( \kappa = \Sigma/\Sigma_c \), the number of lens–source pairs per unit area on the sky will decrease; secondly, that the magnification means that fainter sources will be visible than would have been otherwise, and therefore the number of lens–source pairs per unit area will increase; and thirdly, the magnification changes the...
resolution factors of the source galaxies. The competition between the first two effects can be quantified by $s = \frac{d \log_{10} N(m)}{dm}$, where $N(m)$ is the total number density of source galaxies given a faint magnitude limit of $m$. For the $r > 21$ sample, we must take into account the loss of sources at the bright magnitude limit of 21, and compute

$$s = \frac{d \log_{10} N(m, \ldots m_+)}{dm_+} - \frac{d \log_{10} N(m_-, \ldots m_+)}{dm_-},$$

where $m_+$ is the faint magnitude limit and $m_-$ is the bright magnitude limit. We compute $s$ separately for each source sample: for the $r < 21$ source sample, we find $s = 0.36$; for $21 < r < 21.8$, we find $s = 0.47 - 0.57 = -0.10$ (i.e. when the magnitudes are shifted brighter, we lose more galaxies at the bright end than we gain at the faint end because of our cuts on the shape measurement); and for the LRG sample, we find $s = 0.27$.

The resolution factor dependence of this effect has not been previously evaluated. If we take the Gaussian approximation for the galaxy, $T^{(p)} = T^{(\mu)} + T^{(\alpha)}$, and note that $T^{(\mu)} \propto \mu$ in the weak lensing regime, we find that the effect of magnification is to adjust the resolution factor by

$$\delta R = -\delta \left( \frac{T^{(\alpha)}}{T^{(\mu)}} \right) = 2(1 - R_2)\kappa.$$

The number of galaxies that are gained due to the resolution factor cut is then

$$\frac{\delta N}{N} \propto 2(1 - R_{2, \text{min}}) n(R_{2, \text{min}}) \kappa,$$

where $n(R_{2, \text{min}})$ is the resolution factor distribution normalized to $\int R_{2, \text{min}} n(R_2) dR_2 = 1$. The total change in the number density of galaxies is then

$$\frac{\delta N}{N} = \left[ 5s - 2 + 2(1 - R_{2, \text{min}}) n(R_{2, \text{min}}) \right] \kappa.$$

For the three source samples, using the values of $s$ given above and $n(R_{2, \text{min}})$ from Section 2.2.2, we estimate $\delta N/N = 1.9\kappa, 0.7\kappa$ and 3.1$\kappa$ for $r < 21$, $r > 21$ and LRG samples, respectively. (Without taking into account the effect of the change in $R_2$, we would have had $\delta N/N = -0.2\kappa$, $-2.5\kappa$, and $-0.6\kappa$, so this effect significantly changes our sensitivity to magnification bias.)

The convergence can be simply estimated for roughly power-law profiles $\gamma_r \propto r^{-\alpha}$ as

$$\kappa(r) = \left( \frac{2}{\alpha} - 1 \right) \gamma_r(r) + \kappa(\infty),$$

where $\kappa(\infty)$ represents the mass-sheet degeneracy. We ignore this since our boost factors approach unity at large separations. Since for our galaxies we find $\alpha \approx 0.85$, it then follows that $\kappa \approx 1.4\gamma_r$. We can then use computed values of $\Delta \Sigma$ and $\langle \Sigma^{-1} \rangle$ to estimate $\kappa(r)$. Table 6 shows the best-fitting power-law $\kappa(r)$ and our resulting predictions for $\delta N/N$ for L3–L6; L1 and L2 are not used because the shear is statistically consistent with zero in these bins, and therefore so is $\kappa$.

This effect is not taken into account explicitly, and therefore may lead to a systematic bias in the signal, since we assume that the increase in sources around lenses is due to physically associated pairs. For all samples, we may be overestimating the boosts and therefore $\Delta \Sigma$ (particularly for LRGs). We do not, however, attempt to correct for this effect explicitly, since until a concrete detection is made, these estimates should not be treated as certain. For the 2$\sigma$ confidence intervals, we use the estimates in Table 6 as the expected value, with an expected uncertainty equal to half the magnitude of the estimate.

### Table 6. Parameters of power-law fits $\kappa(r) = \kappa_0 (r/1\ h^{-1}\ Mpc)^{-\alpha}$ to the lensing signal. The resulting value of $\delta N/N$ as a function of lens and source sample is shown as well.

| Lens sample | $10^9\kappa_0$ | $\alpha$ | $\delta N/N$ (30 $h^{-1}$ kpc) | $\delta N/N$ (100 $h^{-1}$ kpc) |
|-------------|---------------|---------|-----------------------------|-----------------------------|
| L3          | 5.5           | 0.68    | 0.011                       | 0.005                       |
| L4          | 4.7           | 0.88    | 0.020                       | 0.007                       |
| L5          | 7.9           | 1.00    | 0.050                       | 0.015                       |
| L6          | 29            | 0.95    | 0.15                        | 0.049                       |

#### 4.8 Intrinsic alignments

Satellite galaxies that are physically associated with a lens will not be lensed; however, it is often difficult to remove them from the source sample in the absence of good redshift information, which is why we must use the boost factors described in the previous section. If the satellites are actually aligned with the lens in some way (radially or tangentially elongated), this alignment will cause a false lensing signal. This effect is well motivated theoretically (Lee & Pen 2001; Hui & Zhang 2002), and there are several claimed detections in the literature of correlations between the galaxy density (or specific features thereof, such as the supergalactic plane) and intrinsic ellipticities (e.g. Flin & Godlowski 1989; Lee & Pen 2002; Navarro, Abadi & Steinmetz 2004). We use $\Delta \Sigma$ to represent the signal due to intrinsic shear, which is the shear signal that would be computed using our shear estimator with a population of physically associated galaxies. Note that the galaxy density–shear correlation that concerns us is distinct from the intrinsic shear–shear correlations that add spurious power to cosmic shear surveys (Croft & Metzler 2000; Heavens, Refregier & Heymans 2000; Lee & Pen 2000; Catelan, Kamionkowski & Blandford 2001; Crittenden et al. 2001; Jing 2002).

In order to place upper limits on contamination due to intrinsic shear, we must determine what fraction $f_c$ of the lens–source pairs are physically associated, and the typical intrinsic tangential shear $\Delta \gamma_{\text{int}}$ of these sources. To compute $f_c$, we use our results for $B(r)$. Limits on $\Delta \gamma_{\text{int}}$ (on the intrinsic ellipticity or position angle) have been obtained by Lee & Pen (2001), Bernstein & Norberg (2002) and H04; we use the results from H04 as these are the tightest constraints on relevant scales. Results for this estimation will be shown in Section 6.3.8.

#### 4.9 Correction for non-volume limited lens sample

In this section, we describe the corrections that account for the fact that the lens samples are not volume limited, when comparing between different source subsamples. These corrections

© 2005 RAS, MNRAS 361, 1287–1322
required us to compute the mean weighted luminosity for each lens–source sample combination. This computation was done as a function of transverse separation in order to check for any systematic effects.

We found that there is a slight variation in the mean weighted luminosity with transverse separation, because brighter lenses are more clustered, so in the regions where physically associated pairs are abundant (i.e. where \( \xi_s(r) \gg 0 \)), the mean weighted luminosity of the sample was higher than at larger radii where \( \xi_s(r) \approx 0 \). We see the same trend in the mean weighted redshift of the sample, as well. However, this effect does not really mean that the mean weighted redshift or luminosity corresponding to the computed signal is larger at small radii, since the excess pairs around the brighter lenses that cause this effect do not contribute to the signal.

We also expect that \( L_{\text{eff}}(r) \) and \( z_{\text{eff}}(r) \) may increase at large \( r \), because for more nearby (less luminous) lenses, the same transverse separation corresponds to a larger angular scale, and therefore survey edge effects may cause nearby lenses to lose pairs faster than distant lenses. The observation of such an effect would indicate a problematic selection effect, but fortunately we do not see it at the maximum transverse separations studied in this paper (2 \( h^{-1} \) Mpc).

The correction when comparing \( \Delta \Sigma \) at slightly different \( L_{\text{eff}} \) is as follows: we use the results of the signal averaged over radius in each luminosity bin to derive a relation \( \Delta \Sigma(L) \). Then, when we compare signals at two slightly different luminosities (typical differences are less than 1 per cent) \( L_{\text{eff}} \) and \( L_{\text{eff}} + \delta L \), we assume that rather than being 1, the ratio of the two signals should be

\[
\frac{\Delta \Sigma(L_{\text{eff}} + \delta L)}{\Delta \Sigma(L_{\text{eff}})} = 1 + \frac{(\Delta \Sigma/L)_{\text{ls}} \delta L}{\Delta \Sigma(L_{\text{eff}})}. \tag{33}
\]

If this correction is non-trivial compared to the statistical error, then we must apply it; otherwise we do not.

5 APPLICATION OF THE SYSTEMATIC TEST

In this section, we describe our test to determine the effects of the redshift distributions and systematic errors described in Sections 3 and 4, and show calculations justifying our assertion that cosmology plays a negligible role.

5.1 Methodology

First, in order to compute bootstrap errors on the results, we divide our lens catalogue into 200 bootstrap subsamples as described in Section 2.4.3. For each lens luminosity bin and source sample, we compute the signal in each of the bootstrap subsamples, using 46 logarithmically spaced radial bins from 20 to 2000 \( h^{-1} \) kpc in order to measure \( \Delta \Sigma(r) \) more sensitively where it varies the most, at small transverse separation. (Plots of signal shown in this paper will have many radial bins averaged at small \( r \) so they are easier to read.)

Next, to account for physically associated pairs and systematic shear, we used random lens catalogues to compute \( \Delta \Sigma_{\text{rand}}(r) \) and \( n_{\text{rand}}(r) \). Twenty-four random catalogues were used, so that these functions were determined reasonably smoothly.

After this procedure, we performed bootstrap resampling, generating 2500 data sets for each lens luminosity bin \( i \) and source sample \( \alpha \) as follows (to avoid confusion, we use Roman letters to denote lens luminosity bins and Greek letters to denote source samples).

(i) For each lens–source sample combination, we use the same set of random numbers (in order to take into account correlations) drawn from a uniform distribution to choose 200 bootstrap subsamples randomly, with replacement. The signal from these subsamples are then averaged to get signal for each of the lens–source sample combinations, \( \Delta \Sigma_{\text{rand}}(r) \).

(ii) We compute boost factors \( B_{i,\alpha}(r) \) using the weighted number of pairs per unit area \( n(r) \) relative to the average weighted number from random catalogues \( n_{\text{rand}}(r) \). The error in \( n(r) \) is taken into account by the bootstrap procedure itself, and the error in \( n_{\text{rand}}(r) \) is taken into account via multiplication by a Gaussian random number of mean 1 and standard deviation equal to the fractional uncertainty in \( n_{\text{rand}}(r) \), where a different random number is used for each \( i,\alpha \), and radial bin.

(iii) We use the measured random catalogue signal, taking into account its uncertainty via multiplication by a Gaussian random number of mean 1 and standard deviation equal to the fractional uncertainty in the random catalogue signal, where a different number is used for each \( i,\alpha \), and radial bin. This procedure yields \( \Delta \Sigma_{\text{rand}}(r) \).

(iv) Finally, we compute the signal for each data set, \( \Delta \Sigma_{i,\alpha}(r) = B_{i,\alpha}(r)(\Delta \Sigma_{\text{rand}}(r) - \Delta \Sigma_{i,\alpha}(r)) \). The random catalogue signal (not just the signal measured with the real lens catalogue) must be multiplied by the boost factor because all galaxies, whether physically associated with the lenses or not, have systematic shear associated with them.

Once the data sets have been generated using this procedure, we can then do the following.

(i) By averaging over the 2500 data sets, we can compute the mean value of \( \Delta \Sigma_{i,\alpha}(r) \) (and other quantities such as \( B_{i,\alpha}(r) \)) and its standard deviation.

(ii) Using the bootstrap errors from the previous step for weighting purposes, we can compute a value of \( \Delta \Sigma_{i,\alpha} \equiv (\Delta \Sigma_{i,\alpha}(r)/1 \ h^{-1} \text{Mpc})^{0.85} \) for each bootstrap-resampled data set. The \( (r/1 \ h^{-1} \text{Mpc})^{0.85} \) is included in order to increase signal-to-noise ratio on the averaged value. Only bins with \( r > 30 \) kpc/h are included in this calculation.

(iii) By averaging over the 2500 values of \( \Delta \Sigma_{i,\alpha} \) we can get the mean value \( \langle \Delta \Sigma_{i,\alpha} \rangle \) and its covariance matrix, where its value for overlapping source samples will have very high correlations (the extreme cases of identical source samples but different methods of computing \( \Sigma^{-1} \) typically have correlation coefficient \( r = 0.99 \)). This \( \langle \Delta \Sigma_{i,\alpha} \rangle \) will be used to compare the signal for different sets of sources, the same lens sample.

(iv) For each luminosity bin \( i \), we can compute ratios of the averaged signal with different sets of sources,

\[
R_{i,\alpha,\beta} = \frac{\langle \Delta \Sigma_{i,\alpha} \rangle}{\langle \Delta \Sigma_{i,\beta} \rangle}, \tag{34}
\]

and its error. The errors on the numerator and denominator, and their covariance, were used to compute confidence intervals on \( R_{i,\alpha,\beta} \) using the full non-Gaussian error calculation, which is important for those cases in which either number may be consistent with zero. (For details of the calculation of confidence intervals on the ratio of two correlated variables, see Appendix B.) Due to the high correlations between some of the source samples, the ratio could often be computed to extremely high accuracy.

