Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **401,** 1399–1408 (2010) doi:10.1111/j.1365-2966.2009.15748.x

Catastrophic photometric redshift errors: weak-lensing survey requirements

Gary Bernstein[1★] and Dragan Huterer[2★]

[1]*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*
[2]*Department of Physics, University of Michigan, 450 Church St, Ann Arbor, MI 48109-1040, USA*

Accepted 2009 September 19. Received 2009 September 18; in original form 2009 February 26

ABSTRACT

We study the sensitivity of weak-lensing surveys to the effects of *catastrophic* redshift errors – cases where the true redshift is mis-estimated by a significant amount. To compute the biases in cosmological parameters, we adopt an efficient linearized analysis where the redshift errors are directly related to shifts in the weak-lensing convergence power spectra. We estimate the number $N_{\rm spec}$ of unbiased spectroscopic redshifts needed to determine the catastrophic error rate well enough that biases in cosmological parameters are below statistical errors of weak-lensing tomography. While the straightforward estimate of $N_{\rm spec}$ is $\sim 10^6$, we find that using only the photometric redshifts with $z \lesssim 2.5$ leads to a drastic reduction in $N_{\rm spec}$ to $\sim$30 000 while negligibly increasing statistical errors in dark-energy parameters. Therefore, the size of the spectroscopic survey needed to control catastrophic errors is similar to that previously deemed necessary to constrain the core of the $z_{\rm s}$–$z_{\rm p}$ distribution. We also study the efficacy of the recent proposal to measure redshift errors by cross-correlation between the photo-z and spectroscopic samples. We find that this method requires $\sim$10 per cent a priori knowledge of the bias and stochasticity of the outlier population, and is also easily confounded by lensing magnification bias. The cross-correlation method is therefore unlikely to supplant the need for a complete spectroscopic-redshift survey of the source population.

Key words: gravitational lensing.

1 INTRODUCTION

Weak gravitational lensing is a very promising cosmological probe that has potential to accurately map the distribution of dark matter and measure the properties of dark energy and the neutrino masses (for reviews, see Bartelmann & Schneider 2001; Huterer 2002; Hoekstra & Jain 2008; Munshi et al. 2008). It is well understood, however, that systematic errors may stand in the way of weak lensing reaching its full potential – that is, achieve the *statistical* errors predicted for future ground- and space-based surveys such as the Dark Energy Survey (DES[1]), Large Synoptic Survey Telescope (LSST[2]) and the Joint Dark Energy Mission (JDEM[3]). Controlling the systematic errors is a primary concern in these and other surveys so that a variety of dark-energy tests (recently proposed and reviewed by the JDEM Figure of Merit Science Working Group; Albrecht et al. 2009) can be performed to a desired high accuracy.

Several important sources of systematic errors in weak-lensing surveys have already been studied. Chief among them is the redshift accuracy – approximate, photometric redshifts are necessary because it is infeasible to obtain optical spectroscopic redshifts for the huge number ($\sim 10^8$–10^9) of galaxies that future surveys will utilize as lensing sources (Ma, Hu & Huterer 2005; Huterer et al. 2006; Abdalla et al. 2008a; Amara & Réfrégier 2008). It is therefore imperative to ensure that statistical errors and systematic biases in the relation between photometric and spectroscopic redshifts, recently studied in depth with real data as well as simulations (Oyaizu et al. 2008; Banerji et al. 2008; Lima et al. 2008; Cunha et al. 2009; Abdalla et al. 2008b), do not lead to appreciable biases in cosmological parameters.

The relation between the photometric and spectroscopic redshifts has been previously modelled as a Gaussian with a redshift-dependent bias and scatter. It is found that both the bias and scatter (i.e. quantities $\langle z_{\rm p} – z_{\rm s}\rangle$ and $\langle (z_{\rm p} – z_{\rm s})^2\rangle^{1/2}$ in each bin of $\Delta z = 0.1$) need to be controlled to about 0.003 or better in order to lead to less than $\sim$50 per cent degradation in cosmological parameter accuracies (Ma et al. 2005; Huterer et al. 2006; Kitching et al. 2008). These constraints are a bit more stringent in the most general case when the redshift error cannot be described as a Gaussian (Ma & Bernstein 2008). These requirements imply that $N_{\rm spec} \lesssim 10^5$

★E-mail: garyb@physics.upenn.edu (GB); huterer@umich.edu (DH)
[1] http://www.darkenergysurvey.org/
[2] http://www.lsst.org/
[3] http://jdem.gsfc.nasa.gov/

spectra are required in order to calibrate the photometric redshifts to the desired accuracy. Gathering such relatively large spectroscopic sample will be a challenge, setting a limit to the useful depth of weak-lensing surveys. (While the redshift errors have been well studied, other systematics are also important, especially theoretical errors in modelling clustering of galaxies at large and small scales, intrinsic shape alignments and various systematic biases that take place during observations; Huterer & Takada 2005; White 2004; Hagan, Ma & Kravtsov 2005; Zhan & Knox 2004; Huterer & White 2005; Zentner, Rudd & Hu 2008; Rudd, Zentner & Kravtsov 2008; Shapiro & Cooray 2006; Shapiro 2009; Heitmann et al. 2005, 2008; Takada & Bridle 2007; Takada & Jain 2009; Kitching et al. 2008; Joachimi & Schneider 2008; Bernstein 2009; Jarvis & Jain 2004; Jain, Jarvis & Bernstein 2006; Ma et al. 2008; Guzik & Bernstein 2005; Stabenau et al. 2007; Heymans et al. 2006; Massey et al. 2007; Paulin-Henriksson et al. 2008; Amara & Réfrégier 2008; Heavens, Refregier & Heymans 2000; Croft & Metzler 2000; Crittenden et al. 2001; Mackey, White & Kamionkowski 2002; Jing 2002; Heymans & Heavens 2003; King & Schneider 2003; Hirata et al. 2004; Hirata & Seljak 2004; Mandelbaum et al. 2006; Bridle & Abdalla 2007; Bridle & King 2007.)

Essentially, all of the aforementioned photo-z requirement studies (e.g. Ma et al. 2005; Huterer et al. 2006), however, have modelled the errors as a perturbation around the z_s–z_p relation. While this perturbation was allowed to be large and to have a non-zero mean and scatter (e.g. Huterer et al. 2006; Abdalla et al. 2008a; Amara & Réfrégier 2008) and even skewness (e.g. Ma et al. 2005), it did not subsume a general, multimodal error in redshift.

In this paper, we would like to remedy this omission by estimating the effect of *catastrophic* redshift errors. Catastrophic errors are loosely defined as cases when the photometric redshift is grossly mis-estimated, i.e. when $|z_p–z_s| \sim O(1)$, and are represented by arbitrary 'islands' in the z_p–z_s plane. We develop a formalism that treats these islands as small 'leakages' (or 'contaminations') and directly estimates their effect on bias in cosmological parameters. We then invert the problem by estimating how many spectroscopic redshifts are required to control catastrophic errors at a level that makes them harmless for cosmology.

This paper is organized as follows. In Section 2, we derive the relevant equations for the bias in cosmological parameters induced by mis-estimated catastrophic redshift errors in a tomographic weak-lensing survey. In Section 3, we apply these methods to a canonical ambitious weak-lensing cosmology project. In Section 4, we ask the following question: how large must a *complete* spectroscopic-redshift survey be in order that the catastrophic photo-z error rates be measured sufficiently well that remnant cosmological biases are well below the statistical uncertainties? Newman (2008) has suggested an alternative mode of measuring the photo-z error distribution, namely the angular cross-correlation of the photometric galaxy sample nominally at z_p with a spectroscopic sample at z_s; in Section 5, we investigate whether systematic errors in the photo-z outlier rates derived from this cross-correlation technique will be small enough to render cosmological biases insignificant. Section 6 discusses the scaling of these results with critical survey parameters, the ramifications for survey design and areas of potential future investigation.

2 FORMALISM

In this section, we establish the formalism that takes us from 'islands' in the z_s–z_p plane to biases in cosmological parameters. First, however, we define the basic observable quantity, the convergence power spectrum and its corresponding Fisher information matrix.

