CATASTROPHIC PHOTO-\textit{z} ERRORS AND THE DARK ENERGY PARAMETER ESTIMATES WITH COSMIC SHEAR

LEI SUN\textsuperscript{1,2}, ZU-HUI FAN\textsuperscript{1}, CHARLING TAO\textsuperscript{2}, JEAN-PAUL KNEIB\textsuperscript{3}, STÉPHANIE JOUVEL\textsuperscript{3}, and ANDRÉ TILQUIN\textsuperscript{2}

\textsuperscript{1} Department of Astronomy, Peking University, Beijing 100871, China
\textsuperscript{2} Centre de Physique des Particules de Marseille, CNRS/IN2P3-Luminy and Université de la Méditerranée, Case 907, F-13288 Marseille Cedex 9, France
\textsuperscript{3} Laboratoire d’Astrophysique de Marseille, UMR 6110, CNRS-Université de Provence-38 rue Frédéric Joliot-Curie, 13013 Marseille, France; sunl@bac.pku.edu.cn

ABSTRACT

We study the impact of catastrophic errors occurring in the photometric redshifts of galaxies on cosmological parameter estimates with cosmic shear tomography. We consider a fiducial survey with nine-filter set and perform photo-\textit{z} measurement simulations. It is found that a fraction of 1% galaxies at $z_{\text{spec}} \sim 0.4$ is misidentified to be at $z_{\text{phot}} \sim 3.5$. We then employ both $\chi^2$ fitting method and the extension of Fisher matrix formalism to evaluate the bias on the equation of state parameters of dark energy, $w_0$ and $w_a$, induced by those catastrophic outliers. By comparing the results from both methods, we verify that the estimation of $w_0$ and $w_a$ from the fiducial five-bin tomographic analyses can be significantly biased. To minimize the impact of this bias, two strategies can be followed: (1) the cosmic shear analysis is restricted to $0 < z < 2.5$, where catastrophic redshift errors are expected to be insignificant; (2) a spectroscopic survey is conducted for galaxies with $3 < z_{\text{phot}} < 4$. We find that the number of spectroscopic redshifts needed scales as $N_{\text{spec}} \propto f_{\text{cata}} \times A$, where $f_{\text{cata}} = 1\%$ is the fraction of catastrophic redshift errors (assuming a nine-filter photometric survey) and $A$ is the survey area. For $A = 1000 \text{ deg}^2$, we find that $N_{\text{spec}} \geq 320$ and 860, respectively, in order to reduce the joint bias in $(w_0, w_a)$ to be smaller than $2\sigma$ and $1\sigma$. This spectroscopic survey (option 2) will improve the figure of merit of option 1 by a factor $\times 1.5$ thus making such a survey strongly desirable.

Key words: gravitational lensing – large-scale structure of universe

1. INTRODUCTION

The weak gravitational lensing effect by large-scale structures, namely, the cosmic shear, is believed to be a powerful tool to measure the cosmological parameters including the equation of state of dark energy. Over the last 10 years, weak lensing observations have already provided increasingly better constraints on the matter density $\Omega_M$ and the amplitude of mass fluctuations on the scale of 8 Mpc $h^{-1}$ (Hoekstra et al. 2002; Jarvis et al. 2003; Rhodes et al. 2004; Heymans et al. 2004; Hoekstra et al. 2006; Semboloni et al. 2006; Massey et al. 2007a; Benjamin et al. 2007; Fu et al. 2008). Their contributions to the constraints on the total neutrino mass have also been investigated (Tereno et al. 2008; Gong et al. 2008; Ichiki et al. 2009; Li et al. 2009). Several more ambitious weak lensing surveys targeting at hundred to thousand times larger areas, such as DES\textsuperscript{4}, LSST\textsuperscript{5}, JDEM\textsuperscript{6}, and EUCLID\textsuperscript{7}, are being planned aiming at exploring the properties of dark energy in great detail.