(v) For many of the tests described below, the ratio test was also done for the signal averaged over luminosity bins as well. This comparison must be done with care; since the relative weights of the signal in the different luminosity bins varies from sample to sample (e.g. all LRGs contribute to each luminosity bin, but there are many more \( r < 21 \) galaxies in the faint bins than in the bright...
ones due to the $z_p > z_l + 0.1$ cut, so for $r < 21$ the fainter bins get more weight than the bright ones relative to the LRG sample), we must be careful that when comparing samples $\alpha$ and $\beta$ averaged over luminosity, we use the same weights for averaging each sample over luminosity. If the weights differ, then we could mistakenly be led to believe that there is a calibration error. So, for each bootstrap data set, we compute the average value of $\Delta \Sigma$ via

$$\Delta \Sigma_{all,\alpha} = \sum_i w_i \Delta \Sigma_{i,\alpha},$$

where

$$w_i = \frac{\Delta \Sigma_{i,\alpha}}{\sum_j \Delta \Sigma_{j,\alpha}}.$$

Here, $\Delta \Sigma_i$ is some approximation to the actual value of $\Delta \Sigma_i$ (in practice, the value averaged over subsamples $\alpha$ and $\beta$), and

$$v_i = \frac{1}{\sigma_i^2 + \sigma_i^2 - 2\text{Cov}(\Delta \Sigma_{i,\alpha}, \Delta \Sigma_{i,\beta})}.$$

The values of $\sigma$ used in this expression are from the covariance matrix already obtained for $\langle \Delta \Sigma_{i,\alpha} \rangle$. By averaging over bootstrap data sets, we can then get values $\langle \Delta \Sigma_{all,\alpha} \rangle$, which allow us to compute ratios

$$R_{all,\alpha,\beta} = \frac{\langle \Delta \Sigma_{all,\alpha} \rangle}{\langle \Delta \Sigma_{all,\beta} \rangle},$$

and obtain non-Gaussian confidence intervals as usual.

### 5.2 Cosmology dependence

In this section, we show why the systems test is essentially cosmology independent. We compute $\Delta \Sigma$ by computing $\gamma_i$, (which is independent of cosmology, except via the weighting scheme since $w_i \propto \Sigma_i^{-2}$, which should not change the results, only the error bars) and $\Sigma_i^{-1}$ with our assumed cosmology. Now, we propose that our cosmological model is wrong, so the true value of $\Delta \Sigma$ can be computed from the same $\gamma_i$ that we computed, but with a different $\Sigma_i^{-1}$ that we will call $\Sigma_i^{-1,\text{true}}$. Our results for a given lens and source sample are then related to the true $\Delta \Sigma$ by the relation

$$\Delta \Sigma_{\text{measured}} = \Delta \Sigma_{\text{true}} \frac{\Sigma_{c,\text{true}}}{\Sigma_{c,\text{guess}}},$$

so we can define a ‘cosmology factor’ $f_{\text{cos}} = \Sigma_{c,\text{true}}/\Sigma_{c,\text{guess}}$, relating the true and measured signal. Our concern, then, is that $f_{\text{cos}}$ for a given lens and source sample will be different for each source sample, so that our comparison of $\Delta \Sigma$ computed with different source samples is invalid.

We do a simple test for three lens redshifts ($z_l = 0.02$, the minimum value; 0.1, the typical value; and 0.25, on the high-redshift tail of the lens redshift distribution), and two source redshifts ($z_s = z_l + 0.1$ and 0.7) that span the range of source redshifts used. Furthermore, we test two cosmologies that are drastically different from the one assumed: a flat cosmology with $\Omega_m = 1$, and an open cosmology with $\Omega_m = 0.3$. For each cosmology, lens redshift, and source redshift, we can compute $f_{\text{cos}}$, relating our measured $\Delta \Sigma$ and the true value, and compare the values of $f_{\text{cos}}$ between the different source samples for a given lens redshift and cosmology. The results of this test are shown in Table 7.

We can see from Table 7 that while the measured value of $\Delta \Sigma$ may be off from the true one by a significant fraction, the differences between $f_{\text{cos}}$ for the same lens redshift but different source redshift are less than 1 per cent, even for these extreme cosmologies for all cases.

In reality, most of the lenses are near $z_s \sim 0.1$, most of the sources are at more intermediate values of redshift and the allowed ranges of cosmology is much smaller than the extreme cases considered here. Consequently, we can state with confidence that cosmology plays a negligible role in the comparisons of source samples for the allowable range of cosmology and lens redshift range probed by the SDSS.

There were several oversimplifications involved in this simple calculation. First, we did not take into account the lens or source redshift distributions; however, since we found that cosmology is not important for the range of lens redshifts used, and for the most extreme values of source redshift, this oversimplification is unimportant. Secondly, we did not account for the change in weighting when $\Sigma_i^{-2}$ changes from the assumed to the true cosmology, but this change simply means that the weighting scheme was not optimal, which would lead to larger error bars but no change in the result. Finally, we did not account for another change in $\Delta \Sigma$ due to the change in cosmology: since it is computed as a function of transverse comoving separation, $r = \theta_D D_\perp(z_l)(1 + z_l)$, cosmology also comes into the computation of $r$ for a given lens–source pair. The ratio $r_{\text{guess}}/r_{\text{true}}$ is non-trivially different from 1 for the two cosmologies considered, and since $\Delta \Sigma \propto r^{-2}$, rescaling $r$ leads to $\Delta \Sigma$ being off by some factor. However, this effect is only a function of lens redshift, so the rescaling will be the same for all source samples, and therefore does not affect our use of the ratio test for systematics.

Finally, it is worth noting that for futuristic surveys with lens redshifts around $z_l \sim 0.5$–0.7 and two source samples at $z_s \sim 1$ and $z_s \sim 1.5$, the differences between $f_{\text{cos}}$ for the two source samples for the cosmologies considered here is more than 1 per cent, so if the systematics are better under control, then this sort of difference may be detectable with future data sets.

### 6 RESULTS

In this section, we present results for DEEP2 redshift distributions and photometric redshift error distributions, compare the different methods of error determination and demonstrate the results of the systematics test.

#### 6.1 Redshift distributions and photometric redshift performance

Here we describe the redshift distributions and photometric redshift error distributions used for the rest of the paper. As mentioned in Section 3.3, several tests were performed using the DEEP2 redshifts. First, they were used to determine the redshift distribution of the

| $z_l$ | $z_s$ | $f_{\text{cos}}$ (flat $\Omega_m = 1$) | $f_{\text{cos}}$ (open $\Omega_m = 0.3$) |
|------|------|---------------------------------|----------------------------------|
| 0.02 | 0.12 | 0.980                           | 0.987                            |
| 0.02 | 0.70 | 0.981                           | 0.988                            |
| 0.10 | 0.20 | 0.910                           | 0.941                            |
| 0.10 | 0.70 | 0.915                           | 0.946                            |
| 0.25 | 0.35 | 0.814                           | 0.878                            |
| 0.25 | 0.70 | 0.821                           | 0.885                            |
source galaxies, and then they were used to study the photometric redshift error distributions.

6.1.1 Redshift distributions

The redshift distribution determination using DEEP2 galaxies was done by choosing a common functional form for the redshift distribution, the Γ distribution
\[ p(z) = \frac{dP}{dz} = \frac{z^{r-1}e^{-z/s}}{\Gamma(r)} \]  

(40)

This probability distribution has mean \( \langle z \rangle = az_i \), variance \( \langle z^2 \rangle - \langle z \rangle^2 = a^2 z_i^2 \) and mode \( (a - 1) z_i \). A maximum-likelihood fit was done for parameters \( a \) and \( z_i \) for \( 1 < r < 21 \) and \( r > 21 \) galaxies separately, with fit results shown in Table 8. The fit was performed by minimizing
\[ \chi^2 = \sum_i [-2 \ln p(z_i)] \]  

(41)

where the summation is over all DEEP2 matches in the appropriate magnitude range. This \( \chi^2 \) function can be minimized to give the best-fitting redshift distribution, but it does not have the same distribution as the usual \( \chi^2 \) function; its statistical properties are summarized in Appendix A. For example, goodness of fit cannot be measured in an absolute sense directly from the \( \chi^2 \), since the ‘zero-point’ is undetermined; only differences between the fits for different models are significant. Note that since the majority of the galaxies are located in the peak of the probability distribution, where \( p(z) > 1 \), the best-fitting \( \chi^2 < 0 \). Table 8 summarizes the fit results, and includes error bars both with and without LSS taken into account; LSS is taken into account as described in Appendix A, assuming angular correlation function \( w(\theta) \) from Connolly et al. (2002). Because \( w(\theta) \) may not be correct for lensing selected galaxies, this model may not perfectly capture the increase in error bars due to LSS, but should compute them to within reasonable accuracy. For the \( r > 21 \) sample, we used amplitude of the angular correlation function for the \( 21 < r < 22 \) sample from Connolly et al. (2002). For the \( r < 21 \) sample, our error bars including LSS are a worst-case estimate, where we combined the amplitudes \( A_1, A_2, \) and \( A_3 \) for \( 18 < r < 19, 19 < r < 20 \) and \( 20 < r < 21 \) sources (with fractions of galaxies \( f_1, f_2, \) and \( f_3 \)) from that work to get an overall amplitude.

Table 8. Parameters of best-fitting redshift distributions for several source samples. For both samples, \( r(z_s, a) = -0.94 \), so the fit parameters are highly correlated. Results for error bars and the variance of the \( \chi^2 \) are shown both with and without large-scale structure taken into account.

| Sample | \( 18 < r \leq 21 \) | \( r > 21 \) |
|--------|----------------|----------------|
| \( N_s \) | 162 | 116 |
| \( z_i \) | 0.100 | 0.105 |
| \( \sigma(z_s) \) | 0.012 | 0.014 |
| \( \sigma_{\text{LSS}}(z_s) \) | 0.047 | 0.017 |
| \( \alpha \) | 3.52 | 4.36 |
| \( \sigma(\alpha) \) | 0.38 | 0.55 |
| \( \sigma_{\text{LSS}}(\alpha) \) | 1.77 | 0.70 |
| \( \langle z \rangle \) | 0.35 | 0.46 |
| \( \sigma(z) \) | 0.19 | 0.22 |
| mode(\( z \)) | 0.25 | 0.35 |
| \( \langle \chi^2 \rangle \) | -117 | -43 |
| \( \langle \chi^2 \rangle_{\text{theory}} \) | -115 | -42 |
| Var(\( \chi^2 \)) | 283 | 194 |
| Var_{\text{LSS}}(\( \chi^2 \)) | 4917 | 361 |

Figure 7. The redshift distribution from COMBO-17 and DEEP2, shown separately for \( 18 < r < 21 \) and \( r > 21 \) galaxies. The lines are, as labelled, the COMBO-17 \( p(z|r) \) averaged over \( p(r) \) for all galaxies, and the best fit \( \Gamma \) distribution from DEEP2 for lensing-selected galaxies. The histograms are derived from the actual data.

\[ A = (f_1 \sqrt{A_1} + f_2 \sqrt{A_2} + f_3 \sqrt{A_3})^2 \], which assumes perfect correlation between the samples. While a perfect correlation is unrealistic, we use this result to place conservative error bars with LSS taken into account.

The resulting distributions for lensing-selected galaxies are shown in Fig. 7, along with the histogram of redshifts from DEEP2 and the COMBO-17 distribution for all galaxies. While the fits are reasonably good, it is apparent that the \( \Gamma \) distribution may not be the best choice of distribution because its shape is not well suited to matching the shape of the histogram of redshifts from DEEP2 for \( r > 21 \). However, we have not found another functional form that would be more appropriate. Note that this plot shows the actual redshift distributions; when the signal is computed, the effective source redshift distribution, which involves the weights used for each lens–source pair, is more important. These weights tend to emphasize the higher redshift portion of the curve via the inclusion of \( \Sigma_e^{-2} \), but that portion is also downweighted due to the fact that more of the higher redshift sources are detected with lower significance, and therefore have higher \( \sigma_e \). All plots of redshift distributions shown in this paper are the unweighted versions. As shown, the COMBO-17 distribution is at slightly higher redshift than the best-fitting DEEP2 distributions; the discrepancy may be attributed to an offset in the magnitudes used to determine the distribution, and the fact that the galaxies involved in its determination were not required to pass our selection criteria.

As compared to DEEP2 matches for the full photometric sample, including those that had failed shape determination and therefore were not in the source catalogue, these redshift distributions are at slightly lower redshift on average, as would be expected.

We can estimate the possible variation of the signal due to statistical error in the redshift distribution determination. This calculation can be done using the statistical error bars from the fits, and the
larger ones that include LSS, for both $r < 21$ and $r > 21$ sources. We determine error bars on $\Sigma^{-1}_c$ via propagation of the fit covariance matrix $C$ for the vector of parameters $\vec{a} = (z, \alpha)$ via

$$\sigma^2_{\Sigma^{-1}_c} = \left( \frac{d\Sigma^{-1}_c}{d\vec{a}} \right)^T \cdot C \cdot \left( \frac{d\Sigma^{-1}_c}{d\vec{a}} \right).$$  \hspace{1cm} (42)

For this computation, we use two-sided derivatives with respect to $z$ and $\alpha$ evaluated at separations $\pm 0.1r$ around the best-fitting parameters $d_{\delta z}$. We can then compute the calibration uncertainty on the signal, $\delta (\Delta \Sigma) / \Delta \Sigma = -\delta (\Sigma^{-1}_c) / \Sigma^{-1}_c$ for a given lens redshift. Results for this uncertainty in the signal due to the redshift distribution statistical uncertainty is larger for higher $z_1$ since those lenses are closer to the source redshift distribution; for lenses at $z_1 = 0.03$, the majority of the distribution is essentially at $z = \infty$, so $d\Sigma^{-1}_c / d\delta$ is small. Furthermore, the errors on the distribution at $r < 21$ have a greater effect than those on the distribution at $r > 21$ because the brighter sources are, on average, closer to the lenses. As shown, the errors for the 18 $< r < 21$ sources with LSS are quite large (large enough for $z_1 = 0.2$ that we may worry about the accuracy of this linear calculation), but that is due to the fact that we used a very conservative estimate of $w(\theta)$ for this sample as described above. Consequently, those error bars are very conservative, and in reality are most likely significantly smaller and closer to the error bars without LSS. Nonetheless, we can see that using the redshift distribution for $r < 21$ may be a bad idea because errors in its determination may have a large effect on $\Delta \Sigma$.