2.1 Basic observables and the Fisher matrix

The convergence power spectrum of weak lensing at a fixed multipole ℓ and for the ith and jth tomographic bins is given by

$$\mathcal{P}^{\kappa}_{ij}(\ell) = \frac{\ell^3}{2\pi^2}\int_0^{\infty} \mathrm{d}z\, \frac{W_i(z)\, W_j(z)}{r(z)^2\, H(z)}\, \mathcal{P}_{\rm mat}\left(\frac{\ell}{r(z)}, z\right), \tag{1}$$

where $r(z)$ is the comoving distance, $H(z)$ is the Hubble parameter and $\mathcal{P}_{\rm mat}(k, z)$ is the matter power spectrum. The weights W_i are given, for a flat universe, by $W_i(\chi) = \frac{3}{2}\,\Omega_{\rm M}\, H_0^2\, g_i(\chi)\,(1+z)$, where $g_i(\chi) = \chi\int_{\chi}^{\infty} \mathrm{d}\chi_s n_i(\chi_s)(\chi_s - \chi)/\chi_s$, χ is the comoving distance and n_i is the comoving density of galaxies if χ_s falls in the distance range bounded by the ith redshift bin and zero otherwise. We employ the redshift distribution of galaxies of the form $n(z) \propto z^2 \exp(-z/z_0)$ that peaks at $2z_0 \simeq 0.9$.

The observed convergence power spectrum is

$$C^{\kappa}_{ij}(\ell) = \mathcal{P}^{\kappa}_{ij}(\ell) + \delta_{ij}\, \frac{\left\langle \gamma^2_{\rm int}\right\rangle}{\bar{n}_i}, \tag{2}$$

where $\langle \gamma^2_{\rm int}\rangle^{1/2}$ is the rms intrinsic shear in each component which we assume to be equal to 0.24 and $\bar{n}_i$ is the average number of galaxies in the ith redshift bin per steradian. The cosmological constraints can then be computed from the Fisher matrix

$$F_{ij} = \sum_{\ell} \frac{\partial \boldsymbol{D}}{\partial p_i}\, \mathrm{Cov}^{-1}\, \frac{\partial \boldsymbol{D}}{\partial p_j}, \tag{3}$$

where Cov^{-1} is the inverse of the covariance matrix between the observed power spectra $C^{\kappa}_{ij}(\ell)$ which have been organized into a vector $\boldsymbol{D}$, and matrix multiplication is implied. For a Gaussian convergence field, its elements are given by

$$\mathrm{Cov}\left[C^{\kappa}_{ij}(\ell), C^{\kappa}_{kl}(\ell)\right] = \frac{\delta_{\ell\ell'}}{(2\ell+1)\, f_{\rm sky}\, \Delta\ell} \times \left[C^{\kappa}_{ik}(\ell) C^{\kappa}_{jl}(\ell) + C^{\kappa}_{il}(\ell) C^{\kappa}_{jk}(\ell)\right], \tag{4}$$

where $f_{\rm sky}$ is fraction of the sky observed and $\Delta\ell$ is the binning of the convergence power spectra in multipole.

We assume the parametrization of the equation of state of dark energy $w(a) = w_0 + (1-a)w_a = w_p + (a_p - a)w_a$, where in the latter equality the pivot scale factor a_p is chosen so that the parameters $w_p \equiv w_0 + (1-a_p)w_a$ and w_a are uncorrelated for a given survey. Our fiducial Supernova Cosmology Acceleration Probe (SNAP) survey described below, without any theoretical systematics, determines w_0 and w_a to accuracies of $\sigma(w_0) = 0.089$ and $\sigma(w_a) = 0.31$, respectively [corresponding to the pivot value determined to $\sigma(w_p) = 0.027$].

2.2 Biases in the Gaussian limit

Consider the general problem of constraining a vector of $N_{\rm par}$ cosmological parameters $\{p_i\}$ organized into a vector $P = \{p_i\}$ based on an observed data vector $D = \{D_\alpha\}$. Here, D_α are the power spectra C_{ij} from equation (2) and α runs over *pairs* of redshift bins. If the observable quantities D_α are distributed as Gaussians with covariance matrix C (defined as Cov in equation 4), then the first-order formula for bias in the ith parameter, Δp_i, induced by a bias ΔD in the data is (e.g. Knox, Scoccimarro & Dodelson 1998; Huterer & Turner 2001)

$$\Delta p_i = \sum_j (F^{-1})_{ij} \sum_{\alpha\beta} \frac{\partial \bar{D}_\alpha}{\partial p_j} (C^{-1})_{\alpha\beta}\, \Delta D_\beta. \tag{5}$$

Here, F is the Fisher matrix for the cosmological parameters defined in equation (3). The above bias can be more concisely expressed as

$$\Delta P = F^{-1} Q\, \Delta D \equiv F^{-1} V, \tag{6}$$

where we have defined the matrix Q and vector V as

$$Q_{ij} \equiv \sum_k \frac{\partial \bar{D}_k}{\partial p_i} (C^{-1})_{kj} \tag{7}$$

$$V \equiv Q\, \Delta D. \tag{8}$$

The induced parameter bias is considered unimportant if it is small compared to the expected statistical variation in the cosmological parameters. In the case where the likelihood in the parameter space is Gaussian, the likelihood of the bias ΔP being exceeded by a statistical fluctuation is determined by

$$\Delta\chi^2 = \Delta P^T F \Delta P = V^T F^{-1} V. \tag{9}$$

Here, $\Delta\chi^2$ indicates the significance of the bias in the full $N_{\rm par}$-dimensional parameter space. The usual χ^2 formulae apply; for example, in a one (two) parameter model, the bias is larger than the statistical error of 68 per cent of the time for $\Delta\chi^2 = 1.0$ (2.3).

In the Appendices, we prove two useful theorems about $\Delta\chi^2$.

(i) The bias significance $\Delta\chi^2$ always *decreases* or stays fixed when we augment the likelihood with (unbiased) prior information, e.g. data from a non-lensing technique.

(ii) $\Delta\chi^2$ always *decreases* or stays fixed when we marginalize over one or more dimensions of the parameter space. In the Gaussian limit, the bias Δp_i is unaffected by marginalization over other parameters.

We will use these results later to argue that our requirements on the control of catastrophic redshift errors are conservative, in the sense that adding other cosmological data (e.g. constraints expected from the Planck experiment) or considering individual cosmological parameter biases will only weaken the requirements.

2.3 The case of catastrophic photo-z errors

For weak-lensing tomography, the data elements D_α are the convergence (or shear) cross-power spectrum elements $C^\kappa_{\alpha\beta}(\ell)$ between photo-z bins α and β at multipole ℓ. Let us examine how these will be biased by photo-z outliers.

We assume a survey with the (true) distribution of source galaxies in redshift $n_{\rm S}(z)$, divided into some number $N_{\rm b}$ of bins in redshift. Let us define the following terms.

(i) *Leakage*: fraction of objects from a given spectroscopic bin that are placed into an incorrect (non-corresponding) photometric bin.

(ii) *Contamination:* fraction of galaxies in a given photometric bin that come from a non-corresponding spectroscopic bin.

One could estimate either of these quantities – after all, when specified for each bin, they contain the same information. Let *leakage* fraction $l_{\rm ST}$ of galaxies in some spectroscopic-redshift bin $n_{\rm S}$ (the 'source' of leakage) end up in some photo-z bin $n_{\rm T}$ (the 'target' of leakage), so that $l_{\rm ST}$ is the fractional perturbation in the source bin. Note that since bins S and T may not have the same number of galaxies, the fractional perturbation in the target bin is not necessarily the same. The *contamination* of the target bin T, $c_{\rm ST}$, is related to the source-bin leakage via

$$c_{\rm ST} = \frac{n_{\rm S}}{n_{\rm T}}\, l_{\rm ST}, \tag{10}$$

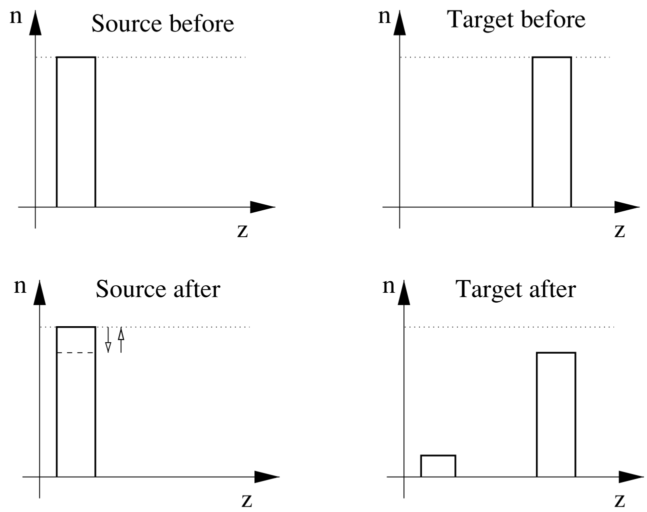

Figure 1. Explanation of how the leakage and contamination operate. In this figure, we assume for simplicity that the number of galaxies in the source and target bin is the same, so that $l_{\rm ST} = c_{\rm ST}$. Because the redshift distribution $n(z)$ is normalized to unit integral in each bin, the source bin's redshift distribution $n_{\rm S}(z)$ does not change; see the bottom left-hand panel. The target bin's redshift distribution, $n_{\rm T}(z)$, does change however, as illustrated in the bottom right-hand panel.

where $n_{\rm S}$ and $n_{\rm T}$ are the absolute galaxy numbers in the source and target bins, respectively.