Future large surveys are expected to be able to control statistical errors to insignificant levels. Therefore, different systemsatics can become dominant sources of contaminations in future weak lensing studies, and should be investigated carefully. These include the effect of the point-spread function (PSF; e.g., Amara & Refregier 2008), the methodology in the shear measurement (Bacon et al. 2001; Erben et al. 2001; Hirata & Seljak 2003; Heymans et al. 2006; Massey et al. 2007b; Miller et al. 2007; Kitching et al. 2008; Bridle et al. 2008; Ngan et al. 2008; Semboloni et al. 2008), redshift errors (Ma et al. 2006; Huterer et al. 2006), intrinsic alignments of source galaxies (see Heavens 2001 for a review, Bridle & King 2007; Fan 2007), uncertainties in the computation of the nonlinear power spectrum (see discussion in Huterer 2002; Van Waerbeke et al. 2002), and so on. In this paper, we focus on the impact of the photometric redshift errors on cosmological studies.

The necessity of measuring vast amount of galaxy redshifts in cosmic shear observations makes the multicolor photometric determination of redshifts the only feasible way in obtaining the redshift information of source galaxies. However, the inaccuracy of photometric redshifts can contaminate considerably the cosmological information embedded in cosmic shears (Huterer et al. 2006; Ma et al. 2006; Amara & Refregier 2007; Wittman et al. 2007; Abdalla et al. 2008; Mardonner & Wittman 2008; Stabenu et al. 2008). Taking into account the bias and the scatter of $z_{\text{phot}}$ with respect to $z_{\text{spec}}$, Ma et al. (2006) analyze the corresponding degradation of the constraints on dark energy. It is found that these errors can completely erase the additional information from tomographic lensing observations. In order to make full use of the advantage of the tomographic binning, both the bias and scatter of $z_{\text{phot}}$ have to be known to better than 0.003–0.01. Amara & Refregier (2007) and Abdalla et al. (2008) pay special attention to the so-called catastrophic errors in $z_{\text{phot}}$, the outliers with $|z_{\text{spec}} - z_{\text{phot}}| > 1$, and explore their effects on the figure of merit (F.o.M) of dark energy parameter constraints.

While most of the existing studies concentrate on the error range of parameter constraints, the systematic bias in the determination of cosmological parameters due to the photometric redshift errors should also be carefully analyzed. A large systematic bias can lead to a completely wrong conclusion on the values of the cosmological parameters although the statistical
errors may be small (Amara & Refregier 2008). Our study presented in this paper emphasizes precisely this aspect of effects on dark energy parameter estimates arising from catastrophic failures in photo-z measurements. We focus on a survey with similar design as described in Aldering et al. (2004), with its nine-filter set described also in Dahlen et al. (2008) and S. Jouvel et al. (2009, in preparation). We analyze in particular the bias on dark energy parameter estimates induced by the catastrophic failures occurred around \( (z_{\text{spec}}, z_{\text{phot}}) \sim (0.4, 3.5) \). And we investigate in detail the requirements of the number of spectroscopic redshift measurements to retrieve the correct estimates.

The outline of the paper is as follows. In Section 2, we introduce the methods to describe the lensing observables and to evaluate the biases. We present the results of bias in Section 3. In Section 4, we discuss the calibration requirements to remove the biases due to the catastrophic failures. Discussions and conclusions are given in Section 5.

2. METHODOLOGY

In this section, we describe the formalism of lensing observables. Based on simulation results, we then model the source galaxy distribution with catastrophic failures taken into account. Finally, we discuss the methods to evaluate the corresponding biases on dark energy parameter estimates, without and with spectroscopic calibration.