### 6.1.2 Photometric redshift errors

In order to be able to use photometric redshifts to eliminate physically associated pairs, we also computed photometric redshift error distributions using the matches from DEEP2. Of the 278 matches, 162 (58 per cent) were at $r < 21$ (with 9 of those, or 6 per cent, having failed photometric redshift determination). Because high-redshift LRGs have different properties from the overall photometric sample, we explicitly exclude them from this comparison, leaving 135 galaxies with which to study photometric redshifts.

First, we define the terminology used in this section. For a set of true redshifts $z$ and photometric redshifts $z_p$, we define the photometric redshift error as $\delta = z - z_p$, and can construct average distributions $p(\delta | z)$ and $p(\delta | z_p)$ in bins. The statement that the photometric redshifts are biased means that

$$\int p(\delta | z) \delta \delta \delta \neq 0.$$  \hspace{1cm} (43)

For our sample of 135 redshifts at $r < 21$, we find that overall there is a small photometric redshift bias, where the overall bias is 0.04, with $\langle z \rangle = 0.34$ and $\langle z_p \rangle = 0.30$. When we look at the subsample of 64 galaxies below the mean redshift, the bias is zero to two significant figures and the scatter is 0.11. When we restrict to the subsample of 71 galaxies above the mean redshift, the mean bias is 0.08 and the scatter is 0.17.

However, the calculation that is actually more relevant for the purpose of this paper is the ‘conditional bias’,

$$\langle \delta \rangle = \int p(\delta | z_p) \delta \delta \delta.$$  \hspace{1cm} (44)

In particular, even if the bias is zero, then for a redshift distribution that peaks at redshift $\langle z \rangle = 0.34$, we will have $\langle \delta \rangle > 0$ for $z_p < \langle z \rangle$ and $\langle \delta \rangle < 0$ for $z_p > \langle z \rangle$ simply due to the scatter in the photometric redshifts. We also define a ‘conditional scatter’,

$$\sigma_\delta = \sqrt{\int p(\delta | z_p) (\delta - \langle \delta \rangle)^2 \delta \delta \delta}.$$  \hspace{1cm} (45)

Table 9 shows the mean conditional bias and scatter as a function of magnitude; Table 10 shows the same information as a function of photometric redshift. Information is included for $r > 21$ in Table 10 for informational purposes, but since we did not use photometric redshift information for those sources due to the large bias and scatter, they are not included in the calculations for Table 11. A likely explanation for the increase in $\sigma_\delta$ at fainter magnitudes is the larger photometric errors that make photometric redshift determination more difficult. The fact that our test of this program on the SDSS spectroscopic sample at $r < 18$, on galaxies with negligible photometric errors, shows $\langle \delta \rangle = -0.003$ and $\sigma_\delta = 0.048$, with $\sigma_\delta$ increasing with $r$ magnitude if the sample is split into subsets, supports this hypothesis. Also, because we require $z_1 > 0.02$ and $z_1 > z_1 + 0.1$, we only use photometric redshifts greater than 0.12 in our analysis, though error information has been included in the tables.

### Table 9. Estimated calibration uncertainty on $\Delta \Sigma$ for two values of $z_1$ and $p(z)$ computed for $r < 21$ and $r > 21$ sources. The uncertainties are shown both with and without LSS taken into account.

| $z_1$ | $\delta (\Delta \Sigma)$ | $\delta (\Delta \Sigma)$ |
|-------|-----------------|-----------------|
|       | (18 < $r < 21$) | ($r > 21$)       |
| 0.03  | without LSS     | ±0.007          |
| 0.03  | with LSS        | ±0.004          |
| 0.20  | without LSS     | ±0.053          |
| 0.20  | with LSS        | ±0.25           |

### Table 10. Conditional bias $\langle \delta \rangle$ and scatter $\sigma_\delta$ as a function of $r$ magnitude for non-LRGs with successful photometric redshift determination.

| Magnitude range | $\langle r \rangle$ | $N_{gal}$ | $\langle \delta \rangle$ | $\sigma_\delta$ |
|-----------------|------------------|----------|----------------|----------|
| $z_p < 0.12$     |                  |          |                |           |
| $z_p > 0.12$     |                  |          |                |           |

© 2005 RAS, MNRAS 361, 1287–1322

Downloaded from https://academic.oup.com/mnras/article-abstract/361/4/1287/1117325
by guest
on 27 July 2018
Figure 8. The redshift distribution for \( r < 21 \) non-LRG sources with successful photometric redshift determination, shown computed with spectroscopic redshifts and with photometric redshifts, with and without error distributions taken into account.

for those less than 0.12. Our finding that photometric redshifts are, on average, biased low is in accordance with the same finding in H04 (based on the fractions of physically associated pairs found using \( z_p < z_l - 0.1 \)). To illustrate the effects of photometric redshift errors, Fig. 8 shows the redshift distribution for the \( r < 21 \) sample computed using photometric redshifts, with and without the errors taken into account according to equation (25), and using the spectroscopic redshift distributions from DEEP2 directly. As shown, the photometric redshift distribution has some strange features, including the peak near \( z = 0 \) and at \( z = 0.38 \); the distribution with the errors taken into account looks significantly more reasonable, and is an excellent match for the histogram of redshifts from DEEP2. In fact, the distribution with photometric redshift errors taken into account appears to be a better match than the best-fitting \( \Gamma \)-distribution. (Any discrepancies between the distributions in Figs 7 and 8 arise from the fact that Fig. 7 includes LRGs in the \( r < 21 \) sample, but Fig. 8 does not.) While we may have been concerned about using so few galaxies to determine the error distribution as a function of photometric redshift (by dividing 135 galaxies into five bins), it is apparent that even with so few galaxies, enough of the error distribution can be determined that this procedure is reasonably successful.

While we have determined the photometric redshift error distributions, we are still far from an understanding of how they affect the lensing signal. There are two effects that must be included.

(i) The conditional bias, \( \delta \), must be included in our computation of \( \Sigma_c^{-1} \). Because \( \delta > 0 \) for low \( z_p \), the corrected redshift is larger than \( z_p \); raising \( \Sigma_c^{-1} \) (lowering \( \Delta \Sigma \)) once we make this correction. This effect is particularly important at low values of \( z_p \), because \( \Sigma_c^{-1} \) is varying more rapidly with source redshift at low redshift. (However, since \( \Sigma_c^{-1} \) is lower at low source redshifts, these sources get less weight, so the error they cause is not as high as one might naively expect.) At high redshift, \( \delta < 0 \), so the correction will lower the redshift and \( \Sigma_c^{-1} \) (raise \( \Delta \Sigma \)), but because \( \Sigma_c^{-1} \) varies so slowly with redshift in that regime, we expect that this will only be a minor correction. The majority of the photometric redshifts used are around 0.35–0.4, where the conditional bias goes to zero, so we may expect that its effects will not be too large.

(ii) The conditional scatter \( \sigma_\delta \) must also be included. At photometric redshifts above \( \sigma_\delta \) above the lens redshift (i.e. for most sources in this work, due to our requirement that \( z_p > z_l + 0.1 \)), because \( \Sigma_c^{-1} \) is an increasing function of source redshift with negative second derivative in this regime, the inclusion of the width of the error distribution will lower the estimate of \( \Sigma_c^{-1} \) (raise \( \Delta \Sigma \)). This effect is counter to the effect of the conditional bias \( \delta \) at low \( z_p \), and in the same direction as it at high \( z_p \). However, at low \( z_p \), close to \( z_l \), the effect of the \( \delta \)-function in the second derivative at \( z_l = z_l \) is that \( \Sigma_c^{-1} \) is actually higher than it would have been without the scatter, i.e. this effect goes in the same direction as the bias at very low photometric redshift.

Fig. 9 shows a plot of \( \Sigma_c^{-1} \) for lenses at several redshifts, as a function of source redshift, for several different possibilities: straight acceptance of source photometric redshift, use of the photometric redshift corrected for the conditional bias, and corrected for the full photometric redshift error distributions (actual, and best-fitting). The fit for the distribution was done by assuming it is a Gaussian with a fixed width but with a mean that is a linear function of \( z_p \). This form was assumed because it is a simple parametrization that can still account for the fact that the error distribution changes sign as we go from low to high \( z_p \). The effects of the mean correction (which raises \( \Sigma_c^{-1} \) for \( z_p < 0.4 \) and lowers it for \( z_p > 0.4 \)) and the width of the error distribution can clearly be seen in this plot.

Vertical lines show our cutoff of \( z_p = z_l + 0.1 \); fortunately, as shown, this limitation on the photometric redshifts means that even just using photometric redshifts directly, without accounting for the errors, is less likely to cause errors. Furthermore, as shown, the errors in \( \Omega_c^{-1} \) for \( z_p \sim 0.1 \) are downweighted in importance due to...
our weighting scheme, with weights \( \propto \Sigma^{-2} \). Consequently, we can expect to see some cancellation, whereby the large increase in \( \Sigma_{\ wandered}^{-1} \) (decrease in signal) for small \( z_p - z_i \), which has small weight due to the weighting scheme, is cancelled by the modest decrease in \( \Sigma_{\ wandered}^{-1} \) (increase in signal) for large \( z_p \) and \( z_i \), which get considerable weight. The extent of the cancellation will be shown in Section 6.3.

The difference in results between the best-fitting Gaussian error distribution versus the actual error distribution is quite small, except for the range \( \Delta z \approx z_i \) that is not used in this work; probably, the best-fitting distribution gives higher values of \( \Sigma_{c}^{-1} \) in this range because the fit distribution underestimates the tails that are present in the actual distribution, and the tails help to lower \( \Sigma_{c}^{-1} \) with the actual error distribution for \( z_p \approx z_i \). This result implies that when restricting to the range \( \Delta z > z_i + 0.1 \) as in this work, it is unimportant whether or not the method of parametrizing the photometric redshift error distribution is able to account for the tails, as long as the basic shape (mean and width) is correct. However, for analyses that include lower values of \( z_p \), a more careful treatment of the error distribution is necessary.

It is clear from these results that for different limits on source photometric redshift (e.g. without requiring \( z_p > z_i + 0.1 \)), the results of photometric redshift errors will vary, and must be re-evaluated if we choose to relax this restriction.

As illustrated by this discussion, the effect of photometric redshift errors is complicated, with several possible effects that will tend to push the lensing signal in opposite directions, and it is difficult to anticipate which effects will turn out to be more important than others for any given lens and source sample. Consequently, we must turn to the systematics test described in Section 5 to determine the effects of photometric redshift errors on \( \Delta \Sigma \).

We also must consider the effects of errors in the photometric redshift error distribution. Statistical (Poisson) error can be determined via a bootstrap resampling of the DEEP2 matches. The effects of LSS and systematic error are more complex, because correlations between photometric redshift errors are more difficult to model than the correlations between the redshifts used to derive \( p(z) \). There are several concerns: statistical errors on photometry; systematic errors on colours and deviations of galaxy SEDs from templates. For example, if the colours of galaxies depend on environments, we may expect different photometric redshift error distributions for field galaxies versus those in groups or clusters.

### 6.2 Error comparison

Because we used three methods to compute error estimates on the signal (the analytic expression for the error, the random catalogue errors and the bootstrap errors), we compare the results of the three methods to ensure that they gave comparable results. In general, the results of the three methods were quite comparable, with the exception of two situations. First, for the lower luminosity bins, the bootstrap and random catalogue errors were larger than those computed analytically by 15–20 per cent. Secondly, for the higher luminosity bins, the radial bins that required a sizable boost had larger errors computed using the bootstrap than via the other methods. Because there is some uncertainty on the boost factors (which was also computed using the bootstrap), the uncertainty on the product \( B(\gamma) \Delta \Sigma(r) \) is approximately \( B(\gamma) \Delta \Sigma(r) \sqrt{\sigma_{\beta}^2 + \sigma_{\Sigma}^2} \) when computed using the bootstrap, raising the fractional error by roughly 5 per cent at small pair separations. The larger errors from the bootstrap are more likely to be correct since they take into account errors on the boost factor itself, so these are the errors that were used for the computation of ratios \( R_{i,a,b} \).

The analytic errors are essentially noiseless since they are computed over millions of lens–source pairs. The random catalogue errors had the most noise, since they were computed with only 24 random catalogues due to the computational expense involved. If the error amplitude \( E(r) \) is fit to a power-law \( Ar^{-1} \), and the noise amplitude is defined as the fractional difference between the actual errors and the best-fitting value, then the random catalogue errors typically had a noise amplitude of 15 per cent. The bootstrap errors have little to no noise, with a maximum noise amplitude of 5 per cent even for L1 and L6 (which had the fewest lenses) and 1 per cent or lower for the others.

We also checked the covariance between the signal in different radial bins using the bootstrap for all luminosity bins and source samples. We can anticipate that these correlations may become important on angular scales for which the lens–source separation is comparable to the lens–lens separation, since on those scales a given source will contribute to the calculation for more than one lens. We find a typical lens–lens separation to be 7 arcmin, which corresponds to comoving transverse separation of 200 \( h^{-1} \) kpc in L1, and as high as 1100 \( h^{-1} \) kpc in L6. Using the bootstrap resampling for 2500 data sets, we find that for L1, there are correlations as high as 0.2–0.3 starting at radii of 800 \( h^{-1} \) kpc between nearby radial bins (generally the three nearest radial bins in either direction) but little covariance between the bins for lower radii. The bootstrap procedure naturally introduces noise with standard deviation \( 1/\sqrt{M} \) (\( M \) is the number of bootstrap regions) in the correlation coefficients when they are \( \ll 1 \) (the noise is less for higher correlations), so the statement that there is little covariance means that the correlation coefficients were consistent with being drawn from a Gaussian distribution \( N(0, 1/M) \).