The redshift distribution of galaxies (normalized to unity) in the source bin, $n_{\rm S}$, does not change since while a fraction of galaxies is lost, the redshift distribution is normalized to unit integral; see Fig. 1. Conversely, things are perturbed in the target bin, since it now contains two populations of galaxies, the original one with fraction $1 - c_{\rm ST}$ and the contamination at incorrect (source-bin) redshift with fraction $c_{\rm ST}$; again, this is clearly shown in Fig. 1. Therefore,

$$n_{\rm S}(z) \rightarrow n_{\rm S}(z) \tag{11}$$

$$n_{\rm T}(z) \rightarrow (1 - c_{\rm ST})\, n_{\rm T}(z) + c_{\rm ST} n_{\rm S}(z), \tag{12}$$

and only the *target*-bin-normalized number density is affected (i.e. biased) by photo-z catastrophic errors.

The effect on the cross-power spectra is now easy to write down. Clearly, only the (α, β) cross-spectra where one of the bins is the target bin – $\alpha = T$ or $\beta = T$ – will be affected:

$$\begin{aligned} C_{\rm TT} &\rightarrow (1 - c_{\rm ST})^2 C_{\rm TT} + 2c_{\rm ST}(1 - c_{\rm ST}) C_{\rm ST} + c_{\rm ST}^2 C_{\rm SS} \\ C_{\alpha T} &\rightarrow (1 - c_{\rm ST}) C_{\alpha T} + c_{\rm ST}\, C_{\alpha S} \qquad (\alpha \neq T) \\ C_{\alpha\beta} &\rightarrow C_{\alpha\beta} \qquad\qquad \text{(otherwise).} \end{aligned} \tag{13}$$

We have checked that ignoring the quadratic terms in $c_{\rm ST}$ leads to no observable effects on the results (for $c_{\rm ST} = 0.001$ contamination).

Equations (13) are a central result in this paper, as they show how the observable power spectra change as a result of catastrophic redshift errors. The biases in the cosmological parameters, Δp_i, can now be computed by assigning the biases in the observables (ΔD_β in equation 5) to be the differences between the right-hand side and the left-hand side in the above equations.

We can simplify our final analytical results as follows. We replace the single index α for data elements in equation (5) with the triplet $\ell\alpha\beta$ so that $D_{\ell\alpha\beta} \equiv C^\kappa_{\alpha\beta}(\ell)$, and we reserve the symbol C for the covariance of the data elements. The bias in data induced by the catastrophic errors is

$$\Delta D_{\ell\alpha\beta} = \sum_{\mu\nu} c_{\mu\nu}[\delta_{\alpha\nu}(D_{\ell\mu\beta} - D_{\ell\alpha\beta}) + \delta_{\beta\nu}(D_{\ell\mu\alpha} - D_{\ell\alpha\beta})]. \quad (14)$$

If we make the further assumption that the convergence is a Gaussian random field, then we have

$$C_{\ell\alpha\beta,\ell'\gamma\delta} = \delta_{\ell\ell'}[D_{\ell\alpha\gamma}D_{\ell\beta\delta} + D_{\ell\alpha\delta}D_{\ell\beta\gamma}] \quad (15)$$

$$\Rightarrow \quad (C^{-1})_{\ell\alpha\beta,\ell'\gamma\delta} = \frac{\delta_{\ell\ell'}}{2}\left(D_\ell^{-1}\right)_{\alpha\gamma}\left(D_\ell^{-1}\right)_{\beta\delta}. \quad (16)$$

Equation (5) simplifies considerably when we invoke equations (14) and (16):

$$\Delta p_i = \sum_{j,\mu\nu}(F^{-1})_{ij} M_{j,\mu\nu} c_{\mu\nu}, \quad (17)$$

where

$$M_{j,\mu\nu} \equiv \sum_\ell \left[(E_i^\ell)_{\mu\nu} - (E_i^\ell)_{\nu\nu}\right], \quad (18)$$

and

$$E_i^\ell \equiv \frac{\partial D_\ell}{\partial p_i} D_\ell^{-1}. \quad (19)$$

As a reminder, the Fisher matrix in the case of a zero-mean Gaussian variable is (Tegmark, Taylor & Heavens 1997)

$$F_{ij} = \frac{1}{2}\sum_\ell \mathrm{Tr}\left(E_i^\ell E_j^\ell\right). \quad (20)$$

In a cosmological application, we will marginalize over all parameters except a subset of interest A. In the Fisher approximation bias is simply projected on to the A subset: $\Delta p_A = P_A \Delta p$, where P_A is the projection matrix (see Appendix B). If F_A is the marginalized Fisher matrix, then the $\Delta\chi^2$ of the bias after marginalization is

$$\Delta\chi^2 = \sum_{\mu\nu}\sum_{\mu'\nu'} c_{\mu\nu}c_{\mu'\nu'} \sum_{ij} M_{i,\mu\nu}\left(F^{-1}P_A^T F_A P_A F^{-1}\right)_{ij} M_{j,\mu'\nu'}. \quad (21)$$

3 APPLICATION TO CANONICAL SURVEYS

For a first study we examine a weak-lensing survey similar to that proposed for SNAP (Aldering et al. 2004), but with the source–galaxy selection restricted to incur a minimal catastrophic error rate. Evaluation of other potential surveys could be performed following the same model.

We take the eight-parameter cosmological model considered by the Dark Energy Task Force (DETF; Albrecht et al. 2006): dark-energy physical density $\Omega_{\rm DE}h^2$; and equation-of-state parameters w_0 and w_a; normalization of the primordial power spectrum A and spectral index n; and matter, baryon and curvature physical densities $\Omega_{\rm M}h^2$, $\Omega_{\rm B}h^2$ and $\Omega_{\rm k}h^2$, respectively. The fiducial values of these parameters are taken from the 5-yr *Wilkinson Microwave Anisotropy Probe* (*WMAP*) data (Hinshaw et al. 2009). We will assume a Planck cosmic microwave background (CMB) prior as specified by the DETF report. Recall that application of further priors can only weaken the requirements on photo-z outliers (see Appendix A).

We assume shear tomography with 20 bins linearly spaced over $0 < z < 4$ with $\Delta z = 0.2$; we have checked that the results are stable with Δz. The redshift distribution and fiducial $c_{\alpha\beta}$ are taken from a simulation of the photo-z performance of SNAP as described in Jouvel et al. (2009). The procedure is to (1) create a simulated catalogue of galaxies, (2) calculate their noise-free apparent magnitudes in the SNAP photometric bands spanning the visible and NIR to 1.6 μm, (3) add the anticipated observational noise to each magnitude, (4) determine a best-fitting galaxy type and redshift with the template-fitting program LE PHARE[4] and (5) examine the 95 per cent confidence region $z_{\rm l} < z_{\rm p} < z_{\rm h}$ determined by LE PHARE and retain only those galaxies satisfying (Dahlen et al. 2008)

$$D95 \equiv \ln\left(\frac{1+z_{\rm h}}{1+z_{\rm l}}\right) \le 0.15. \quad (22)$$

This strict cut results in a catalogue of ≈70 galaxies per arcmin^2, with a median redshift of $z_{\rm m} \approx 1.2$. The Weak Lensing (WL) survey is assumed to cover $f_{\rm sky} = 0.1$ of the full sky, with a shape noise of $\sigma_\gamma = 0.24$ per galaxy. We consider only shear tomography at the two-point level, as this will likely maximize the bias imparted by redshift errors. We also ignore systematic errors other than redshift outliers, which will likely maximize the statistical significance of the outlier bias.