2.1. Lensing Observables

The convergence power spectrum \( P_{ij}(\ell) \) for the \( ith \) and \( jth \) redshift bin can be expressed as (see Kaiser 1992; Kaiser 1998; Ma et al. 2006)

\[
P_{ij}(\ell) = \int_0^\infty dz \, W_i(z) \, W_j(z) \, \frac{H(z)}{D(z)} \, P(k_\ell, z),
\]

where \( H(z) \) is the Hubble parameter and \( D(z) \) is the comoving angular diameter distance. The matter power spectrum \( P(k_\ell, z) \) with \( k_\ell = \ell / D(z) \) is computed from the linear transfer function of Bardeen et al. (1986), and then the nonlinear fitting function of Peacock & Dodds (1996). The weighting function \( W_i \) is given by

\[
W_i(z) = \frac{3}{2} \, \Omega_m \, \frac{H_0^2}{H(z)} \, D(z) \left(1 + z\right) \int_0^\infty dz' \, n_i(z') \, \frac{D_{L,i}(z, z')}{D(z')},
\]

where \( D_{L,i}(z, z') \) is the angular diameter distance between \( z \) and \( z' \), and \( n_i \) is the galaxy distribution (normalized to unity) of the \( ith \) redshift bin. We employ the overall galaxy distribution of the form (Smail et al. 1994)

\[
n(z) \propto z^\alpha \exp[-(1.41/z_{\text{med}})^\beta],
\]

where \( z_{\text{med}} \) is the median redshift of a survey and we take \( \alpha = 2, \beta = 1.5 \).

Considering only the shot noise and the Gaussian sample variance, the covariance matrix of the lensing observables can be written as (e.g., Huterer et al. 2006; Ma et al. 2006)

\[
\text{Cov} \left[ C_{ij}(\ell'), C_{kl}(\ell) \right] = \frac{\delta_{\ell'} \ell}{(2\ell + 1) f_{\text{sky}} \Delta \ell} \times \left[ C_{ij}^s(\ell) C_{kl}^s(\ell) + C_{ij}^s(\ell) C_{jl}^s(\ell) \right],
\]

where \( \Delta \ell \) is the band width in multipole \( \ell \) and \( f_{\text{sky}} \) is the fractional sky coverage of a survey. The total power spectrum \( C_{ij}^s \) is given by

\[
C_{ij}^s = P_{ij}^s + \delta_{ij} \sigma^2_{\gamma} n_g^i, \quad \text{(5)}
\]

where \( \sigma^2_{\gamma} \) is the variance of each component of the intrinsic ellipticity of source galaxies and \( n_g^i \) is the surface number density of galaxies in the \( ith \) bin.

In our study, we consider a fiducial space-based survey with \( f_{\text{sky}} = 0.025 \), corresponding to 1000 deg\(^2\), the summation of \( n_g^i \), i.e., the total surface number density of source galaxies \( n_g^\text{tot} \), is taken to be 100 arcmin\(^{-2}\), \( \sigma_{\gamma} = 0.22 \), and \( z_{\text{med}} = 1.26 \).

We are aware of the likely extension of the survey area. Thus, in analyzing the requirement of the spectroscopic redshift calibration, we derive a scaling relation that allows us to estimate the corresponding calibration requirement for any survey area and any catastrophic error fraction.

Investigations by Ma et al. (2006) show that tomographic analyses with multi-bins of source galaxies (denoted as \( N_{\text{bin}} \)) can improve the statistical errors on cosmological parameter constraints considerably in comparison with those of two-dimensional analyses. The improvement increases with the increase of \( N_{\text{bin}} \). On the other hand, we cannot gain much further improvement with \( N_{\text{bin}} > 5 \). Hence, throughout the paper we consider lensing tomography with five equally spaced redshift bins within the range of \( z = [0, 3] \). All galaxies with \( z > 3 \) are simply added to the last bin. The resulting redshift binning is illustrated in Figure 1.

2.2. Modeling Galaxy Distribution with Catastrophic Failures

Photometric estimations of galaxy redshifts through multi-wavelength observations largely depend on the characteristic features of the spectral energy distribution (SED) of galaxies, such as the Lyman break and Balmer/4000 Å break. The locations of the features give us the redshift information of the galaxies. Therefore, the accuracy of photometric redshifts is sensitive to the wavelength coverage of observations. It is also affected by the observational signal-to-noise ratio and the estimation method. Besides, the spectroscopic calibration plays an important role in improving the accuracy of photo-z, which shall be discussed specifically in Section 4.

To realistically assess the accuracy of photometric redshifts for our survey, we first consider a nine-filter set which covers a wavelength range from visible blue to infrared with \( \lambda \sim 390 \) nm to \( \sim 1700 \) nm (for specific definition, see Dahlen et al. 2008).