For L6, no covariance is observed within the noise, even for the outermost bins that should be most correlated, since these correspond to a smaller angular scale than for the fainter (more nearby) lenses. Note that our procedure described in Section 5 automatically takes any covariance between the radial bins into account because it uses the bootstrap.

We also found that our measure of the signal in a given bootstrap-resampled data set, \( \Delta \Sigma \) as defined in Section 5, has a significant covariance between luminosity bins. This result is unsurprising considering that many of the same sources contribute to the result in different luminosity bins, though the weighting may differ significantly. We found correlations between the results in different luminosity bins as high as 0.30 for adjacent luminosity bins, 0.10 for luminosity bins that were separated by one (e.g. L4 and L6) and consistent with zero for further separations. This kind of correlation must be taken into account (as it is with the bootstrap) for any analysis that utilizes the results in several luminosity bins simultaneously.

As an additional test, we also examined the distribution of values of \( \xi \) for each luminosity bin and source sample, computed from the 2500 bootstrap-resampled data sets. This distribution was found to be Gaussian to high accuracy. The Gaussianity is quantified via calculation of the skewness

\[
x_3 = \frac{\langle (\Delta \Sigma - \langle \Delta \Sigma \rangle)^3 \rangle}{(\text{Var}(\Delta \Sigma))^3} \]

and kurtosis excess

\[
x_4 = \frac{\langle (\Delta \Sigma - \langle \Delta \Sigma \rangle)^4 \rangle}{(\text{Var}(\Delta \Sigma))^2} - 3,
\]

both of which should be zero for a Gaussian distribution within the noise. For \( M \) bootstrap subregions and \( N \) resampled data sets, the expected variance of the skewness is \( \text{Var}(x_3) = 6(M^{-1} + N^{-1}) \approx 6/M \) (\( M = 250 \) and \( N = 2500 \) in our case) and of the kurtosis is
Var(s₄) = 24(M⁻¹ + N⁻¹) ≈ 24/M (Cramer 1946). The values of s₃ and s₄ calculated for a number of different sources samples and luminosity bins were consistent with being drawn from random distributions with mean zero and approximately those variances, and therefore consistent with Gaussianity.

6.3 General systematic tests

In this section, we present results of tests for systematics described in Section 4.

6.3.1 Random points test

As mentioned in Section 4.1, we found a non-zero signal computed around random points in the survey region, indicating the presence of a systematic shear γsys in the source catalogue that must be accounted for. The shape and amplitude of this signal varies with luminosity bin and source sample, because it is related to angular scale (and therefore different transverse separations for lens samples at different redshifts) and likely varies with source R2 as well.

In principle, if the systematic shear is uniform across the survey, then it will only be revealed on large scales. If the distribution of sources around lenses is circularly symmetric, then the contribution of the systematic shear to the signal averages out, so only edge effects on large scales will show the presence of systematic shear. The effect of this spurious ellipticity is most noticeable for the narrow southern stripes, of approximate width 2.5°; our decision not to use them for the work in this paper reduced the observed random catalogue signal by approximately a factor of 4. However, it is unlikely that the systematic shear is perfectly uniform, and if it fluctuates on small scales then we may also see a random catalogue signal at small transverse separations.

As an illustration, Fig. 10 shows the mean random catalogue signal for all six luminosity bins, computed with r > 21 sources. As shown in this figure, there is a clear signal from the random catalogues at large radius; the statistical error due to this signal can be taken into account by the bootstrap. For the brighter sources, r < 21, the random catalogue signal has slightly different shape, and is shown in Fig. 11. As shown there, the signal is zero at low transverse separation, then becomes positive in the fainter bins, decreasing to zero at larger radii. While this occurs at different radii for L1–L3, the corresponding angular scale is same: roughly 25 arcmin. This number is of significance in the SDSS as the separation between adjacent camcols; however, the reason it is showing up here is not entirely clear. For the characteristic redshift of L4, 25 arcmin corresponds to 2.2 h⁻¹ Mpc, which is why the signal does not go to zero in that bin; however, it is clearly declining by the maximum radius of 2 h⁻¹ Mpc.

If the systematic shear is constant across the survey, then one way to eliminate or at least reduce it would be to only keep lenses that have a reasonably circular distribution of sources around them, since the systematic shear along the scan direction will then cancel out of the calculation of ΔΣ. We implemented this cut, requiring the ellipticity of the distribution of sources around the lens to be less than 0.2, but found that this cut did not significantly decrease the random catalogue signal; consequently, we conclude that the systematic shear fluctuates on a scale smaller than the typical distribution of sources around a lens, and did not use this cut for the work in this paper.

6.3.2 45° test

In this subsection we describe results of the 45° test described in Section 4.2. This test was performed using the bootstrap with 2500 subsamples. In order to check for a systematic dependent on transverse separation, we did the test in three radial bins: 30–100 h⁻¹ kpc, 100–600 h⁻¹ kpc and 600–2000 h⁻¹ kpc, as well as for the full range 30–2000 h⁻¹ kpc. The test was done for each lens luminosity sample separately, and the results were also checked for the combined result averaged over lens luminosity.

Note that the 45° random catalogue signal was non-zero for all lens–source subsamples, and therefore had to be subtracted just
Table 12. $\Delta \Sigma_{45}$, its standard deviation $\sigma_{45}$, and the probability to exceed the measured value by chance if it is truly consistent with zero, for all source samples and radial ranges. The results are shown for the signal averaged over lens luminosity. Because the $r > 21$ sample includes some LRGs, the results shown for those two samples are not independent; the results for all source samples combined take their covariance into account.

| Radial range (h$^{-1}$ kpc) | $\Delta \Sigma_{45}$ (h M$_{\odot}$ pc$^{-2}$) | $\sigma_{45}$ | p-value |
|-----------------------------|------------------------------------------|-------------|----------|
| $r < 21$ sources, without LRGs |                                         |             |          |
| 30 $< r < 100$              | 1.70                                     | 2.45        | 0.49     |
| 100 $< r < 600$             | $-0.03$                                  | 0.45        | 0.96     |
| 600 $< r < 2000$            | 0.38                                     | 0.24        | 0.12     |
| 30 $< r < 2000$             | 0.33                                     | 0.22        | 0.13     |
| $r > 21$ sources             |                                         |             |          |
| 30 $< r < 100$              | $-1.72$                                  | 3.04        | 0.57     |
| 100 $< r < 600$             | $-1.17$                                  | 0.55        | 0.04     |
| 600 $< r < 2000$            | $-0.13$                                  | 0.25        | 0.62     |
| 30 $< r < 2000$             | $-0.25$                                  | 0.23        | 0.26     |
| LRGs                        |                                         |             |          |
| 30 $< r < 100$              | 6.37                                     | 5.31        | 0.23     |
| 100 $< r < 600$             | $-0.17$                                  | 0.91        | 0.85     |
| 600 $< r < 2000$            | $-0.21$                                  | 0.45        | 0.04     |
| 30 $< r < 2000$             | $-0.19$                                  | 0.42        | 0.66     |
| All                         |                                         |             |          |
| 30 $< r < 100$              | 1.44                                     | 1.90        | 0.45     |
| 100 $< r < 600$             | $-0.47$                                  | 0.34        | 0.17     |
| 600 $< r < 2000$            | 0.09                                     | 0.18        | 0.63     |
| 30 $< r < 2000$             | 0.02                                     | 0.17        | 0.88     |

Figure 12. 45° rotated signal for each of the three source samples, averaged over all luminosities. Many radial bins are averaged to reduce the noise and make the plot easier to understand.

6.2.4 Seeing dependence of calibration

As described in Section 4.4, we also used the systematics test to compare the signal computed with sources that had PSF sizes greater than and less than the median. The results of the test using L3–L6 lenses are shown in Table 13. As shown there, the signals are statistically consistent with each other, indicating that we do not have to worry about seeing-dependent calibration of the shear.

6.3.5 $R_2$ dependence of calibration

As described in Section 4.5, we used the systematics test to compare the signal computed with sources with $R_2 < 0.55$ versus with $R_2 > 0.55$. This exercise was done in both the r and i bands separately, and results are shown in Table 13. As shown there, there is no evidence at the 2σ level for a systematic calibration offset that depends on resolution factor $R_2$.

© 2005 RAS, MNRAS 361, 1287–1322
This result is important because recent tests have indicated that the source catalogue used in H04 did have an $R_2$-dependent calibration, with the calibration changing by as much as 50 per cent from the lowest $R_2 = 1/3$ to the highest. When averaging over all $r < 21$ sources, this effect caused the calibration for that paper to be too low by approximately 15 per cent (only $r < 21$ sources were used for that work; since the LRG sample and the $r > 21$ sample is, on average, at lower $R_2$, the offset is even worse for those samples). Tests to determine the source of the problem were not definitive, but seemed to rule out selection biases. The basic conclusions of that paper, namely the 99.9 per cent stat+sys confidence intervals on the intrinsic shear, are essentially unaffected by this problem, since we assumed up to 20 per cent shear calibration bias when computing the confidence intervals. However, this finding highlights the importance of our results here, which is that this catalogue does not have a statistically significant $R_2$-dependent calibration.

### 6.3.6 Systematic differences between bands

While for most tests we used the shape measurements averaged over both $r$ and $i$ bands, we also compared the signal computed using only one band versus the signal computed using the other. This test was done separately for $r < 21$ and $r > 21$ sources. As shown in Table 13, there is no sign of a systematic discrepancy between the signal computed using the shape measurements in either band individually, so to improve statistics we will henceforth use the shape measurement averaged over bands.

### 6.3.7 Boosts

The boost factors $B_{\alpha,\beta}(r)$ generally follow expected trends, decreasing with radius and increasing with luminosity. Furthermore, they are largest with the $r < 21$ sources, which have the largest overlap in redshift range with the lens sample; less with the $r > 21$ sources; and the smallest with the LRG sample, since those are chosen specifically to avoid contamination from low-redshift ($z < 0.35$) galaxies.

A systematic effect at the smallest radii ($r < 30 h^{-1}$ kpc) is revealed by the boost factors. While they should, in principle, be monotonically decreasing, we found that for the brightest lens samples, they actually increase from 20 to 30 $h^{-1}$ kpc; then follow the expected trend of decreasing for $r > 20 h^{-1}$ kpc. In H04, this trend appeared for even larger radii; however, for that work, our source catalogue did not include deblended children, so this result is expected. This source catalogue does include deblended children, so we do not expect a loss in lens-source pairs for small transverse separation due to the loss of deblended children. However, very bright galaxies, for which this effect is noticeable, are large enough that at those separations, their light could make it impossible to detect much smaller, faint galaxies. This could explain the fact that this effect is more noticeable for the fainter $r > 21$ sources than it is for the $r < 21$ sources. In order to avoid strange selection effects due to this problem, this work only uses $r > 30 h^{-1}$ kpc for the analysis.

Plots of the boost factor are shown in Fig. 13. The top plot shows the trends of the boost with luminosity bin for the $r > 21$ sources;
A close examination of the boost factors reveals a systematic effect at small transverse separations that affects any analysis of galaxy–galaxy weak lensing done using SDSS data. From plots of the boost factor at small transverse separation computed using LRGs as sources, it became apparent that there is a deficit in lens–source pairs below a certain radius. This deficit manifests itself in the boost factor actually dipping below 1 (i.e. $\xi < 0$), which is not physical. This deficit also is present for the $r < 21$ and $r > 21$ samples, though it is harder to notice because of the fact that there are physically associated pairs at the affected radii. We determined that this effect is due to either an instrumental effect or something in the PHOTO pipeline associated pairs at the affected radii. We determined that this effect is due to either an instrumental effect or something in the PHOTO pipeline (rather than due to some selection effect in our catalogue, error in signal computation, or actual physics) due to the following factors: (i) it occurs at a particular angular (not physical) scale, most noticeably around 50 arcsec though it extends up to about 90 arcsec, (ii) the effect is noticeable with two different catalogues, the one from this work and from H04, which were processed with completely independent software pipelines from the PHOTO outputs, (iii) it is most noticeable for lenses that are bright in apparent magnitude, (iv) it is most noticeable for low surface-brightness sources and (v) we see the effect even if we compute the signal around bright stars instead of around galaxies. While a hint of this effect can be seen for $B(r)$ from LRGs in Fig. 13, it is more obvious in Fig. 14, which shows a plot of the correlation function between bright ($r < 19$) stars and high-redshift LRGs, where there is a non-zero correlation function at small separations because of low-redshift quasar contamination in the star sample. This plot was computed using only low surface-brightness sources (those with a value below the median), and is shown as a function of star magnitude.

The effect turns out to be due to a problem in determination of the sky level around bright objects (stars or galaxies). The sky level is determined in overlapping boxes of size 256 pixel (approximately 100 arcsec) and linearly interpolated on the 128 pixel (50-arcsec scale), using a $2.3\sigma$ clipped median.\(^5\) We expect the sky level near bright lenses (within about 1 arcmin) to be higher than the global sky because light from the lens galaxies extends out to a large angular separation, and we do not want that extra light to be included in the flux from the sources in that region. So, we do want the sky level to be influenced by the presence of the bright objects. However, our results suggest that the sky level determination is overly influenced by the presence of bright objects, with the sky level being set too high near them. The problem is different for each band. This change in the sky level has several effects: (i) some sources are not detected at all by the PHOTO software, (ii) some are detected, but due to the faulty sky level, the magnitudes, colours, and shape measurements determined by model fits to the light profile are off, and therefore they are not included in our source catalogue because of magnitude or $R_2$ cuts or OBJC_TYPE failure, or are included but with these quantities computed wrong and (iii) our computed redshift distribution of these sources is likely wrong because of the incorrect fluxes. Furthermore, because this effect can vary from band to band, the colours are incorrect as well, affecting both photometric redshift determination and selection of the high-redshift LRG sample.