We will henceforth in this paper define a redshift outlier to satisfy

$$\left|\ln\frac{1+z_{\rm p}}{1+z_{\rm s}}\right| > 0.2 \qquad \text{(catastrophic outlier definition).} \quad (23)$$

In the simulated photo-z catalogue, 2.2 per cent of the source galaxies are outliers by this criterion. In our analyses below, we will only consider biases from photo-z errors meeting this outlier criterion, i.e. we assume that the 'core' of the photo-z distribution is well determined.

Fig. 2 illustrates the canonical model and the sensitivity to redshift outliers in this model. The left-hand panel shows the quantity $c_{\rm sp}/\Delta z$ versus $z_{\rm s}$ and $z_{\rm p}$. (We scale the contamination by Δz to produce a quantity that is independent of the choice of bin size Δz.) The highest contaminations are in two 'islands' as follows. The first at $z_{\rm p} > 2.5$, $z_{\rm s} < 0.6$ is probably due to confusion between high-z Lyman breaks and low-z 400-nm breaks. Because true $z > 2.5$ galaxies are relatively rare, a small leakage rate from $z_{\rm s} \sim 0.5$ can produce a high contamination fraction. The LE PHARE code run for this simulation does not incorporate a magnitude prior for the photo-z; doing so might reduce the size of this island.

A second high-contamination region is $z_{\rm s} \approx 1$, $z_{\rm p} < 0.2$. Again, the contamination rate is high because the target-bin density is much lower than the source-bin density.

The right-hand panel in Fig. 2 shows $\Delta\chi^2$ evaluated using equation (21) for this case of catastrophic errors. We calculate the significance $\Delta\chi^2_{\rm 2D}$ of the bias after marginalization of the cosmology on to the two-dimensional (hence 2D) w_0–w_a plane. We simplify by considering the bias arising from contamination in a single bin. This is

$$\Delta\chi^2_{\rm 2D} = (h_{\mu\nu}\Delta z)^2 c^2_{\mu\nu}, \quad (24)$$

$$h^2_{\mu\nu} \equiv \sum_{ij} M_{i,\mu\nu}\left(F^{-1}P_A^T F_A P_A F^{-1}\right)_{ij} \frac{M_{j,\mu\nu}}{(\Delta z)^2}. \quad (25)$$

Again, the inclusion of the Δz factor defines an $h_{\mu\nu}$ that is invariant under rebinning. The interpretation of $h_{\mu\nu}$ is as follows: if there is an 'island' of outliers that spans a range $\delta z_{\rm p}$ of photo-z bins, and contains a fraction $\bar{c}$ of the galaxies in these photo-z bins, then the 2D significance of the resultant bias will be

$$\sqrt{\Delta\chi^2} \approx h_{\mu\nu}\delta z_{\rm p}\bar{c}. \quad (26)$$

[4] www.oamp.fr/people/arnouts/LE PHARE.html

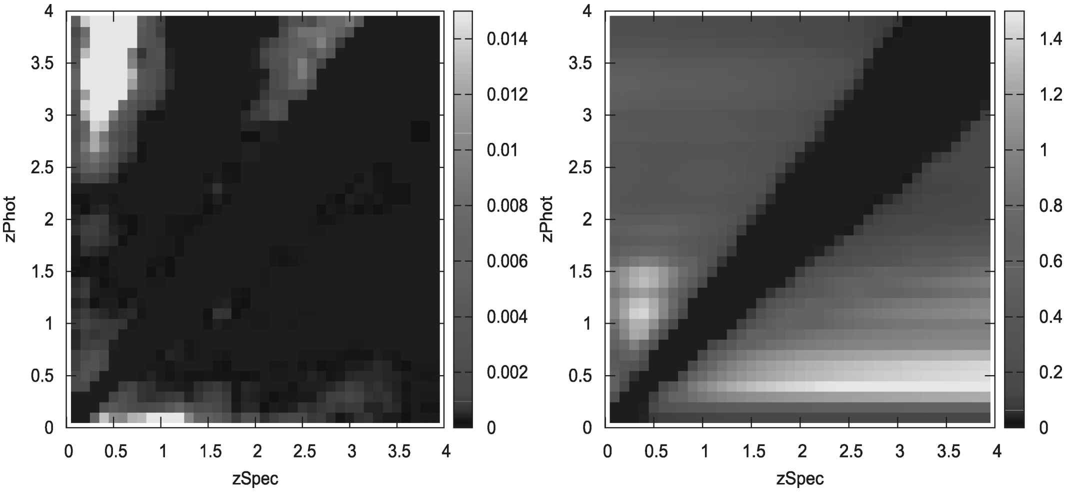

Figure 2. Left: the contamination rate $c_{\rm sp}/\Delta z_{\rm s}$ of the photo-z bin per unit redshift in spectro-z is plotted for our example case. Note that the contamination is highest at $z_{\rm p} > 2.5$ and $z_{\rm p} < 0.2$, where there are relatively few source galaxies and hence a small number of outliers can become a large fractional contamination. Right: the quantity $h_{\rm sp}$ which specifies the significance of the w_0–w_a bias caused by a contamination rate of $0.001/\delta z_{\rm p}$ across a range $\delta z_{\rm p}$ of photo-z. This plot indicates that the outlier contamination rates must be known to one to three parts per thousand over all photometric redshift bins, most sensitively at $0.3 < z_{\rm p} < 1.5$.

Fig. 2 indicates the bias significance of a contamination rate $\bar{c} = 0.001/\delta z_{\rm p}$. We desire $\Delta\chi^2_{\rm 2D} \ll 2.3$ to keep the bias well within the 68 per cent confidence contour. The most severe constraint on $\bar{c}$ would be to take the peak value $h_{\mu\nu} \approx 1300$ and a very wide island, $\delta z_{\rm p} \approx 0.5$, in which case the criterion for small bias becomes

$$\bar{c} < 1/(h_{\mu\nu}\delta z_{\rm p}) \approx 1/(1300 \times 0.5) = 0.0015. \qquad (27)$$

The contamination rate into any island of outliers must be known to 0.0015 or better to avoid significant cosmological bias. This conclusion is independent of the nominal outlier rate. The tolerance on the outlier rate will scale with sky coverage as $f_{\rm sky}^{-1/2}$.

3.1 Photo-z probability distributions

Throughout this paper we have assumed that the photo-z algorithm assigns a single $z_{\rm p}$ to each galaxy, and that the WL analysis bins galaxies on the basis of this estimate. Most photo-z algorithms will, however, produce some posterior likelihood distribution $P(z_{\rm p})$ over redshift. For many astrophysical measurements, selection of a single maximum-likelihood $z_{\rm p}$ will induce biases that can be alleviated by properly considering the full $P(z_{\rm p})$ distribution for the galaxies placed into a bin (Padmanabhan et al. 2007; Mandelbaum et al. 2008). The results of this paper can be applied to an analysis that retains $P(z_{\rm p})$ information as follows: first, we assume that some criterion has been used to divide the galaxies into WL tomography bins, labelled by index T. Then the $P(z_{\rm p})$ distributions of all galaxies in bin T are summed to give the overall estimated distribution $P_{\rm T}(z_{\rm p})$ of the galaxies in the bin. The estimated contamination fraction $\hat{c}_{\rm ST}$ is the integral of $P_{\rm T}(z_{\rm p})$ over a source redshift bin near $z_{\rm s}$. The tolerances on $c_{\rm ST}$ in this paper apply to how accurately these estimates from the photo-z posterior must match the true contamination rates. All the equations derived below for validation via spectroscopic sampling should still apply to the use of photo-z distributions.

Various selection criteria might be applied to galaxies on the basis of their $P(z_{\rm p})$, with the intent of reducing the outlier bias – our D95 cut (equation 22) being just one example. The requirements for calibration of the outlier rate will apply to the galaxy sample *after* such cuts or weights are applied, and of course the calibration requirements will in general depend on these cuts and weights.

4 CONSTRAINT VIA SPECTROSCOPIC SAMPLING

The most obvious way to determine the contamination rate $c_{\alpha\beta}$ is to conduct a *complete* spectroscopic-redshift survey of galaxies in photo-z bin β. It is of course essential that the spectra be of sufficient quality to determine redshifts even for the outliers in the sample.

Let us now estimate the total number of spectra $N_{\rm spec}$ required in order to keep the total bias below some desired threshold. We will assume that each redshift drawn from the spectroscopic survey is statistically independent. In this case the distribution of $N_{\alpha\beta}$, the number of galaxies in photo-z bin β that have spectro-z in bin α, will be described by a multinomial distribution. When the outlier rates are small, the number of spectra in each outlier bin tends towards independent Poisson distributions.