Then, a series of simulations are carried out. The details of the simulations will be presented in S. Jouvel et al. (2009, in preparation). Here, we give a brief description. The simulations are performed with the “Le Phare” software. First, we generate galaxies according to the GOODS luminosity function by Dahlen et al. (2005), each with an assigned spectrum based on the Coleman Extended templates (Coleman et al. 1980). This set of templates contains 66 spectra obtained by interpolating between five standard spectra: elliptical; Sbc; Scd; irregular; and star-forming galaxies. We then apply the filtering process to the simulated spectra with the nine-filter set. We thus obtain the SED for each simulated galaxy. The magnitudes of galaxies are randomly assigned in accordance with the properties of our fiducial survey. The photometric redshift for each galaxy is then computed using the LePhare photometric redshift code. The results are shown in Figure 2.
are simply added into the last bin. Five tomographic bins are equally spaced
between 0 < z < 3, denoted by dashed vertical lines. All galaxies with z > 3
are simply added into the last bin.

It is seen that scatters and biases around $z_{\text{spec}}$ exist for $z_{\text{phot}}$.
Besides, there are noticeable islands in the $z_{\text{spec}}$-$z_{\text{phot}}$ plot. The
dominant one is located at $z_{\text{spec}} \sim 0.4$ and $z_{\text{phot}} \sim 3.5$, which
accounts for more than 50% of the galaxies with $|z_{\text{phot}} - z_{\text{spec}}| \geq
1$. This island obviously comes from the confusion of the 4000 Å
break with the Lyman break in galaxy SEDs due to the limited
wavelength coverage. It is found that the fraction of galaxies
with such large catastrophic errors is about 1% of the total
number of galaxies. In our study, we particularly analyze the
effect of this island on the constraints of dark energy parameters.

To characterize the true distribution of galaxies whose photo-
metric redshifts suffer from catastrophic failures, we use the
following Gaussian distribution with the form

$$A_{\text{cata}} \exp \left[ -\frac{(z - zm)^2}{2\sigma^2} \right], \quad (6)$$

where the “true” mean redshift of the island $zm = 0.398$ and
the standard deviation $\sigma = 0.108$ are obtained from fitting to
the spectral(true)-$z$ distribution histogram of the island. The
parameter $A_{\text{cata}}$ is determined by normalizing the Gaussian
distribution to the overall fraction of catastrophic failures
$f_{\text{cata}} = 0.01$. Then, in a tomographic division according to
galaxies’ photometric redshifts, this part will be mistakenly
shifted from low-$z$ bin to another high-$z$ bin. Specifically, for
the $N_{\text{bin}} = 5$ tomography we consider, the 1% galaxies with
their true-$z$ falling in the first bin are assigned to the last (5th)
bin whose redshift range is $z_{\text{phot}} > 2.4$.

2.3. Evaluation of Biases on Dark Energy Parameter Estimates

Once the distribution of catastrophic failure fraction is known,
we can evaluate the biases it induces on dark energy parameter
estimates.

As mentioned in the previous subsection, we need to be aware of
the fact that observationally source galaxies are assigned to
different redshift bins in terms of their “observed” photometric
redshifts rather than their true redshifts. Hence, a fraction of
low-redshift galaxies can be mistakenly distributed to high-
redshift bins due to the catastrophic errors. As a consequence,
weak lensing signals from those relevant tomographic bins are
perturbed and biases on cosmological parameter estimates
may arise, if models used to fit the data were not modified
properly.

Concerning the fiducial survey in this study, we know from
simulations that there will be $f_{\text{cata}} = 0.01$ galaxies being
misplaced into the 5th bin (denoted as “Bin-5”), whilst their
true redshifts are within the range of the first bin (denoted
as “Bin-1”). The true distribution of this catastrophic failure
part can be described with Equation (6) and will be labeled as
Bin-cata for convenience in the following discussion. Note that
the true-$z$ range of Bin-cata is within the $z$-range of Bin-1.