To test that this effect is indeed the cause of the observed effect, we used the PHOTO reductions to compute the averaged sky level as a function of angular separation from bright galaxies (we except the effect to differ around stars versus galaxies, because galaxies are extended and therefore affect more pixels). This computation was done in each band, and we verified that the sky flux level $f_{\text{sky}}$ around bright galaxies is indeed higher than the baseline value computed around faint galaxies, by a quantity we call $\Delta_{\text{sky}}$ in each band. For such small angular separations, we expect that $f_{\text{sky}}$ should be higher than that baseline, so much of the rise at small angular separations is not spurious. Fig. 15 shows a plot of the average sky flux in the $r$ band as a function of angular separation from lens galaxies with apparent magnitudes as labelled on the plot. We can make a naive calculation of the effect of this problem (assuming that all of $\Delta_{\text{sky}}$ is spurious and that the only effect was to change the surface brightness of the faint source without changing its size) at 40 arcsec for a circular source galaxy 4 arcsec in diameter at magnitude 21.5. The flux for such an object is 2.5 nanomaggies, and since its area is 12.5 arcsec\(^2\), then if the lens is at $r \sim 16$, with $\Delta_{\text{sky}} = 0.4$ nanomaggies per square arcsecond, its true flux should have been 7.5 nanomaggies, a full factor of 3 higher, giving it a true magnitude of 20.3, a highly significant offset from the observed value. For lenses around $r \sim 17.5$, a large fraction of the sample, $\Delta_{\text{sky}}$ is smaller, but we still expect a ‘true’ magnitude of 21. Of course, this is just a crude estimate, and a more careful calculation of what happens to the fits to the light profiles when the sky level is changed is necessary. This calculation is actually quite complicated because the change in sky flux changes the effective radius of the source from the fits, so not only is the flux underestimated within the area of the galaxy, light is also lost due to the underestimated size. However, when we tried applying this naive correction, we found that we overcompensated for this effect, so it seems clear that some of $\Delta_{\text{sky}}$ is not spurious (as would be expected for such large, bright lenses, for which some of the light does indeed extend to these radii).

There are several implications of this effect for our work. First, while Fig. 14 shows what looks like a dip in the boost factor below 90 arcsec and most prominently at 50 arcsec, with it then rising

\(^5\)For more information about the sky level determination, see http://www.sdss.org/dr3/algorithms/sky.html.

---

**Figure 14.** Correlation function between bright stars and the high-redshift LRG sample, for low surface-brightness LRGs only, as a function of star apparent magnitude.
In future work, we will attempt a more detailed modelling of this problem. For most of our results, we do not attempt to correct the boost factors, which are clearly underestimated, though we will show what happens to the results if we try a naive correction; since only small separations are involved, the change in the signal averaged over radius due to this effect is small. Furthermore, while we have estimated rather large values for the magnification bias at small separations for the \( r < 21 \) and LRG samples, this sky level problem makes it effectively impossible to make any detection of magnification bias with the SDSS using MAIN spectroscopic sample lenses and working at angular separations of less than approximately 90 arcsec.

The final test performed in this section was to compare the signal computed for the bright and faint halves of the brightest bin (L6) separately and then averaging them (i.e. taking the average value of \( B(r)[\Delta \Sigma(r) - \Delta \Sigma_{\text{rand}}(r)] \) versus the signal computed by averaging over the whole luminosity bin, as we usually do, to get the average values of \( B(r) \) and \( \Delta \Sigma(r) - \Delta \Sigma_{\text{rand}}(r) \)). The concern is that due to the non-uniformity of the boost factor, with both the boost factor and signal rapidly increasing in this bin, our averaging process may bias the signal. Our tests indicate that for the both \( r < 21 \) and \( r > 21 \) sources, the effect of this averaging is less than a 1 per cent effect even for L6; consequently, it must be negligible for the other bins, for which the signal and boosts are varying less rapidly.

### 6.3.8 Intrinsic alignments

When computing the expected contamination due to intrinsic alignments, we must do the computation separately for each luminosity bin \( i \) and source sample \( \alpha \) because the contributions may vary with lens mass (and therefore luminosity) and with source sample (more distant source samples will have fewer physically associated lens–source pairs). Furthermore, the procedure should be done as a function of radius, since limits on the intrinsic shear \( \Delta y_{\text{int}} \) vary with radius. We split the data into three radial ranges \( j \) with limits \( r_{\text{min},j} \) and \( r_{\text{max},j} \). This estimate assumes that the boost factors \( B(r) \) are entirely due to physically associated pairs, and ignores the magnification bias estimates in Table 6.

The signal contamination due to intrinsic alignments can be computed as follows. First, we determine the fraction of lens–source pairs that are physically associated. For each lens sample \( i \), source sample \( \alpha \) and radial bin \( j \), we compute

\[
 f_c^{i(\alpha)}(r_j) = \frac{\sum_{i=1}^{r_{\text{max},j}} |B_{\alpha,i}(r_j) - 1|A(r_j)\eta^{i(\alpha)}_{\text{rand}}(r_j)}{\sum_{i=1}^{r_{\text{max},j}} A(r_j)\eta^{i(\alpha)}_{\text{rand}}(r_j)}, \tag{46}
\]

where \( A(r) \) is the area of the annular bin centred at \( r \), and the value of \( f_c \) is determined via the bootstrap.

Then, for a given radial range, we can compute the estimated contribution of intrinsic alignments to the lensing signal via

\[
 \Delta \Sigma_{\text{int}}^{i(\alpha)}(r_j) = f_c^{i(\alpha)}(r_j) \frac{\Delta y_{\text{int}}^{i(\alpha)}(r_j)}{(\Delta \Sigma)^{-1}}. \tag{47}
\]

In order to avoid excessively low (or even non-physical negative) values of \( f_c \), we had to correct for the sky level effect that suppresses \( B(r) \) at small radii. This correction was done as follows. For the LRG sample, for which this effect is most noticeable, we examined \( B(r) \) for all lens samples. We found the minimum value \( B_{\text{min}} \) at \( r = r_{\text{min}} \), and assumed that the effect was actually the same at all \( r < r_{\text{min}} \) (and could not be observed because \( B(r) > 1 \) for those radii). So, for \( r < r_{\text{min}} \), we multiplied the signal and \( B(r) \) by \( 1/B_{\text{min}} \); for \( r > r_{\text{min}} \),

---

We thank Ryan Scranton for providing the necessary masks and mask utilities to allow us to perform this test.
we multiplied the signal and \(B(r)/B(r)\) if \(B(r) < 1\), and by 1 otherwise. This correction is very rough, since it assumes that the sky level effect is the same for all source samples, and that the sky level effect only changes the boosts without having an effect on the shear or redshift calibration. However, our estimate of contamination from intrinsic alignments in Table 14 is still conservative, because we ignore the fact that magnification bias contributes to the boost factors at small radii. For the purpose of this calculation, the 99.9 per cent stat+sys confidence intervals from H04 have been reduced by a factor of 0.6 so they will be 2\(\sigma\) rather than 3.4\(\sigma\) bounds.

For this calculation, we assumed that \(\Delta \gamma_{\text{int}}\) for the \(r > 21\) and LRG samples is the same as that for the \(r < 21\) sample. While this assumption is not fully justifiable, and means that the actual estimates of the contamination should not be taken too seriously, the size of the confidence intervals are likely still trustworthy.

Here we discuss the results in Table 14. The entries with \(\pm \infty\) as bounds are those for which H04 could not place any constraint on \(\Delta \gamma_{\text{int}}\). As shown, because of the poor constraints at low luminosities (due to the fact that the contamination fractions can be very low), errors on weak lensing signals are due to intrinsic alignments placed in H04. However, the fact that the average signal (\(\Delta \Sigma\)) for these ranges is entirely reasonable suggests that the contamination by intrinsically associated pairs is not causing a large problem. For \(r < 21\), use of \(p(z)\) rather than photometric redshifts leads to very large contamination fractions, and therefore very loose constraints on \(\Delta \Sigma_{\text{int}}\). The fact that the constraints on intrinsic alignments for \(r < 21\) and LRG samples are quite different, generally tighter on LRGs and looser on \(r > 21\), yet the results agree quite well (as will be shown in Section 6.4.4) suggests once again that we do not have a major problem with contamination by intrinsic alignments. For small radii, the \(r < 21\) (with \(z_p\)) and \(r > 21\) samples, we assign 2\(\sigma\) uncertainties due to intrinsic alignments of \(\pm 20\) per cent for L3–L5 and \(\pm 60\) per cent for L6; for LRGs, \(\pm 5\) per cent for L3–L5 and \(\pm 15\) per cent for L6. For large radii, the \(r < 21\) (with \(z_p\)) sample, we assign 2\(\sigma\) uncertainties of \(\pm 5\) per cent for L3–L5 and \(\pm 15\) per cent for L6; and negligible for LRGs. These constraints were assigned by comparing the 2\(\sigma\) constraints on \(\Delta \Sigma_{\text{int}}\) to the characteristic \(\Delta \Sigma\) in the table.

6.3.9 Corrections for non-volume limited sample

As described in Section 4.9, we must consider the effects of differences in the mean weighted luminosity of the subsample when computed with different source samples or redshift distributions for the same sample. Using a simple model for \(\Delta \Sigma(L)\), we can compute the expected change in the signal from equation (33), and expected

\[ R_{\alpha, \beta} = 1 + \frac{\delta(\Delta \Sigma)}{\Delta \Sigma}. \]  

(48)

We have no need for these corrections for L1 and L2, since for those samples all source samples are essentially at infinite redshift.
6.4 Redshift systematics tests

In this section, we describe our application of the systematics test to assess which methods of redshift distribution determination are optimal. First, for each group of sources (r < 21 without LRGs, r > 21 and LRGs), we compared the various methods of determining the redshift distribution. Next, we compared the signal determined using each of these three samples, to make sure that our results are consistent. In each case, this test was done individually for each luminosity bin, because we can expect that different lens redshift samples will sample different parts of the source redshift distribution, and therefore results may differ systematically across the luminosity bins. However, we also use the results averaged over luminosity bins for greater statistical power. Plots of the signal will be shown in Section 6.4.4.

All error bars shown in this section are statistical error bars from the computation of signal only; they do not include systematic error on the shear or redshift distributions, or statistical error in the determination of redshift distributions or photometric redshift error distributions.

6.4.1 Bright (r < 21) sources

For the r < 21 sources, several methods of redshift distribution determination are compared as follows:

(i) the redshift distribution p(z|r) from COMBO17,
(ii) the average redshift distribution p(z) from DEEP2,
(iii) photometric redshifts from KPHTOY directly,
(iv) photometric redshifts, corrected for the conditional bias (δp|z(r)) only,
(v) photometric redshifts, corrected for the best-fitting error distribution,
(vi) photometric redshifts, corrected for the actual (noisy) error distribution.

The first two methods involve averaging over a redshift distribution to determine Σ−1 without the use of photometric redshift information to eliminate physically associated pairs. The final four methods use photometric redshifts to select sources with z_p > z_l + 0.1 to eliminate physically associated source–source pairs. The two methods involving distributions use all r < 21 galaxies, including those for which photometric redshifts were not available and those passing the high-redshift LRG cuts; the last four methods only involve non-LRGs, since the photometric redshift error distributions for non-LRGs and LRGs are quite different.

In this section, we compare the results using 30–2000 h^−1 kpc, without subdividing this region, since changes in the redshift distribution only affect the amplitude of the signal, and not its slope. Furthermore, we only use luminosity bins L3–L6 since the majority of the signal is in those bins; adding L1 and L2 actually lowers the S/N slightly. Since changes in redshift distribution will tend to have a greater effect on the higher redshift lenses, we are also interested in the results as a function of luminosity bin. Table 15 shows the results for the ratios R_{i,a,p} for the selected ranges of radii, including 95 per cent confidence intervals computed via the non-Gaussian error computation for correlated variables (correlations between the signal computed using method 1 or 2 and methods 3–6 were of order 0.75, and correlations among methods 3–6 were of order 0.98). Because of the high correlations, the ratios were determined to high accuracy. Results are shown for L3, L6 (so that comparison can illustrate trends in lens luminosity) and the average signal in bins L3–L6.

We can see from Table 15 that at 95 per cent CL, the results from these different ways of computing Σ−1 all give results that have overall calibration within 15 per cent of each other. First, we compare the two redshift distributions shown in the table; this comparison is
on line 5 of the table. As shown, the distributions from COMBO-17 and DEEP2 give results that have no difference in overall calibration, and this calibration can only differ by $-2$ to $+3$ per cent at the 95 per cent CL when averaged over luminosity bins. Since Fig. 7 shows how similar these distributions are for $r < 21$, this result is not too surprising.

Next, we compare the results for the different ways of using photometric redshifts. First, we can consider the difference between using photometric redshifts directly and correcting for their mean bias (but not the width of the error distribution), shown in line 7. As demonstrated by the results for L3–L6, correcting for the mean bias lowers the signal by about 3 per cent, and the results are statistically inconsistent with the results without any correction for photometric redshift error at the 95 per cent CL. We expected the bias to lower the signal because $(\delta) < 0$ for low photometric redshift, so correcting for the mean bias will raise the assumed source redshift and $\Sigma_{c}$, lowering $\Delta \Sigma$ (while $(\delta) > 0$ for high photometric redshift, so correcting for it should raise the signal, the effect at low photometric redshift is apparently more important since $\Sigma_{c}^{-1}$ varies more strongly with $z_p$ at low $z_p$ despite the weights proportional to $\Sigma_{c}^{-2}$ that make higher redshift sources more important). We next consider what happens when we go from including the mean bias only, to including the full error distribution (line 13). As shown, including the width of the error distribution then raises the signal by 5 per cent; the results corrected for only the mean bias versus using the full error distribution (bias and scatter) are statistically inconsistent at approximately the 4$\sigma$ level, which means that including the width of the error distribution is very important. This increase in signal when we include the scatter is consistent with our expectation that scatter will lead to decreased $\Sigma_{c}^{-1}$, at least for the values $z_p > z_l + 0.1$ used here (at lower $z_p$ it may actually raise $\Sigma_{c}^{-1}$). The next comparison is using photometric redshifts directly versus including the full error distribution, line 8; consistent with combining the aforementioned results from lines 7 and 13, we find that including the full error distribution increases the signal compared to no correction at all by approximately 2 per cent, though the results are marginally consistent at the 2$\sigma$ level. This near agreement is due to the effects from the mean bias and width of the distributions cancelling out; the nearly exact cancellation was impossible to predict in advance, which is why this systematics test is so useful.