We would like to relate the contamination uncertainties δc_{ij} to the required number of galaxy spectra. Let N_β be the number of spectra drawn from the photometric redshift bin β so that $N_{\rm spec} = \sum_\beta N_\beta$. In this case, $\langle N_{\alpha\beta}\rangle = c_{\alpha\beta}N_\beta$ and the variance of the contamination estimate is

$$\delta c^2_{\alpha\beta} = \frac{(\delta N_{\alpha\beta})^2}{\langle N_\beta\rangle^2} = \frac{c_{\alpha\beta}}{N_\beta}. \qquad (28)$$

Since the Poisson errors between different outlier bins are uncorrelated, the expected bias significance becomes

$$\langle\Delta\chi^2\rangle = \sum_{\alpha\beta}(h_{\alpha\beta}\Delta z)^2\left\langle\delta c^2_{\alpha\beta}\right\rangle = \sum_{\alpha\beta}(h_{\alpha\beta}\Delta z)^2 c_{\alpha\beta}/N_\beta. \qquad (29)$$

We would like to quote a total number of spectroscopic redshifts rather than the number per photo-z slice (N_β above) in order to make our findings more transparent. We consider two cases: first, a slitless or untargeted spectroscopic survey will obtain redshifts in proportion to the number density N_β of source galaxies in each redshift bin: $N_\beta = N_{\rm spec}n_\beta/n$. Then we will consider a targeted survey, in which the number N_β can be chosen bin by bin to produce the minimal bias for given total $N_{\rm spec}$.

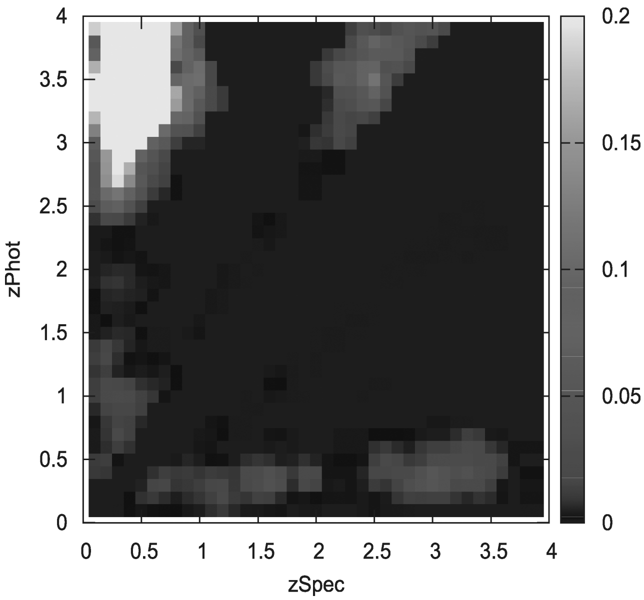

Figure 3. The number of calibration spectra required to attain an outlier bias significance of $\Delta\chi^2_{2D} = 1$ is given by the integral of this plot over the z_s–z_p space. The quantity plotted is, from equation (31), $h^2_{sp}c_{sp}n/n_p$, where n_p is the source density in the photo-z bin and n is the total source density. The plot is in units of 10^6 galaxies. Note that the required number of calibration spectra is strongly driven by the need to constrain the outliers with photo-z $z_p > 2.5$ but true redshifts $z_s < 0.8$. Note further that this plot is essentially the product of the left-hand panel and the (square of the) right-hand panel of Fig. 2.

4.1 Untargeted spectroscopic survey

In the untargeted case, the number of galaxies to follow up in each bin is proportional to the number density in that bin:

$$N_\beta = N_{\rm spec} n_\beta / n, \tag{30}$$

and the condition $\Delta\chi^2 \ll 1$ becomes (from equation 29)

$$N_{\rm spec} \gg \sum_{\alpha\beta} (\Delta z)^2\, h^2_{\alpha\beta} c_{\alpha\beta} n / n_\beta. \tag{31}$$

Fig. 3 plots the summand of this expression in the z_s–z_p plane. The required $N_{\rm spec}$ is hence the sum over values in this plane (note that we omit the bins near the diagonal that do not meet the 'outlier' definition). We find that $\Delta\chi^2_{2D} \lesssim 1$ requires $N_{\rm spec} \gtrsim 8 \times 10^5$, and in the full 8D parameter space (i.e. considering $\Delta\chi^2$ for the 8D parameter ellipsoid), we need $N_{\rm spec} \gtrsim 2 \times 10^6$.

These requirements are daunting, potentially larger than the $N_{\rm spec}$ that are needed to constrain the core of the photo-z distribution as determined by Ma & Bernstein (2008). Note however that the requirement is strongly driven by the region $z_s > 2.5$, $z_p < 0.8$. This is because contamination rates are high here in the nominal photo-z distribution (Fig. 2), and these bins are sparsely populated (small n_p), meaning that many spectra must be taken in order to accumulate a strong enough constraint on these contamination coefficients.

This suggests a strategy of omitting the $z_p > 2.5$ galaxies from the tomography analysis entirely. Omitting $z_p > 2.5$ from the sum (equation 31) produces a much weaker requirement: for $\Delta\chi^2_{2D} \lesssim 1$, we need $N_{\rm spec} \gtrsim 2.8 \times 10^4$, a 30-fold reduction. This strategy would eliminate ≈8 per cent of the source galaxies in the SNAP model and reduce the dark-energy constraint power by 18 per cent, as measured by the DETF figure of merit (FoM). This would be an acceptable strategy to reduce the outlier bias if one were unable to eliminate the high-z_p island of outliers by refinements to the photo-z methodology.

4.2 Targeted spectroscopic survey

If we wish to minimize $\Delta\chi^2$ in equation (29) for a given total $N_{\rm spec}$, a simple optimization yields

$$N_\beta \propto \sqrt{\sum_\alpha h^2_{\alpha\beta} c_{\alpha\beta}}, \tag{32}$$

and thus

$$N_{\rm spec}\Delta\chi^2 = \left(\sum_\beta \Delta z \sqrt{\sum_\alpha h^2_{\alpha\beta} c_{\alpha\beta}} \right)^2. \tag{33}$$

Optimized targeting reduces the requirement for $\Delta\chi^2_{2D} \lesssim 1$ to be $N_{\rm spec} \gtrsim 1.2 \times 10^5$. Eliminating the $z_p > 2.5$ sources reduces the requirement sixfold, $N_{\rm spec} \gtrsim 2.0 \times 10^4$.

Note that the targeted redshift requires seven times fewer calibration redshifts than the untargeted survey, if we are using the full source redshift range, but only 1.4 times smaller $N_{\rm spec}$ if we restrict $z_p < 2.5$ in the lensing analysis.

4.3 Scaling and robustness

The required $N_{\rm spec}$ to reduce outlier-rate biases to insignificance scales with the sky coverage and the mean outlier rate as

$$N_{\rm spec} \propto f_{\rm sky}\bar{c} \tag{34}$$

when most of the information is coming from shear tomography, and the depth of the survey is held fixed. We have verified that $N_{\rm spec}$ varies little as the number of tomography bins grows ($\Delta z < 0.2$) and the information content of the tomography saturates. The two bias theorems imply that the required $N_{\rm spec}$ will drop if we add additional unbiased prior information or if we marginalize down to a single dark-energy parameter.

More precisely, we find that the ratio $N_{\rm spec} \times [\sigma(w_p) \times \sigma(w_a)] \equiv N_{\rm spec}/\mathrm{FoM}$, featuring the well known 'FoM' (Huterer & Turner 2001; Albrecht et al. 2006), is roughly the same with several alternative survey specifications we consider (and is exactly the same if only the sky coverage $f_{\rm sky}$ is varied).

We have used two independent codes to verify the robustness of results to the myriad of assumptions made and check for the presence of unwanted numerical artefacts. The two codes agree to roughly a factor of 2 in $N_{\rm spec}$, which is satisfactory given the differences between the implementations, e.g. whether curvature and/or neutrino masses are free to vary and whether the fiducial redshift distribution is smoothed over the cosmic variance in the simulated catalogue.