For Bin-1, when Bin-cata galaxies within it are misplaced in
Bin-5 due to catastrophic redshift errors, the number of galaxies
contained in Bin-1 is reduced. The redshift distribution of the
galaxies left in it also changes somewhat in comparison with the
true distribution without the catastrophic errors. The number
change in Bin-1 does not affect the lensing signal because the
galaxy distribution within the bin is always normalized to unity.
The shape change of the redshift distribution in the bin may
have some effects. On the other hand, however, this change may
already be included in the modeled overall redshift distribution
of galaxies derived from photometric redshifts. Thus, in our
consideration we assume that the misplacement of Bin-cata does
not affect the lensing signal from Bin-1.

For bins in between Bin-1 and Bin-5, obviously the lensing
signals should be kept unperturbed. Combining the previous
analysis for Bin-1, we thus get $C_{ij}(\ell) = C_{ij}^{\ell}$, where
$i, j = 1, \ldots, 4$.

To the contrary, the lensing signals from Bin-5 get perturbed.
As shown in the bottom-left panel of Figure 3, it in fact consists of
two parts of galaxies, one with their true redshifts falling in
Bin-5 and the other being a contamination from Bin-cata.
Correspondingly, besides the auto spectra of Bin-5, those cross-
spectra with one of the bins being Bin-5 will also be affected. In
this case, those convergence power spectra can be expressed as

$$P_{55}(\ell) = \int_0^\infty dz \ W_i(z) W_j(z) \frac{H(z)}{D(z)^2} \ P(k, z) \quad (i = 1, \ldots, 5) \quad (7)$$

with

$$W_5(z) = \cdots \int_z^\infty dz' n_5(z', f_{\text{cata}}, zm, \sigma) \frac{D_5(z', z')}{D(z')} \quad (8)$$

Figure 1. Source galaxy distribution $n(z)$. The solid curve is the overall galaxy
distribution defined in Equation (3). Figure 2. Result of LePhare simulation for our
fiducial survey with a nine-filter set. A magnitude cut of 26 AB is set to fit the median
redshift of the survey. The two gray lines show the borders inside which the relation
$|z_{\text{spec}} - z_{\text{phot}}| < 1$ holds.
where we omit the antecedent terms identical to those in Equation (2). The true (spectral) galaxy distribution of Bin-5 can be written as

\[
n_5(z, f_{\text{cata}}, z_m, \sigma) = \frac{1}{f_5 + f_{\text{cata}}} \left[ n_5(z) + n_{\text{cata}}(z, f_{\text{cata}}, z_m, \sigma) \right] = \frac{1}{f_5 + f_{\text{cata}}} \left[ n_5(z) + A_{\text{cata}} \exp \left[ -\frac{(z - z_m)^2}{2\sigma^2} \right] \right],
\]

where \(1/(f_5 + f_{\text{cata}})\) is to normalize \(n_5(z, f_{\text{cata}}, z_m, \sigma)\) to be unity, with \(f_5\) being the overall fraction of galaxies with their true redshifts falling in Bin-5. Here, the right-side terms \(n_5(z)\) and \(n_{\text{cata}}\) are normalized to their overall fraction \(f_5\) and \(f_{\text{cata}}\), respectively. \(A_{\text{cata}}\) is thus determined by \(f_{\text{cata}}\), as mentioned in the paragraph following Equation (6). With the substitution \(n_5 \rightarrow n_5 + n_{\text{cata}}\) into Equation (5), finally we get the “observed” total spectra \(C_5^\ell(\ell, f_{\text{cata}}, z_m, \sigma)\), where \(f_{\text{cata}} = 0.01, z_m = 0.398, \sigma = 0.108\), and \(\ell = 1, \ldots, 5\).