We can also use these results to place approximate 2$\sigma$ bounds on the calibration uncertainty due to statistical error in the photometric redshift error distribution. We place these bounds using the comparison between no correction at all, correction for $(\delta)$, and correction for the full error distribution, which gives approximately $\pm 3$ per cent calibration uncertainty (95 per cent CL). This result is comparable to the 2$\sigma$ Poisson error bars (determined via bootstrap resampling of the photometric redshift error distribution) which are $\pm 2.6$ per cent for lens redshift $\Delta z_{\text{resamp}}$, lower for lower lens redshift; we use the $\pm 3$ per cent figure to include extra uncertainty due to the effects of LSS.

Another informative test is the comparison between the results using the best-fitting error distribution versus the actual error distribution. When plotted against each other, the best-fitting distribution did show some significant differences from the actual one; for example, it did not account for the tails in the distribution very well. This comparison is on line 11 of Table 15; as shown, use of the best-fitting error distribution rather than the actual one increases the signal by about 3 per cent, and the two results are inconsistent at the 3$\sigma$ level. The sign of the discrepancy is in accordance with the results for $\Sigma_{c}^{-1}$ in Fig. 9. While the magnitude of this bias is clearly not large, since the difference is statistically significant, it seems that it would be preferable to use the actual error distribution to avoid this bias altogether.

Finally, we compare the results using redshift distributions versus using photometric redshift error distributions, lines 4 (DEEP2) and 15 (COMBO-17). We find that the signal is 3 per cent lower when we use either redshift distribution than when we use the photometric redshift error distributions; however, the two results only differ by about 1$\sigma$, so the difference in calibration is not statistically significant. However, as mentioned previously and demonstrated in Fig. 13, there is significantly more contamination from physically associated pairs when we use a distribution for $r < 21$ rather than using photometric redshift information to remove that contamination. Since there is no clear discrepancy due to the use of the redshift distributions versus photometric redshifts with error distributions, but the use of redshift distributions can lead to higher systematic error due to intrinsic alignments, we will henceforth use photometric redshifts with the actual error distributions for the $r < 21$ sample when comparing against the signal from other source subsamples.

Another systematics comparison that was completed was the computation of signal with $z_p > \langle z_p \rangle$ versus with $z_p < \langle z_p \rangle$ without correcting for photometric redshift errors. If our understanding of photometric redshift errors is correct, then we expect the signal at high $z_p$ to be lower than the signal at low $z_p$ by a statistically significant amount. Indeed, the ratio $R_{z_{\text{lower}}}$ is 0.79; the 3$\sigma$ confidence interval is [0.63, 1], so the discrepancy between the two samples is 3$\sigma$ and has the correct sign. When we correct for photometric redshift error, the discrepancy still exists at the 2$\sigma$ level; it is 0.86, with 2$\sigma$ confidence interval [0.74, 1]. The fact that correcting for photometric redshift errors helps reduce the discrepancy indicates that we are understanding these errors properly. The fact that there is still a discrepancy at the 2$\sigma$ level is not too alarming, because as mentioned previously, the errors used here are purely statistical. Once we add the overall calibration uncertainty, the discrepancy between the two samples is only 1$\sigma$ and therefore not of concern.

6.4.2 Faint ($r \geq 21$) sources

For the $r > 21$ sample, two methods of redshift distribution determination were considered:

(i) the redshift distributions from COMBO-17, $p(z|r)$, and

(ii) the average redshift distribution $p(z)$ from DEEP2 for $21 < r < 21.8$.

In both the cases, distributions were used, so we had no way of eliminating physically associated lens–source pairs; however, since the $r > 21$ sample is, on average, at higher redshift than the lenses, this procedure is less important than for the $r < 21$ sample. In both cases, high-redshift LRGs are included in the sample. As in the previous subsection, we consider results from 30 to 2000 h$^{-1}$ kpc. Results for $R_{\alpha,\beta}$ are shown in Table 16; since only one comparison is being made, we show results for all luminosity bins.

Not surprisingly, the minor discrepancy ($< 2\sigma$) that exists for L3–L4 becomes worse for the higher redshift lenses, since they are closer to the peak of the distribution and are therefore more sensitive to details of the distribution. For L5, the discrepancy is 3$\sigma$ and for L6, 4$\sigma$, with the result averaged over luminosities a 5 per cent discrepancy, and the two signals being different at the 4$\sigma$ level. Consequently, while the difference in signals is relatively small, we are able to use the systematics test to determine it to very high accuracy. The fact that the distribution from DEEP2 gives slightly higher results implies that it gives on average slightly lower
(\Sigma_c^{-1}) and therefore is at lower redshift on average. This result is not entirely surprising, since the DEEP2 distribution is specific to our lensing catalogue, but the COMBO-17 distribution was for photometric galaxies in general, and our lensing cuts will tend to eliminate higher redshift galaxies more than lower redshift ones. This understanding is in accordance with Fig. 7, which shows both distributions for \( r > 21 \). Consequently, for the remainder of this paper, the DEEP2 redshift distribution will be used for the \( r > 21 \) sample.

6.4.3 LRG sources

For LRGs, the signal was computed in the following two ways.

(i) Using photometric redshifts directly, for \( 0.4 < z_p < 0.65 \).

(ii) Using redshift distributions \( p(z|z_p) \) determined via the inversion procedure from Padmanabhan et al. (2004) for \( 0.45 < z_p < 0.65 \).

Besides these two methods of comparing the redshift distribution, we did several tests to ensure that our cuts on the LRG sample were optimal. Using method (i) to determine redshifts, we recompute the results with \( dz < 0.5 \) and then with \( dz > 0.55 \); these stricter cuts should help eliminate more of the low-redshift contamination (this contamination is a concern because we would be vastly overestimating its redshift and \( \Sigma_c^{-1} \), and underestimating \( \Delta \Sigma \)). If the signal computed using these cuts is similar to that computed with our cut, \( dz > 0.45 \), then we can be assured that the role of low-redshift contamination is minimal. In addition, we also compute the signal using photometric redshifts directly from \( 0.45 < z_p < 0.65 \) to ensure consistency with the full sample, in case there is contamination in the \( 0.4 < z_p < 0.45 \). Results for the ratio test are shown in Table 17.

We can come to several conclusions from this table. Lines 1–3 show the results of the systematics tests: using photometric redshifts directly for the full sample with \( dz > 0.45 \) and \( 0.4 < z_p < 0.65 \), versus requiring \( dz > 0.5 \) (line 1), \( dz > 0.55 \) (line 2) or \( 0.45 < z_p < 0.65 \) (line 3). As shown, we see no sign of any systematic discrepancy. This result is useful because it suggests that our photometric redshift and colour cuts are sufficient to eliminate low-redshift contamination to a negligible level, so there is no need to impose stricter cuts such as those listed above, which also lower the statistical power of this subsample. Line 4 shows a comparison of the results using photometric redshifts directly for \( 0.4 < z_p < 0.65 \) versus using the inversion method from Padmanabhan et al. (2004) for \( 0.45 < z_p < 0.65 \). As shown, the results are completely statistically consistent. This result is not surprising, since the LRG sample is at much higher redshift than the lenses, and therefore small changes in the redshift distribution have little effect on the final results. Because of the larger numbers of LRGs in the \( 0.4 < z_p < 0.65 \) sample, we will henceforth use the full LRG sample with photometric redshifts directly when comparing against other subsamples.

When considering the variation in calibration between the samples in Table 17, we conclude that the calibration uncertainty in the LRG signal due to redshift distribution determination is \( \pm 10 \) per cent at the 2\( \sigma \) level.

6.4.4 \textit{All}

In this subsection we compare the results from the three source samples (\( r < 21, r > 21 \) and LRGs) to check that they are consistent. As mentioned in the previous sections, we will use the photometric redshifts with the full error distributions from DEEP2 for \( r < 21 \) (without LRGs), the redshift distribution from DEEP2 for the \( r > 21 \) sample, and photometric redshifts directly for the LRG sample.

Furthermore, in contrast to the previous sections, our concern here is not merely overall calibration; since the different source samples have different boost factors and different random catalogue signals, it is important to check for radius-dependent effects. In principle, we could also measure the ratios \( R_{i,\alpha,\beta} \) for several radial ranges. However, the \( S/N \) for these comparisons was not high enough to give reasonable error bars. Consequently, we only consider larger ranges of 150–2000 \( h^{-1} \) kpc and 30–2000 \( h^{-1} \) kpc. The reason for using the 150–2000 \( h^{-1} \) kpc range is that even for L6 it excludes the angular separations <90 arcsec for which the boost factor has the difficult to quantify systematic due to sky subtraction, and it excludes the radii for which magnification bias may be important. Consequently, 150–2000 \( h^{-1} \) kpc is a more systematic-free ranges. Results are shown in Table 18.

We can draw several conclusions from this table. First, the fact that the ratios using 30–2000 and 150–2000 \( h^{-1} \) kpc are so similar illustrates that even with our concern about the systematic in the boosts at low values of transverse separation, use of the full radial range does not cause obvious problems. Secondly, there seems to be some luminosity-dependence of the overall calibration difference between the samples (especially \( r < 21 \) versus \( r > 21 \)), but the error bars on individual luminosity bins are large enough that no definite statement about this can be made. Finally, when averaged over all luminosities, the results for the three source subsamples agree well within error bars. We see no suggestion that there is anything fundamentally wrong with any of these three samples, either in the shear calibration, redshift distributions, or other systematics, and consider the consistency of the samples to be established at the 10 per cent level (1\( \sigma \)).

Here we show plots of the signal. Fig. 16 shows the signal from L5 for the three samples to illustrate our finding from the systematics test that there is no systematic calibration offset between the three samples. Fig. 17 shows the signal averaged over

\begin{table}
\centering
\begin{tabular}{l|l|l}
\hline
Luminosity & \( R_{i,\alpha,\beta} \) with 95 per cent CL bounds \\
\hline
L3 & 1.02^{+0.07}_{-0.06} \\
L4 & 1.04^{+0.05}_{-0.05} \\
L5 & 1.05^{+0.04}_{-0.05} \\
L6 & 1.09^{+0.05}_{-0.06} \\
L3–L6 & 1.05^{+0.02}_{-0.03} \\
\hline
\end{tabular}
\end{table}

\begin{table}
\centering
\begin{tabular}{l|l|l}
\hline
Sources \( \alpha \) & Sources \( \beta \) & \( R_{i,\alpha,\beta} \) \\
\hline
\( z_p, dz > 0.45 \) & \( z_p \) & 1.02^{+0.06}_{-0.05} \\
\( z_p, dz > 0.55 \) & \( z_p \) & 1.06^{+0.08}_{-0.07} \\
\( z_p \) & \( 0.45 < z_p < 0.65 \) & 0.98^{+0.10}_{-0.07} \\
\( z_p \) & \( p(z|z_p) \) & 1.00^{+0.10}_{-0.08} \\
\hline
\end{tabular}
\end{table}
Table 18. A comparison of the overall calibration of the signal when computed using the three main source samples, all luminosity bins with significant signal. The results for the two radial ranges are, of course, highly correlated. We show $R_{\alpha,\beta}$ and the 95 per cent CL.

| Lenses | $30–2000\ h^{-1}\ kpc$ | $150–2000\ h^{-1}\ kpc$ |
|--------|----------------|----------------|
|        | $r < 21$ without LRGs vs $r > 21$ | |
|        | $r < 21$ without LRGs vs LRGs | |
|        | $r > 21$ versus LRGs | |

Figure 16. $\Delta\Sigma(r)$ in L5 for each source sample as labelled on the plot. Points have a slight horizontal offset so that error bars on all three signals are visible.

Figure 17. $\Delta\Sigma(r)$ averaged over source subsample as a function of lens luminosity. For L1 and L2, which have signal statistically consistent with zero, the vertical scale is not logarithmic, and the zero level is shown as a dashed line; for the other luminosity bins, a logarithmic scale is used for $\Delta\Sigma$. The error bars shown are statistical only; the lines are the results of halo model fits as described in the text.

Table 19. For reference, $\langle \Sigma^{-1} \rangle$ values for all luminosity bins and source samples.

| Lenses | $10^3 \langle \Sigma^{-1} \rangle$, pc$^2$ h$^{-1}$ M$_\odot^{-1}$ |
|--------|--------------------------------------------------|
|        | $r < 21$ | $r > 21$ | LRG |
| L1     | 5.2      | 5.3     | 5.5  |
| L2     | 7.2      | 7.4     | 7.8  |
| L3     | 10.2     | 10.5    | 11.4 |
| L4     | 13.3     | 13.5    | 15.4 |
| L5     | 16.1     | 15.9    | 19.0 |
| L6     | 18.1     | 17.1    | 22   |

In order to allow the $\Delta\Sigma$ values shown on the plots to be converted to shear $\gamma_r$, Table 19 includes $\langle \Sigma^{-1} \rangle$ for each lens and source sample combination.

7 CONCLUSIONS

Using the systematics test, we have estimated the effects of redshift distribution errors and other systematics on the lensing signal. As a summary, Table 20 lists those systematics that are significant at approximately the 1 per cent level or higher. Systematics are only estimated in this table for the redshift determination methods for each sample that we have selected to use in this and future works (e.g. we do not show results for the use of redshift distributions for the $r < 21$ sample). We must emphasize that the estimates of the calibration uncertainty due to these effects are generally only relevant for this work, and their effects for other works must be assessed using the lens/source catalogues used for those works. For those systematics that have some non-trivial dependence on lens sample and/or radial range, rather than listing a value in this table, we refer the reader to the appropriate table/section. We have not included the
following systematic effects because they are expected to be at the 0.1 per cent level: atmospheric refraction, camera shear, the density-systematic shear correlation \(\langle \delta n \cdot \gamma_{\mathrm{sys}} \rangle\), and non-uniformity of the boost factor.