4.4 Correlated outlier errors and incompleteness

So far we have considered the case where bin-to-bin fluctuations in contamination $\delta c_{\alpha\beta}$ are uncorrelated. Contamination-rate errors $\delta c_{\alpha\beta}$ that are *correlated* from bin to bin might arise if the spectroscopic survey systematically misses outliers in certain redshifts islands (or if the spectroscopy is not done at all!). We can set a specification on the maximum allowable systematic contamination-rate error $\delta\bar{c}$ in an island of photo-z width δz_p using equation (26). Our previous results for the canonical survey show that the contamination rate in the island should be known to

$\delta\bar{c} \lesssim 0.0015$. In other words, a spectroscopic-redshift failure rate of only 0.15 per cent in some range of z_p can cause a significant cosmology bias if all of these missed redshift are outliers in a particular island. A 99.9 per cent success rate has rarely if ever been achieved in a spectroscopic-redshift survey.

5 CONSTRAINING OUTLIER RATES USING GALAXY CORRELATIONS

The above requirements on $N_{\rm spec}$ and completeness may be too expensive to accomplish, particularly for fainter galaxies. We now examine the possibility that one could make use of a spectroscopic galaxy sample that does *not* fairly sample the photo-z galaxies (Newman 2008). The idea is to cross-correlate a photo-z sample at nominal redshift z_p with a spectroscopic sample known to be confined to a distinct bin z_s. The amplitude of this cross-correlation will tell us something about the contamination rate $c_{\rm sp}$, since there is no intrinsic correlation between the galaxy densities at the two disparate redshifts.

Newman (2008) calculates the errors in this estimate that would be induced by shot noise in the sample (for a somewhat related work, see Schneider et al. 2006). Here we assume that statistical errors will be negligible and attend to two systematic errors that will arise.

5.1 Outlier bias and correlation coefficients

First, we define g_s and g_p to be the fluctuations in the sky density of the projected distributions of the spectroscopic-survey galaxies in bin s and the photometric-redshift galaxies in bin p. We set m_s to be the fractional fluctuations in the projected mass density in redshift bin s. The bias is defined by $\langle g_s^2\rangle = b_s^2\langle m_s^2\rangle$. The amplitude of the measured cross-correlation between the angular distributions g_s and g_p of the spectroscopic and photometric samples can be written as

$$\langle g_s g_p\rangle = c_{\rm sp} b_s b_{\rm sp} r_{\rm sp}\left\langle m_s^2\right\rangle = c_{\rm sp}\frac{b_{\rm sp} r_{\rm sp}}{b_s}\left\langle g_s^2\right\rangle, \quad (35)$$

where $c_{\rm sp}$ is the contamination rate, $b_{\rm sp}$ is the bias of the *outlier population* and $r_{\rm sp}$ is the correlation coefficient between the density fluctuations of the spectroscopic and the outlier populations. The outlier population is those minority of sources in photo-z bin p that have spectroscopic redshifts in bin s.

In a large survey, shot noise in $\langle g_s g_p\rangle$ and in $\langle g_s^2\rangle$ might become small, and b_s could be determined to good accuracy through a study of the redshift-space power spectrum of the spectroscopic targets. There is, however, no reliable way to infer the bias $b_{\rm sp}$ of the outlier population. The photometric data do not allow us to isolate the outlier population, so we cannot measure the angular autocorrelation of outliers in a (z_s, z_p) cell without being overwhelmed by the correlations of the galaxies in the 'core' of the photo-z error distribution. We also cannot assume that the outliers have the same bias as the other galaxies in their source redshift bin, since the outliers are not necessarily a fair sample. So in general, $b_p \neq b_{\rm sp} \neq b_s$.

The correlation coefficient $r_{\rm sp}$ also has no alternative observable signature we have identified. Hence, there will be a systematic-error floor on $\delta c_{\rm sp}$ arising from the finite a priori knowledge of the product $b_{\rm sp} r_{\rm sp}$.

5.2 Lensing magnification

The second complication to the cross-correlation method is that gravitational lensing magnification bias will induce a correlation between the spectroscopic and photometric samples even if there is *no* contamination. Let us assume that the spectroscopic sample is in the foreground of the photo-z 'core'; a similar analysis can be done when using cross-correlation to search for contamination by background galaxies.

There are two types of magnification-induced correlations. First, it is possible that the spectroscopic sample at z_s and the photo-z sample at z_p are both being lensed by a foreground mass distribution at $z_l < z_s, z_p$. Following the notation of Hirata & Seljak (2004), this would be called a 'GG' correlation. The apparent densities of both z_s and z_p samples are modulated by lensing magnification bias; hence, they are correlated.

The second type of lensing-induced correlation is the 'GI' contribution. In the case $z_s < z_p$, the photo-z sample is lensed by mass fluctuations at redshift $z_l = z_s$. The lensing modulates the sky density of the background z_p sample, while the foreground z_s galaxy density is correlated with its local mass density and hence with the induced magnification bias. The GI effect is also present for $z_p < z_s$, due to mass at $z_l = z_p$.

The GI correlation is as follows: Let the spectroscopic bin z_s span a range $\Delta\chi_s$ in comoving radial distance. The matter fluctuations m_s induce a lensing convergence on the photo-z bin at z_p of

$$\kappa_p = \frac{3\omega_m}{2}\Delta\chi_s\frac{\chi_p - \chi_s}{\chi_p} m_s, \quad (36)$$

where $\omega_m = \Omega_m h^2 = 0.127$ is the comoving matter density, χ_p and χ_s are the comoving angular diameter distances to z_s and z_p, respectively, and we have assumed a flat Universe.

The lensing magnification will induce apparent density fluctuations in the background sample as

$$g_p^{\rm lens} = q_p\kappa_p, \quad (37)$$

where q_p is a magnification-bias factor for the galaxies in the photo-z bin. For instance if the selection criteria for the bin were a simple flux limit, and the intrinsic flux distribution were a power law $dn/df \propto f^{-s}$, then we would have $q_p = 2s - 2$. In general, q_p will be of the order of unity.

The foreground galaxy distribution g_s has a correlation coefficient r_s with the mass m_s; hence, a covariance between populations results:

$$\langle g_s g_p\rangle_{\rm GI} = \frac{3\omega_m}{2}\Delta\chi_s\frac{\chi_p - \chi_s}{\chi_p} q_p b_s r_s\left\langle m_s^2\right\rangle \quad (38)$$

$$= \frac{3\omega_m}{2}\Delta\chi_s\frac{\chi_p - \chi_s}{\chi_p} q_p\frac{r_s}{b_s}\left\langle g_s^2\right\rangle \quad (39)$$

(we have ignored shot noise in the galaxy autocorrelation). This lensing contamination will have to be subtracted from $\langle g_s g_p\rangle$ in order to extract the information on contamination $c_{\rm sp}$. Even if all the cosmological factors are well determined, the magnification coefficient q_p will have to be empirically estimated. Finite accuracy in this estimate will increase $\delta c_{\rm sp}$.

The GG correlation scales as follows: let κ_s be the convergence induced on the foreground (spectroscopic) sample by mass at $z < z_s$. This produces density fluctuations $g_s = q_s\kappa_s$. This mass induces convergence $\kappa_p \approx \kappa_s r(\chi_s, \chi_p)$ on the background (photo-z) source population, where r is an integral involving the distributions of foreground mass which must satisfy $r \geq 1$. Not concerning ourselves with details, we take $r = 1$. The induced angular correlation will be

$$\langle g_s g_p\rangle_{\rm GG} = q_s q_p r\left\langle\kappa_s^2\right\rangle. \quad (40)$$

Typical RMS values κ_s are 0.01–0.02 at cosmological distances. The GG lensing correlation must be removed from the signal to retrieve the contamination fraction, and again even if there is no shot noise and all distances and lensing amplitudes are known perfectly, the values of q_s and q_p will only be known to finite precision.

5.3 Estimate of systematic errors

Summing the GG, GI and intrinsic contributions, the cross-correlation between spectroscopic and photometric samples is

$$\langle g_s g_p \rangle = \left\{ \frac{3\omega_m}{2} \Delta\chi_s \frac{\chi_p - \chi_s}{\chi_p} q_p \frac{r_s}{b_s} + c_{sp} \frac{b_{sp} r_{sp}}{b_s} \right\} \left\langle g_s^2 \right\rangle + q_s q_p r \left\langle \kappa_s^2 \right\rangle \quad (41)$$

$$\Rightarrow c_{sp} = \frac{1}{b_{sp} r_{sp}} \left[\frac{b_s \langle g_s g_p \rangle}{\left\langle g_s^2 \right\rangle} - \frac{b_s q_s q_p \left\langle \kappa_s^2 \right\rangle}{\left\langle g_s^2 \right\rangle} - \frac{3\omega_m}{2} \Delta\chi_s \frac{\chi_p - \chi_s}{\chi_p} r_s q_p \right]. \quad (42)$$

All of the right-hand quantities are potentially well measured from the survey data itself or from other cosmological probes, except the outlier covariance factor $b_{sp} r_{sp}$ and the magnification coefficients q_p and q_p. Uncertainties in the a priori assumed values of these factors will propagate into the contamination coefficient as

$$(\delta c_{sp})^2 \approx \left[\delta(b_{sp} r_{sp})\right]^2 c_{sp}^2 + \delta q_p^2 \left(\frac{3\omega_m}{2} \Delta\chi_s \frac{\chi_p - \chi_s}{\chi_p} \right)^2 + \left(\delta q_p^2 + \delta q_s^2\right) \left(\frac{q \left\langle \kappa_s^2 \right\rangle}{\left\langle g_s^2 \right\rangle} \right)^2. \quad (43)$$

Here we assume $b \approx r \approx 1, q_s \approx q_p$.