In practical analysis, when calculating the “theoretically expected” lensing signals for Bin-5, we have no knowledge, in the case without spectroscopic calibration, of the fraction of low-redshift galaxies and regard all the galaxies being at high redshifts; or in the case with inadequate spectroscopic calibration, we have inaccurate knowledge, therefore can make unprecise estimation of the fraction and distribution of low-z galaxies (see Figure 3). The former can be denoted as \(C_5^\ell(\ell, 0, c, c)\), with \(c\) being an arbitrary constant. The latter case gives \(\tilde{C}_5^\ell(\ell, \tilde{f}_{\text{cata}}, \tilde{z}_m, \tilde{\sigma})\), where the estimations \(\tilde{f}_{\text{cata}}, \tilde{z}_m\), and \(\tilde{\sigma}\) can deviate from their true values depending on the calibration size.

Consequently, systematic biases may arise when the theoretically expected signals \(C_5^\ell\) are compared with the observed ones \(\tilde{C}_5^\ell\) to constrain cosmological parameters. As an approximation, a simple extension of the Fisher matrix formalism can be used to compute the bias in cosmological parameters with (see Huterer et al. 2006; Amara & Refregier 2008)

\[
\delta p_i = F_{ij}^{-1} \sum_{\ell, a, \beta} \left[ C_a^\ell(\ell) - \tilde{C}_a^\ell(\ell) \right] \text{Cov}^{-1}_{ij} \left[ C_a^\ell(\ell), C_\beta^\ell(\ell) \right] \frac{\partial \tilde{C}_\beta^\ell(\ell)}{\partial p_j},
\]

where \(a\) and \(\beta\) each denote a pair of redshift bins. The term \(\left[ C_a^\ell(\ell) - \tilde{C}_a^\ell(\ell) \right] \) reflects the bias on the shear signals. The covariance of the cross power spectra is given in Equation (4). \(F_{ij}^{-1}\) is the inverse of the Fisher matrix

\[
F_{ij} = \sum_\ell \left( \frac{\partial C}{\partial p_i} \right)^T \text{Cov}^{-1} \frac{\partial C}{\partial p_j},
\]

where \(C\) is the column matrix of the convergence power spectra and \(\text{Cov}^{-1}\) is the inverse of the covariance matrix.

Equation (10) is a linear approximation around the fiducial model. On the other hand, the true extent of bias can be quantified. On the other hand, the true extent of bias can be quantified by a direct chi-square fitting method, by minimizing

\[
\chi^2 = \sum_{\ell, a, \beta} \left[ C_a^\ell(\ell) - \tilde{C}_a^\ell(\ell) \right] \text{Cov}^{-1} \left[ C_a^\ell(\ell), C_\beta^\ell(\ell) \right] \left[ C_a^\ell(\ell) - \tilde{C}_a^\ell(\ell) \right].
\]

In our analyses, we will utilize both bias evaluation methods. The comparison of their results permits us to quantify the extent to which the linear approximation is valid. We shall come back to this in the following section, associated with the specific study for our fiducial survey.

Throughout our study, a seven-parameter fiducial flat cosmological model is assumed. For equation of state of dark energy, we consider the form \(w(z) = w_0 + w_a z/(1 + z)\) (Chevallier & Polarski 2001). The fiducial values of the seven parameters are...
f

1% \( f_{\text{cata}} \): no calibration

Figure 4. Biases on the dark energy parameters \( w_0 \) and \( w_a \) in five \( z \)-bins tomography, from \( \chi^2 \) fitting (left panel) and from Fisher matrix calculations (right panel), respectively. The fiducial value is denoted by the oblique cross, and the plus represents the “best-fit” value. The error contours (1, 2, and 3\( \sigma \), inside-out) are given at the best-fit value for the fit result and at the fiducial value for Fisher matrix calculations, respectively. Gaussian priors \( \sigma(\Omega_b) = 0.01 \) for \( \Omega_b \) and \( \sigma(p_i) = 0.05 \) are applied upon all other hidden parameters.

\( \Omega_m = 0.27, \ w_0 = -1, \ w_a = 0, \ h = 0.72, \ \sigma_b = 0.78, \ \Omega_b = 0.0446, \) and \( n = 0.96, \) respectively.

All the weak lensing related calculations are done with the ICOSMO package (A. Refregier, ver. 2005), with modifications from two-dimensional weak lensing to tomographic form for our specific study. The \( \chi^2 \) fitting tool is MPFIT\(^9\) from C. Markwardt.