The first section of Table 20 includes a summary of the findings regarding the redshift distribution calibration, and ends with a calculation of the overall calibration uncertainty for each sample due to redshift distribution-related factors. The second section includes a list of other factors that lead to calibration uncertainty that have been discussed in this paper. The table ends with an estimate of the overall uncertainty at the 2\(\sigma\) level for various lens, source combinations and ranges of radii. The estimates are somewhat rough, in the sense that they cannot include all details of the error estimates completed in the text, and consequently a better idea of the magnitudes of some of the errors can be found by reading the appropriate sections of the text.

In this table, we refer to ‘small’ versus ‘large’ radii, where the distinction between the two is that on ‘large’ radii the estimates only include the calibration uncertainties such as stellar contamination, shear calibration and redshift distributions, but for ‘small’ radii, we must also be concerned about systematics such as intrinsic alignments, sky subtraction effects and magnification bias. The definition of small versus large radii varies by luminosity bin, and the approximate transition between the two occurs at 40, 60, 80, 100, 140 and 200 \(h^{-1}\) kpc for L1–L6, respectively.

The overall uncertainty estimates in Table 20 were computed using one of several possible methods. The difficulty is that by definition systematic errors are those with unknown probability distributions, so it is not clear how to combine confidence intervals for multiple effects. The first possible method, the most conservative, would be to linearly add together the upper bounds for all the independent sources of error to get an overall upper bound, and similarly for the lower bound. This method does not require us to make any assumptions about the probability distributions for these systematic errors, most of which are very likely not remotely Gaussian. The second approach would be to assume a uniform distribution of the systematic error between the 2\(\sigma\) bounds \([a, b]\) given in the table. With this assumption, we can find the mean \((a + b)/2\) and the standard deviation \((b - a)/\sqrt{12}\) for each systematic error, average the means and add the standard deviations in quadrature, and use those results to find combined 2\(\sigma\) bounds. This approach is also justifiable, since the use of a uniform distribution leads to rather generous confidence intervals, but the results are still smaller than the very conservative bounds from the first

Table 20. Summary of the main systematics investigated (and found to be significant at the 1 per cent level or greater) in this work, and estimates of their significance. Bounds on the error are given at the 2\(\sigma\) level for each effect and for all systematics combined.

| Systematic                                           | Lenses | Sources | Radii | 2\(\sigma\) bounds |
|------------------------------------------------------|--------|---------|-------|-------------------|
| Error in \(p(\delta_{\mathrm{rp}})\)                  | All    | \(r < 21\) | All   | \(0.03\)          |
| Redshift failures from DEEP2 (Table 4)               | All    | \(r > 21\) | All   | \([-0.14, +0.025]\) |
| Statistical error in \(p(r)\)                        | All    | \(r > 21\) | All   | Table 9           |
| LRG redshift distribution failure                    | All    | LRGs    | All   | \(0.10\)          |
| Overall redshift calibration                         | All    | \(r < 21\) | All   | \(0.03\)          |
| Overall redshift calibration                         | All    | \(r > 21\) | All   | \([-0.14, 0.07]\)  |

Other systematic effects

| Systematic                                           | Lenses | Sources | Radii | 2\(\sigma\) bounds |
|------------------------------------------------------|--------|---------|-------|-------------------|
| Shear calibration bias (Section 2.2.2)              | All    | \(r < 21\) | All   | \([-0.05, 0.12]\)  |
| Shear calibration bias                                | All    | \(r > 21\) | All   | \([-0.08, 0.18]\)  |
| Stellar contamination (Section 6.3.3)                | All    | LRGs    | All   | \([-0.06, 0.19]\)  |
| Stellar contamination                                 | All    | LRGs    | All   | \([0.040, 0.0]\)   |
| Stellar contamination                                 | All    | LRGs    | All   | \([-0.068, 0]\)    |
| Intrinsic alignments (Section 6.3.8)                 | L3–L6  | \(r < 21\) | All   | Table 14          |
| Intrinsic alignments                                  | L3–L6  | \(r > 21\) | All   | Table 14          |
| Magnification bias (Section 4.7.3)                   | L3–L6  | LRGs    | Small | Table 6           |
| Magnification bias                                    | L3–L6  | LRGs    | Small | Table 6           |
| Magnification bias                                    | L3–L6  | LRGs    | Small | Table 6           |
| PHOTO sky level error (Section 6.3.7)                | All    | All     | Small | \([-0.15, 0]\)     |
| Overall uncertainty (L1–L4)                          | L1–L4  | \(r < 21\) | Small | \([-0.15, 0.16]\)  |
| Overall uncertainty (L5–L6)                          | L5–L6  | \(r < 21\) | Large | \([-0.09, 0.11]\)  |
| Overall uncertainty (L5–L6)                          | L5–L6  | \(r > 21\) | Small | \([-0.41, 0.42]\)  |
| Overall uncertainty (L5–L6)                          | L5–L6  | \(r > 21\) | Large | \([-0.13, 0.14]\)  |
| Overall uncertainty (L5–L6)                          | L5–L6  | \(r > 21\) | Small | \([-0.22, 0.21]\)  |
| Overall uncertainty (L5–L6)                          | L5–L6  | \(r > 21\) | Large | \([-0.19, 0.19]\)  |
| Overall uncertainty (L5–L6)                          | L5–L6  | \(r > 21\) | Small | \([-0.44, 0.44]\)  |
| Overall uncertainty (L5–L6)                          | L5–L6  | \(r > 21\) | Large | \([-0.27, 0.26]\)  |
| Overall uncertainty (L1–L4)                          | L1–L4  | LRGs    | Small | \([-0.18, 0.19]\)  |
| Overall uncertainty (L1–L4)                          | L1–L4  | LRGs    | Large | \([-0.16, 0.18]\)  |
| Overall uncertainty (L1–L4)                          | L1–L4  | LRGs    | Small | \([-0.21, 0.23]\)  |
| Overall uncertainty (L1–L4)                          | L1–L4  | LRGs    | Large | \([-0.16, 0.18]\)  |
approach. The final approach would be to add the lower bounds in quadrature, and add the upper bounds in quadrature. For Table 20, we used the bounds from the second method (assuming a uniform distribution); the results using the first method are about 50–100 per cent larger, and using the third method are about 0–10 per cent larger.

In this paper, we have sought to understand the various sources of calibration uncertainty and other systematics in the galaxy–galaxy weak lensing signal computed using our lens and source catalogues. As more data are collected, we will be able to perform these tests to constrain the systematics with greater precision. The understanding we have gained from the systematics test will allow us to use this catalogue for many scientific applications (studies of the relation between mass and luminosity, mass and stellar masses, halo profiles as a function of environment and others) with an understanding of the implication of systematic errors on the results. Similar investigations of systematics will also be needed for weak lensing autocorrelation analyses, once the statistical errors become as small as they are for our SDSS galaxy–galaxy lensing analysis.

ACKNOWLEDGMENTS

We gratefully acknowledge Edd Edmondson, Christian Wolf and Lance Miller for their work on the redshift distributions from COMBO-17, which they allowed us to use. We also acknowledge James Gunn, Erin Sheldon and David Johnston for useful discussion regarding systematics in the lensing signal. We thank the entire DEEP2 team for allowing us to use their redshift data in the Groth strip, and are particularly thankful for the help and feedback from Doug Finkbeiner, Alison Coil and Jeffrey Newman.

RM is supported by an NSF Graduate Research Fellowship, and CH by NASA grant NGT5-50383. US was supported by a fellowship from the David and Lucile Packard Foundation, NASA grants NAGS-1993 and NASA NAGS-11489, and NSF grant CAREER-0132953.

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 US 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.

Funding for the DEEP2 survey has been provided by NSF grants AST-0071048 and AST-0071198. Some of the data presented herein were obtained at the W.M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California and the National Aeronautics and Space Administration. The Observatory was made possible by the generous financial support of the W.M. Keck Foundation. The DEEP2 team and Keck Observatory acknowledge the very significant cultural role and reverence that the summit of Mauna Kea has always had within the indigenous Hawaiian community and appreciate the opportunity to conduct observations from this mountain.

REFERENCES

Abazajian K. et al., 2003, AJ, 126, 2081
Abazajian K. et al., 2004, AJ, 128, 502
Abazajian K. et al., 2005, AJ, 129, 1755
Bernstein G., Jain B., 2004, ApJ, 600, 17
Bernstein G. M., Jarvis M., 2002, AJ, 123, 583
Bernstein G. M., Norberg P., 2002, AJ, 124, 733
Blanton M. R., 2003a, ApJ, 592, 819.
Blanton M. R., Lin H., Lupton R. H., Maley F. M., Young N., Zehavi I., Loveday J., 2003b, AJ, 125, 2276
Blanton M. R. et al., 2003c, AJ, 125, 2348
Blanton M. R. et al., 2005, AJ, 129, 2562
Bliss C. L., 1935a, Ann. Appl. Biol., 22, 134
Bliss C. L., 1935b, Ann. Appl. Biol., 22, 307
Brunner R. T., Blandford R. D., Small I., 1996, ApJ, 466, 623
Brown M. L., Taylor A. N., Bacon D. J., Gray E. M., Dye S., Meisenheimer K., Wolf C., 2003, MNRAS, 341, 100
Budavári T., Szalay A. S., Connolly A. J., Csabai I., Dickinson M., 2000, AJ, 120, 1588
Catelan P., Kambournakis M., Blandford R. D., 2001, MNRAS, 320, L7
Coil A. L. et al., 2004a, ApJ, 609, 525
Coil A. L., Newman J. A., Kaiser N., Davis M., Ma C-P., Kocevski D. D., Koo D. C., 2004b, ApJ, 617, 765
Connolly A. J. et al., 2002, ApJ, 579, 42
Crittenden R. G., Natarajan P., Pen U., Theuns T., 2001, ApJ, 559, 552
Croft R. A. C., Metzler C. A., 2000, ApJ, 545, 561
Cramer H., 1946, Mathematical Methods of Statistics. Princeton Univ. Press
Csabai I., Connolly A. J., Szalay A. S., Budavári T., 2000, AJ, 119, 69
Csabai I. et al., 2003, AJ, 125, 580
Davis M. et al., 2003, in Guhathakurta P., ed., Proc. SPIE Vol. 4834, Discoveries and Research Prospects from 6- to 10-meter-class Telescopes. SPIE, p. 161
Davis M., Gerke B., Newman J. A., 2004, Observing Dark Energy: NOAO Workshop (astro-ph/0408344)
Eisenstein D. J. et al., 2001, AJ, 122, 2267
Fielker E. C., 1954, J. R. Stat. Soc. B, 16, 175
Finkbeiner D. et al., 2004, AJ, 128, 2577
Fischer P. et al., 2000, AJ, 120, 1198
Flin P., Godlowski W., 1989, Astronomy, Cosmology and Fundamen
tal Physics: Proc. Third ESO-CERN Symposium (A90-44077 19-90), Kluwer, Dordrecht, Netherlands and Boston, MA, p. 418.
Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, AJ, 111, 1748
Giavalisco M. et al., 2004, ApJ, 600, L93
Gunn J. E. et al., 1998, AJ, 116, 3040
Guzik J., Seljak U., 2002, MNRAS, 335, 311
Hamilton A. J. S., Tegmark M., 2004, MNRAS, 349, 115
Heavens A., Refregier A., Heymans C., 2000, MNRAS, 319, 649
Heymans C., Brown M., Heavens A., Meisenheimer K., Taylor A., Wolf C., 2004, MNRAS, 347, 895
Heymans C. et al., 2005, MNRAS, 361, 160
Hirata C. M., Seljak U., 2003, MNRAS, 343, 459
Hirata C. M. et al., 2004, MNRAS, 353, 529
Hoekstra H., 2003, in Colless M., Staveley-Smith L., eds, IAU Symp. 216, Maps of the Cosmos. Astron. Soc. Pac., San Francisco, p. 11
Hoekstra H., 2004, MNRAS, 353, 351
Hoekstra H., 2005, MNRAS, 340, 609
Hoekstra H., 2006, MNRAS, 365, 1067
Hogg D. W., Finkbeiner D. P., Schlegel D. J., Gunn J. E., 2001, AJ, 122, 2129
Hudson M. J., Gwyn S. D. J., Dahle H., Kaiser N., 1998, ApJ, 503, 531
Hui L., Zhang J., 2002, preprint (astro-ph/0205512)
Ishak M., Hirata C. M., 2005, Phys. Rev. D, 71, 023002
Ivezic Z. et al., 2004, Astron. Nachr., 325, 585
Jain B., Taylor A., 2003, Phys. Rev. Lett., 91, 1302

Systematic errors in weak lensing

1319
APPENDIX A: $\chi^2$ FOR REDSHIFT DISTRIBUTION FITS

The redshift distribution fits were obtained by minimizing the $\chi^2$ function

$$\chi^2 = -2 \sum_{i=1}^{N_s} \ln p(z_i),$$

(A1)

where $N_s$ is the number of spectra, and $p(z_i) = (dP/dz)|_{z=z_i}$ is the normalized redshift distribution. Equation (A1) is a $\chi^2$ in the sense that the likelihood for a given model $p(z)$ is proportional to $e^{-\chi^2/2}$ in the case where the galaxy redshifts are independent. While minimizing $\chi^2$ is a reasonable way to get the redshift distribution, it is not trivial to test for a ‘goodness of fit’ since (i) even in the case of independent redshifts, equation (A1) does not follow the standard $\chi^2$ distribution and (ii) there is LSS that introduces correlations between different redshifts, particularly in a narrow survey such as DEEP2. This Appendix is devoted to calculating the statistical properties of $\chi^2$ and the fit parameters obtained via its minimization.