Earlier we showed that contamination into an outlier 'island' should be known to $\delta c_{sp} \leq 0.0015$ to avoid significant parameter bias. Can such a small contamination be measured using the cross-correlation technique?

(i) If the nominal outlier rate is $c_{sp} \approx 10\delta c_{sp} \approx 0.015$, then we require a prior estimate of outlier bias/covariance accurate to $\delta(b_{sp}\ r_{sp}) < 0.1$. Little will be known about the outlier population besides its luminosity range, and the outliers may tend to be active galaxies or those with unusual spectra whose clustering properties might be deviant as well. We would consider a 10 per cent prior knowledge on outlier bias to be optimistic but perhaps attainable.

(ii) For the second (GI) term, if we take the distance factors to be $\approx$1, and the outlier population to span a range $\Delta\chi_s \approx 0.3$, then the magnification-bias coefficient must be known to an accuracy of $\delta q_p \lesssim 0.025$. This accuracy in q_p will be challenging to achieve. If the galaxy selection is by a simple magnitude cut, then the slope of the counts yields q_p and potentially could be measured to high precision. Weak-lensing samples typically have more complex cuts and weightings, however, than a simple flux cut-off. Surface brightness, photo-z accuracy and ellipticity errors are involved, making estimation of q_p more difficult.

(iii) The third (GG) term places constraints on δq_p and δq_s that will generally be weaker than those from the GI term.

If the cross-correlation technique is to determine outlier contamination fractions to an accuracy that renders them harmless, then we will need to know the $b_{sp} r_{sp}$ product of the outlier population to 0.1 or better, and also must know the magnification-bias coefficients q of our populations to 2 per cent accuracy. This is true regardless of the sample size, and these tolerances will scale as $f_{sky}^{-1/2}$. The demands on $\delta(b_{sp} r_{sp})$ also become more stringent linearly with the photo-z outlier rate.

We have not considered the possibility of extraneous angular correlations induced by dust correlated with the foreground galaxy sample or by dust in front of both samples (Ménard et al. 2009). This signal will be present to some degree, and may perhaps be diagnosed with colour information.

In summary, while we have shown here that the cross-correlation technique proposed by Newman (2008) is sensitive to catastrophic redshift errors, we found that measuring these errors (i.e., the contamination coefficients c_{sp}) will be difficult using this technique alone.

6 DISCUSSION AND CONCLUSION

In this paper, we have considered the effects of the previously ignored catastrophic redshift errors – cases when the photometric redshift is grossly mis-estimated, i.e. when $|z_p – z_s| \sim O(1)$, and are represented by arbitrary 'islands' in the z_p–z_s plane. We developed a formalism, captured by equations (13), that treats these islands as small 'leakages' (or 'contaminations') and directly estimates their effect on bias in cosmological parameters. We then inverted the problem by estimating how many spectroscopic redshifts are required to control catastrophic errors at a level that makes them harmless for cosmology. In the process, we have proven two general-purpose theorems (in the Appendices): the bias due to systematics always decreases or stays fixed if (1) (unbiased) prior information is added to the fiducial survey or (2) we marginalize over one or more dimensions of the parameter space.

We found that, at face value, of order million redshifts are required in order not to bias the dark-energy parameter measurements (i.e. in order to lead to $\Delta\chi^2 \lesssim 1$ in the w_0–w_a plane). However, the requirement becomes significantly (30 times) less stringent if we restrict the survey to redshift $z < 2.5$; in that case, N_{spec} is only of the order of a few tens of thousands. Essentially, leakage of galaxies from lower redshift to $z > 2.5$ is damaging since there are few galaxies at such high redshift and the relative bias in the galaxy number is large. Therefore, using only galaxies with $z \lesssim 2.5$ helps dramatically by lowering the required N_{spec} while degrading the dark-energy (FoM) constraints by mere ~20 per cent.

We have studied two approaches for a spectroscopic survey: the untargeted one where the number of spectra at each redshift is proportional to the number of photometric galaxies (Section 4.1) and the targeted one where the number of spectra is optimized to be minimal for a given degradation in cosmological parameters (Section 4.2). For the case where galaxies with $z_p > 2.5$ are dropped, the targeted survey gave only a modestly (~40 per cent) smaller required N_{spec}.

We do not imply that these N_{spec} requirements apply to all proposed surveys to high accuracy, although the $O(10^{-3})$ required knowledge on catastrophic rates is robust. The calculation should be repeated with the fiducial photo-z outlier distribution, survey characteristics and cosmological parameters of interest to a particular experiment.

Our work demonstrates for example that efforts to reduce the 'island' of catastrophic mis-assignment from $z \approx 0$ to $z \approx 3$, such as magnitude priors, could greatly reduce the required N_{spec}. Since $N_{spec} \propto f_{sky}\bar{c}$ (with $\bar{c}$ being the mean rate of catastrophic contamination), it is clear that a photo-z survey with an improved signal-to-noise ratio (S/N) and wavelength coverage to reduce the total catastrophic rate will also require lower N_{spec} to calibrate these rates.

Another practical implication of these results is that the spectroscopic-redshift surveys must be of very high completeness –99.9 per cent if there is a possibility that all failures could be in an outlier island, but less if some fraction of the failures are known to be in the core of the error distribution.

If an outlier island is known to exist at a particular (z_s, z_p) location, it may be possible to include the contamination c_{sp} as a free parameter in the data analysis and marginalize over its value. It is possible that self-calibration may reduce the bias in cosmological parameters. It is likely infeasible, however, to leave the c_{sp} values over the full (z_s, z_p) plane as free parameters. We leave self-calibration of outlier rates for future work.

We have also studied whether the technique proposed by Newman (2008), which correlates a photometric sample with a spectroscopic one, can be used to measure, and thus correct for, catastrophic redshift errors. The advantage of this approach is that the spectroscopic survey need not be a representative sampling of the photometric catalogue. While we found that the cross-correlation technique is sensitive to catastrophic errors (specifically, the contamination coefficients c_{sp}), the contamination coefficient is degenerate with the value $b_{sp}r_{sp}$ of the bias and stochasticity of the outlier population. Furthermore, there is a correlation induced by lensing magnification bias that spoofs the contamination signal. It will therefore be difficult to use the cross-correlation technique to constrain outlier rates to the requisite accuracy.

Overall, we are very optimistic that the catastrophic redshift errors can be controlled to the desired accuracy. We have identified a simple strategy that requires only of order 30 000 spectra out to $z \simeq 2.5$ for the calibration to be successful for a SNAP-type survey. Coincidentally, this number of spectra required for the catastrophic errors is of the same order of magnitude as that required for the non-catastrophic, 'core' errors (Ma et al. 2005; Huterer et al. 2006; Ma & Bernstein 2008). The total spectroscopic requirements of a survey will be based on the greater of requirements of these two error regimes.

ACKNOWLEDGMENTS

GB is supported by grant AST-0607667 from the NSF, DOE grant DE-FG02-95ER40893 and NASA grant BEFS-04-0014-0018. DH is supported by the DOE OJI grant under contract DE-FG02-95ER40899, NSF under contract AST-0807564 and NASA under contract NNX09AC89G. We would like to thank Jeff Newman for useful discussions.