3. THE BIAS ON DARK ENERGY PARAMETER ESTIMATES

To demonstrate clearly their significance, in this section, we present the results of bias on dark energy parameters due to catastrophic failures if there is no spectroscopic calibration for the problematic Bin-5. We apply both methods by \( \chi^2 \) fitting and by computation with the extension of Fisher matrix formalism. In the next section, we will focus on the requirements for the spectroscopic calibrations.

First, we employ the \( \chi^2 \) fitting method. As described in the previous section, for Bin-5, the observed signal is a combination of the two parts illustrated in the bottom-left panel of Figure 3.

Our calculations are done for a survey area \( A = 1000 \text{deg}^2 \), corresponding to a sky coverage \( f_{\text{sky}} = 0.025 \). Increasing the survey area reduces the statistical errors with \( \sigma \propto 1/\sqrt{f_{\text{sky}}} \).

According to Hu & White (2001), the bin width should be at least twice as big as \( \Delta \ell_{\text{field}} \) in order to reduce the covariance between the bins arising from the survey geometry. We then take 10 bins, uniform in logarithmic scale over the range \( \ell = [50, 3000] \).

In the right panel of Figure 4, we show the resulting bias on the equation of state of dark energy obtained from \( \chi^2 \) fitting. Gaussian priors \( \sigma(\Omega_b) = 0.01 \) for \( \Omega_b \) and \( \sigma(p_i) = 0.05 \) for others are applied upon the hidden parameters in our seven-parameter fiducial model. These priors are consistent with the current constraints. It is found that the 1% catastrophic failure fraction can bring large biases on \( w_0 \) and \( w_a \) in the tomographic five \( z \)-bins case, highly significant compared to their statistical errors. The “best-fit” values are \( w_0 = -3.5 \) and \( w_a = 3.8 \) in comparison with the fiducial ones \( w_0 = -1 \) and \( w_a = 0 \).

The results from the extended Fisher matrix formalism are shown in the right panel of Figure 4. The same \( \ell \)-binning process and Gaussian priors are applied. The “best-fit” values are found to be \( w_0 = -4.2 \) and \( w_a = 6 \). Thus, the linear approximation results in a noticeably larger bias than those of the \( \chi^2 \) fitting. This significant discrepancy reveals clearly the necessity to go beyond the linear approximation from the Fisher matrix formalism.

In summary, our analyses show that for our fiducial survey with a nine-filter set, the catastrophic photo-\( z \) errors with \( f_{\text{cata}} \sim 1% \) can bias the dark energy parameters at a level far exceeding the statistical errors if no further spectroscopic calibration is conducted.

Our calculations are done for a survey area \( A = 1000 \text{deg}^2 \), corresponding to a sky coverage \( f_{\text{sky}} = 0.025 \). Increasing the survey area reduces the statistical errors with \( \sigma \propto 1/\sqrt{f_{\text{sky}}} \). On the other hand, the absolute biases on cosmological parameters should not change much. This can be seen clearly from Equation (10) in which \( f_{\text{sky}} \) enters through \( F_i^{-1} \) and \( \text{Cov}^{-1} \).

Since both are proportional to \( f_{\text{sky}}^{-1} \), their dependences on \( f_{\text{sky}} \) cancel out. Thus, with a larger sky coverage the effect of the bias relative to the statistical errors would be even larger if the catastrophic fraction \( f_{\text{cata}} \) remains \( \sim 1% \).

4. REQUIREMENTS OF SPECTROSCOPIC CALIBRATIONS

The large bias seen in Figure 4 clearly reveals the necessity to take extra measures to reduce the effects of the catastrophic redshift errors. Among others, the spectroscopic calibration can play a crucial role in this regard. It is known that the spectroscopic calibration can be done either by taking spectra of a subsample of source galaxies after photo-\( z \) measurements or by employing a sophisticated technique where the spectral information is included in the determination of photometric redshifts. We consider the former case in our analyses, and explore

\(^8\) For a latest version, see http://www.icosmo.org/ and Refregier et al. (2008).