A1 Mean and variance of $\chi^2$

The mean of equation (A1) is

$$\mu_{\chi^2} = \langle \chi^2 \rangle = 2N_s S,$$

(A2)

where the ‘entropy’ $S$ is the expectation value of $- \ln p(z)$ for a single galaxy, i.e.

$$S = - \int p(z) \ln p(z) \, dz.$$

(A3)

The entropy is a property of the redshift distribution only; e.g. for the $\Gamma$-distribution (equation 40),

$$S = \alpha + \ln \Gamma(\alpha) - (\alpha - 1) \psi(\alpha) + \ln z_*,$$

(A4)

where $\psi(\alpha) = d \ln \Gamma(\alpha)/d\alpha$ is the digamma function. Note that $S$ can be positive, zero, or negative depending on the distribution; it is larger for wide distributions.

Equation (A2) is the mean of a simple sum over galaxies and hence it is valid regardless of LSS. The variance is more complicated; it is

$$\sigma_{\chi^2}^2 = \langle (\chi^2 - \mu_{\chi^2})^2 \rangle = \sum_{i=1}^{N_s} \langle [\ln p(z_i) + S][\ln p(z_i) + S] \rangle.$$

(A5)

The expectation value here contains two galaxy redshifts, and hence depends on both the redshift distribution and the redshift correlation function. Defining the redshift correlation function $\xi_1$ by

$$p(z_i, z_j) \approx p(z_i) p(z_j) [1 + \xi_1(z_i, z_j)],$$

(A6)

we find

$$\sigma_{\chi^2}^2 = 4N_s S_{11} + 4N_s (N_s - 1) S_2,$$

(A7)

where

$$S_{11} = \int p(z) [\ln p(z) + S]^2 \, dz,$$

(A8)

and

$$S_2 = \int p(z_1) p(z_2) \xi_1(z_1, z_2) [\ln p(z_1) + S] \times [\ln p(z_2) + S] \, dz_1 \, dz_2.$$  

(A9)

The ‘independent redshift’ contribution to $\sigma_{\chi^2}^2$ comes from the $S_{11}$ term, which does not involve the correlation function. For the $\Gamma$-distribution,

$$S_{11} = (\alpha - 1)^2 \psi'(\alpha) - \alpha + 2.$$  

(A10)

A2 Correlation function and $S_2$

The evaluation of the clustering contribution $S_2$ to the variance of $\chi^2$ is more complicated. We begin by assuming that the three-dimensional correlation function of the sources is a power law, $\xi(r) = (r_0/r)^{\gamma}$. Written in polar coordinates, this becomes

$$\xi(r_1, r_2, \theta) \approx \frac{r_0^\gamma}{(r_1 - r_2)^2 + r_1^2 r_2^2} r_1^{\gamma/2},$$

(A11)

where the approximation is valid for small $\theta$ (the relevant case here). The correlation falls off at large $|r_1 - r_2|$ as $|r_1 - r_2|^{-\gamma}$, hence for $\gamma > 1$ the integral is finite and $\xi$ is non-zero only for $r_1 \approx r_2$. We thus make a Limber approximation,

$$\xi(r_1, r_2, \theta) \approx r_0^\gamma \delta(r_1 - r_2) \int_{-\infty}^{\infty} \frac{d\Delta r}{(\Delta r^2 + r_1^2 r_2^2)^{\gamma/2}}.$$

(A12)

\[B \left( \frac{\gamma - 1}{2}, \frac{1}{2} \right) \cdot \left( r_0^\gamma (r_1 r_2)^{-\gamma} \delta(r_1 - r_2) \right).\]
The resulting correlation function $\xi(z_1, z_2)$ is obtained by converting from comoving distance to redshift, and averaging over the angular separations $\theta$ in the survey,
\[
\xi(z_1, z_2) = A \mathcal{R}(\theta) \frac{dz}{d\theta} \delta(z_1 - z_2), \tag{A13}
\]
where for the purposes of obtaining a rough estimate of the importance of $S_2$ we assume $A = B(y^{-1} \frac{1}{2})r_3^0$ to be a constant.\(^7\) The result for $S_2$ is
\[
S_2 = A(\theta^{-1}) \int [p(z)]^2 r^{1-\gamma} H(z) [\ln p(z) + S]^2 dz, \tag{A14}
\]
where we have used $dz/d\theta = H(z)$. This may be divided into $y$ by comparing the observed angular two-point correlation function of the galaxies; the theoretical prediction is
\[
\omega(\theta) = \int \xi(r_1, r_2, \theta) \frac{dp}{d\theta}(r_1) \frac{dp}{d\theta}(r_2) dr_1 dr_2 = A\theta^{-1} \int [p(z)]^2 r^{1-\gamma} H(z) dz. \tag{A15}
\]
The angular correlation function in SDSS (Connolly et al. 2002) is fit by a power law with $\gamma = 1.7$. The DEEP2/SDSS overlap region can be approximated by a rectangle of dimensions 0.25 x 0.6 deg, hence averaged over all pairs of points in this region (obtained via the obvious Monte Carlo procedure) is $(\theta^{-1}) \approx 3.7 \text{ deg}^{-0.7}$.\(^7\)

\section*{A3 Parameter uncertainties}

The uncertainty of the redshift distribution parameters is greater than the naive result from $\Delta \chi^2$ contours because of source clustering. We begin our analysis of this effect by assuming that the redshift distribution depends on a vector of $M$ parameters $a = (a^1, \ldots, a^M)$; in the case of the $\Gamma$-distribution, $M = 2$ and $a = (z, \alpha)$, and we are then estimating the parameters $a$ by minimizing $\chi^2(a)$. We denote the true set of parameters by $a_0$.

We begin by Taylor expanding $\chi^2$ around $a_0$,
\[
\chi^2(a) = \chi^2(a_0) + G \cdot (a - a_0) + (a - a_0) \cdot K(a - a_0) + \ldots, \tag{A16}
\]
where $G$ and $K$ are the derivatives of $\chi^2$ at $a_0$. The minimum value of $a$ in this approximation is then
\[
\hat{a} = a_0 - \frac{1}{2} K^{-1} G. \tag{A17}
\]
We can calculate $G$ and $K$ in terms of the measured source redshifts,
\[
G_\mu = -2 \sum_{i=1}^{N} \left. \frac{\partial}{\partial a^\mu} \ln p(z_i | a) \right|_{a_0} \tag{A18}
\]
and
\[
K_{\mu \nu} = - \sum_{i=1}^{N} \left. \frac{\partial^2}{\partial a^\mu \partial a^\nu} \ln p(z_i | a) \right|_{a_0}. \tag{A19}
\]

\section*{APPENDIX B: NON-GAUSSIAN CONFIDENCE INTERVALS FOR CORRELATED VARIABLES}

In this section, we explain the method used to compute non-Gaussian confidence intervals on the ratio of correlated variables. First, we consider the case of the ratio of two uncorrelated variables with Gaussian noise, then generalize to the correlated case.

When computing the error on the ratio of two variables $R = y/x$, where $y$ and $x$ are Gaussian variables with standard deviations $\sigma_y$ and $\sigma_x$, the naive result $\sigma_R = \sqrt{(\sigma_y/x^2) + (\sigma_x/y^2)}$ is unreliable in the case when $x$ may be consistent with zero even at the many-$\sigma$ level. Instead, the non-Gaussian confidence intervals should be used. If we want to find the confidence intervals on $R$ at the $z\sigma$ level

\[
\frac{\partial^2}{\partial \ln y \partial \ln x} \frac{\partial}{\partial \ln y} \frac{\partial}{\partial \ln x} f \left( \frac{y}{x} \right) \bigg|_{y_0} = \frac{\sigma_y^2}{x^2} + \frac{\sigma_x^2}{y^2}, \tag{A20}
\]

\section*{Appendix}

We now estimate the statistical properties of $\hat{a}$ using equation (A17), approximating $K$ by its expectation value for simplicity:
\[
\langle K_{\mu \nu} \rangle = -N_i \left. \frac{\partial^2 \ln p(z_i)}{\partial \alpha^\mu \partial \alpha^\nu} \right|_{a_0}, \tag{A20}
\]

\section*{References}

\(^7\)Technically, equation (A13) should contain an additive constant to account for the integral constraint $\int p(z_i, z_2) dz_i dz_2 = 1$ in equation (A6). An additive constant in $\xi(z_1, z_2)$ would not affect $S_2$ because its contribution in equation (A9) multiplies $\int p(z_1)p(z_2)[\ln p(z_1) + S][\ln p(z_2) + S]dz_1dz_2 = 0$.\(^7\)

\(^8\)The second equality in equation (A20) can be found by differentiating the relation $\int p(z) dz = 1$ twice to get $\langle \partial^2 \ln p(z)/\partial a^\mu \partial a^\nu \rangle / p(z) = 0$, and then rewriting this result in terms of $\ln p(z)$.\(^8\)
(i.e. \( z = 1 \) would give the 68 per cent confidence interval), then the result (Bliss 1935a,b; Fieller 1954) is as follows:

\[
y \pm \frac{x}{\sqrt{2}} \sqrt{\frac{z}{e^2 - e^2}} = \frac{y}{\sqrt{2}} \sqrt{k\sigma_x^2} + \frac{z}{\sqrt{2}} \sqrt{k\sigma_y^2},
\]

where we define

\[
k = \sqrt{\left(\frac{x}{\sigma_x^2}\right)^2 + \left(\frac{y}{\sigma_y^2}\right)^2}.
\]

If \( y \) and \( x \) are correlated, with non-zero correlation coefficient \( \rho = \text{Cov}(x, y)/(\sigma_x \sigma_y) \), then equation (B1) is no longer valid. In order to get confidence intervals on \( R \), we do the following manipulations:

we change variables so that we are once again dealing with two uncorrelated variables, in this case \( w = y - mx \), where \( m = \rho \sigma_y / \sigma_x \). Here \( w \) is clearly uncorrelated with \( x \), and has error \( \sigma_w^2 = \sigma_x^2 (1 - \rho^2) \). Furthermore, \( R = w/x + m \). We can then use the procedure for uncorrelated variables to compute confidence intervals on \( w/x \), so that the upper and lower limits on \( y/x \) are related to those on \( w/x \) by the addition of \( m \).

**APPENDIX C: CALIBRATION BIAS FROM SIGNIFICANCE CUT**

There is a contribution to the shear calibration bias originating from selection effects from the signal-to-noise cut in our catalogue. We can investigate the magnitude of this effect by the same methods used in Section 2.2.2 PHOTO convolves the observed image \( I \) with the PSF \( P \) to create a new image \( J = I \otimes P \), and finds pixels in \( P \) above some threshold \( J_x \), set to \( \nu_c \sigma(J(x)) \) above the sky background, where the signal-to-noise threshold is \( \nu_c = 5 \). This convention is chosen because, in the case where \( P \) is symmetric under 180° rotation, \( J(x) \) represents the optimal statistic for searching for a point source at \( x \). Objects above \( J_x \) are designated as BINNED1 by PHOTO; for detection, we require that this flag be set in the \( r \) and \( i \) bands. The threshold is

\[
J_x = \nu_c \sqrt{n \int P(x)^2 dx} = \begin{cases} 0.47\nu_c^{1/2} \sigma_{\text{FWHM}}^{-1} \quad \text{(Gaussian)} \\ 0.57\nu_c^{1/2} \sigma_{\text{FWHM}}^{-1} \quad \text{(Kolmogorov)} \end{cases}
\]

where \( n \) is the noise variance per pixel and \( \theta_{\text{FWHM}} \) is the full-width at half-maximum of the PSF. We have shown both the case of the Gaussian PSF and the Kolmogorov turbulence-induced PSF \( \ln P(k) \propto k^{3/2} \). For typical parameters, \( J_c = 0.13 \) nmgy arcsec\(^{-2} \) in \( r \) band and 0.22 nmgy arcsec\(^{-2} \) in \( i \) band.

The value of \( J(0) \), where we have translated the object’s centroid to the origin 0, is given by

\[
J(0) = \int I(x)P(x) d^2x = \int f(x)K(x) d^2x,
\]

where \( K = P \otimes P \) and we have assumed \( P(x) = P(-x) \). The simplest way to evaluate \( J \) is in the case of a circular Gaussian PSF and elliptical Gaussian galaxy, for which

\[
J(0) = \frac{F}{2\pi} \left[ \text{det} \left( M^{(j)} + 2M^0 \right) \right]^{-1/2}, \quad (C3)
\]

where \( F \) is the galaxy flux. Under a shear leaving \( \sigma^{(j)} \) constant, this varies by

\[
\frac{\delta J(0)}{J(0)} = -\frac{2(1 - R_2 \epsilon \cdot \gamma)}{4R_2^2 - 4 + R_2 - e^2}.
\]

The resulting shear selection bias can then be computed by the same method as was used in Section 2.2.2. It yields

\[
\frac{\delta f}{f} = -\left[ \frac{(1 - R_2 e^2)}{(4R_2^2 - 4 + R_2 - e^2)R} \right] \frac{dN(j)}{d \ln j} \bigg|_{j = J_c}, \quad (C5)
\]

where \( N(j) \) is the (weighted) fraction of galaxies with \( J(0) < j \). The coefficient of \( dN(j)/d \ln j \) has a mean value (averaged over the \( R_2 \) distribution) of \(-0.024, -0.028 \) and \(-0.026 \) for the \( r < 21, r > 21 \) and LRG samples, respectively, in \( r \) band, and \(-0.023, -0.026 \) and \(-0.024 \) in \( i \) band. Conservative values of \( dN(j)/d \ln j \) (obtained by maximizing over \( j \), since we have not calculated the threshold exactly) are 1.40, 2.37 and 1.42, respectively, in \( r \) band, and 1.23, 2.01 and 1.46 in \( i \) band. We conservatively estimate the calibration bias due to the signal-to-noise cut using the \( r \)-band values, and obtain \(-0.034, -0.066 \) and \(-0.037 \) for \( r < 21, r > 21 \) and LRG samples, respectively.

This paper has been typeset from a TEX/LaTeX file prepared by the author.