REFERENCES

Abdalla F. B., Amara A., Capak P., Cypriano E. S., Lahav O., Rhodes J., 2008a, MNRAS, 387, 969
Abdalla F. B., Banerji M., Lahav O., Rashkov V., 2008b, preprint (arXiv:0812.3831)
Albrecht A. et al., 2006, preprint (astro-ph/0609591)
Albrecht A. et al., 2009, preprint (arXiv:0901.0721)
Aldering G. et al., 2004, preprint (astro-ph/0405232)
Amara A., Réfrégier A., 2008, MNRAS, 391, 228
Banerji M., Abdalla F. B., Lahav O., Lin H., 2008, MNRAS, 386, 1219
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Bernstein G. M., 2009, ApJ, 695, 652
Bridle S., Abdalla F. B., 2007, ApJ, 655, L1
Bridle S., King L., 2007, New J. Phys., 9, 444
Crittenden R. G., Natarajan P., Pen U.-L., Theuns T., 2001, ApJ, 559, 552
Croft R. A. C., Metzler C. A., 2000, ApJ, 545, 561
Cunha C. E., Lima M., Oyaizu H., Frieman J., Lin H., 2009, MNRAS, 396, 2379
Dahlen T., Mobasher B., Jouvel S., Kneib J.-P., Ilbert O., Arnouts S., Bernstein G., Rhodes J., 2008, ApJ, 136, 1361
Guzik J., Bernstein G., 2005, Phys. Rev. D, 72, 043503
Hagan B., Ma C.-P., Kravtsov A. V., 2005, ApJ, 633, 537
Heavens A., Refregier A., Heymans C., 2000, MNRAS, 319, 649
Heitmann K., Ricker P. M., Warren M. S., Habib S., 2005, ApJS, 160, 28
Heitmann K., White M., Wagner C., Habib S., Higdon D., 2008, preprint (arXiv:0812.1052)
Heymans C., Heavens A., 2003, MNRAS, 339, 711
Heymans C. et al., 2006, MNRAS, 368, 1323
Hinshaw G. et al., 2009, ApJS, 180, 225
Hirata C. M., Seljak U., 2004, Phys. Rev. D, 70, 063526
Hirata C. M. et al., 2004, MNRAS, 353, 529
Hoekstra H., Jain B., 2008, Annu. Rev. Nucl. Particle Science, 58, 99
Huterer D., 2002, Phys. Rev. D, 65, 063001
Huterer D., Takada M., 2005, Astropart. Phys., 23, 369
Huterer D., Takada M., Bernstein G., Jain B., 2006, MNRAS, 366, 101
Huterer D., Turner M. S., 2001, Phys. Rev. D, 64, 123527
Huterer D., White M. J., 2005, Phys. Rev. D, 72, 043002
Jain B., Jarvis M., Bernstein G., 2006, J. Cosmology Astropart. Phys., 0602, 001
Jarvis M., Jain B., 2004, preprint (astro-ph/0412234)
Jing Y. P., 2002, MNRAS, 335, L89
Joachimi B., Schneider P., 2008, A&A, 488, 829
Jouvel S. et al., 2009, preprint (arXiv:0902.0625)
King L. J., Schneider P., 2003, A&A, 398, 23
Kitching T. D., Amara A., Abdalla F. B., Joachimi B., Refregier A., 2008, preprint (arXiv:0812.1966)
Kitching T. D., Taylor A. N., Heavens A. F., 2008, MNRAS, 389, 173
Knox L., Scoccimarro R., Dodelson S., 1998, Phys. Rev. Lett., 81, 2004
Lima M., Cunha C. E., Oyaizu H., Frieman J., Lin H., Sheldon E. S., 2008, MNRAS, 390, 118
Ma Z., Bernstein G., 2008, ApJ, 682, 39
Ma Z., Bernstein G., Weinstein A., Sholl M., 2008, PASP, 120, 1307
Ma Z.-M., Hu W., Huterer D., 2005, ApJ, 636, 21
Mackey J., White M. J., Kamionkowski M., 2002, MNRAS, 332, 788
Mandelbaum R., Hirata C. M., Ishak M., Seljak U., Brinkmann J., 2006, MNRAS, 367, 611
Mandelbaum R. et al., 2008, MNRAS, 386, 781
Massey R. et al., 2007, MNRAS, 376, 13
Ménard B., Scranton R., Fukugita M., Richards G., 2009, preprint (arXiv:0902.4240)
Munshi D., Valageas P., Van Waerbeke L., Heavens A., 2008, Phys. Rep., 462, 67
Newman J. A., 2008, preprint (arXiv:0805.1409)
Oyaizu H., Lima M., Cunha C. E., Lin H., Frieman J., 2008, ApJ, 689, 709
Padmanabhan N. et al., 2007, MNRAS, 378, 852
Paulin-Henriksson S., Amara A., Voigt L., Refregier A., Bridle S. L., 2008, A&A, 484, 67
Rudd D. H., Zentner A. R., Kravtsov A. V., 2008, ApJ, 672, 19
Shapiro C., 2009, ApJ, 696, 775
Shapiro C., Cooray A., 2006, J. Cosmology Astropart. Phys., 0603, 007
Schneider M., Knox L., Zhan H., Connolly A., 2006, ApJ, 651, 14
Stabenau H. F., Jain B., Bernstein G., Lampton M., 2007, preprint (arXiv:0710.3355)
Takada M., Bridle S., 2007, New J. Phys., 9, 446
Takada M., Jain B., 2009, MNRAS, 395, 2065
Tegmark M., Taylor A., Heavens A., 1997, ApJ, 480, 22
White M. J., 2004, Astropart. Phys., 22, 211
Zentner A. R., Rudd D. H., Hu W., 2008, Phys. Rev. D, 77, 043507
Zhan H., Knox L., 2004, ApJ, 616, L75

APPENDIX A: EFFECT OF UNBIASED PRIORS ON BIAS SIGNIFICANCE

Will a bias get worse or better (more or less significant) when additional unbiased prior information is added to the likelihood? It is intuitive that biases ΔP should decrease when unbiased information is added. However $F \to \mathrm{F} + G$ for some new non-negative-definite prior Fisher matrix G, meaning that the statistical errors also shrink. So which effect wins out? We prove here that addition of unbiased prior information *cannot increase* the significance of parameter bias.

The proof is straightforward: equation (9) gives the significance $\Delta\chi^2$ of a bias in terms of the original positive-definite Fisher matrix F and the vector V. If G is a non-negative-definite prior, then the change significance of the bias is

$$\Delta\chi^2(\text{with prior}) - \Delta\chi^2(\text{without prior}) = V^T(F+G)^{-1}V - V^T F^{-1} V. \tag{A1}$$

This quantity cannot be positive. If it were, then there would some $0 < \lambda_0 < 1$ and a positive-definite matrix $H = F + \lambda_0 G$ such that

$$0 < \frac{\partial}{\partial\lambda}\left[V^T(F+\lambda G)^{-1}V\right]_{\lambda_0} = -(H^{-1}V)^T G H^{-1} V. \tag{A2}$$

If G is non-negative definite, this situation cannot occur. We hence conclude that the $\Delta\chi^2$ of some bias is always reduced (or stays the same) by addition of an unbiased prior.

APPENDIX B: EFFECT OF MARGINALIZATION ON BIAS SIGNIFICANCE

We can ask the following question: if we calculate the significance of a bias induced over a parameter space, then marginalize away parameter vector B to leave parameter vector A, how might the significance differ in the smaller space? We show that in the Gaussian limit, *marginalization always reduces (or leaves unchanged) the $\Delta\chi^2$ assigned to the bias, although the $\Delta\chi^2$ per DOF may increase*. To see this, first we note that marginalization over B does not change the biases in the parameters A if the distribution is Gaussian. So the bias in A is simply a projection matrix P_A times ΔP: $\Delta P_A = P_A F^{-1} V$. The $\Delta\chi^2$ after marginalization down to the A space is determined by the marginalized Fisher matrix, $F' = [(F^{-1})_{AA}]^{-1}$. So we have

$$(\Delta\chi^2)_A = V^T F^{-1} P_A^T F' P_A F^{-1} V$$
$$= V^T F^{-1} V - (P_B V)^T (F_{BB})^{-1} (P_B V) \tag{B1}$$

$$= \Delta\chi^2 - (P_B V)^T (F_{BB})^{-1} (P_B V). \tag{B2}$$

The equivalence in equation (B.1) can be derived from manipulation of the common expression for the inverse of a matrix decomposed into a 2×2 array of submatrices. Because F_{BB} and its inverse must be non-negative definite, the last term is negative, so we are assured that $(\Delta\chi^2)_A \le \Delta\chi^2$. Equality is, however, easily obtained, e.g. if there is no bias in the B parameters. We thus know that $\Delta\chi^2/N_{\mathrm{DOF}}$ can potentially increase.

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