\(^9\) http://cow.physics.wisc.edu/~craigm/idl/fitting.html
the requirements of the spectroscopic calibration. Specifically, we study the needed number of galaxies with measured spectroscopic redshifts in order to retrieve the unbiased parameter estimates.

Since we particularly analyze the “island” catastrophic failures around \((z_{\text{spec}}, z_{\text{phot}}) \sim (0.4, 3.5)\) seen in Figure 1, we consider the calibration for galaxies with their photo-z in the range \(3 < z_{\text{phot}} < 4\). We assume that our simulation results shown in Figure 1 reflect the real relation between photo-z and the true redshifts for observed galaxies. In other words, we take the simulation results as the fiducial distribution with an overall \(f_{\text{cata}} = 1\%\). We then investigate how many spectra are needed for the subsample of galaxies with \(3 < z_{\text{phot}} < 4\) to obtain some information on the catastrophic contamination, mainly the value of \(f_{\text{cata}}\).

Our analyzing procedures are as follows. We use the simulated galaxies with \(3 \leq z_{\text{phot}} \leq 4\) as our parent fiducial sample. Then, for a given \(N_{\text{spec}}\), where \(N_{\text{spec}}\) is the number of galaxies chosen for spectral measurements, we perform a large number of Monte Carlo random samplings of \(N_{\text{spec}}\) galaxies from the parent catalog. For each realization, we first compute the estimated fraction of catastrophic failures \(\bar{f}_{\text{cata}}\) by counting the number of galaxies among \(N_{\text{spec}}\) that are identified in the realization to be in Bin-cata, taking into account the overall fraction of galaxies that have \(3 \leq z_{\text{phot}} \leq 4\). With this subsample of Bin-cata galaxies, we then fit their distribution to the form given by Equation (6) to obtain the estimated \(\bar{z}_{\text{m}}\) and \(\bar{\sigma}\), denoted as \(\bar{z}_{\text{m}}\) and \(\bar{\sigma}\). With these calibrated values of \(\bar{f}_{\text{cata}}, \bar{z}_{\text{m}}, \) and \(\bar{\sigma}\), we are able to calculate the model-predicted signals. By comparing them with those from the fiducial distribution, we therefore can evaluate the biases on cosmological parameters for each individual realization. The biases arise from the differences between the calibrated \((\bar{f}_{\text{cata}}, \bar{z}_{\text{m}}, \bar{\sigma})\) and the fiducial \((f_{\text{cata}}, z_{\text{m}}, \sigma)\). With the results from all the realizations, we can further evaluate statistically the level of bias from \(N_{\text{spec}}\) calibrations.

Figure 5 shows the results on \(w_0\) and \(w_a\) of 100 realizations for \(N_{\text{spec}} = 100\) (upper panels) and \(N_{\text{spec}} = 500\) (lower panels), respectively. The symbols are the “best-fit” values from individual realizations, and the statistical errors are calculated at the fiducial values for the purpose of clarity. The left panels are for the results of full calculations including both the auto- and cross-correlations between different redshift bins. The right panels are the results considering only the autocorrelations within individual redshift bins. In each panel, the pluses and triangles correspond, respectively, to the results from \(\chi^2\) fitting and from the Fisher matrix calculations of Equation (10). First, when the bias is large and the fitted values are outside the \(3\sigma\) contour, the results from the linear approximation of Equation (10) start to deviate from the \(\chi^2\) fitting results, as seen most evidently in the upper right panel. This shows the limitation of the linear approximation. If in some cases very large biases are expected, it is necessary to perform a full \(\chi^2\) fitting to evaluate quantitatively the biases. Otherwise, an overestimate of the bias on \(w_0\) and \(w_a\) would be obtained from the linear analysis.
is in good accordance with the results shown in Figure 4. On the other hand, when the biases are within the $3\sigma$ statistical contour, the two sets of results are in excellent agreement. Thus, in this regime the linear approximation of Equation (10) provides us an accurate and convenient way to estimate the bias effects.

We further find that the triangles from the linear calculations are almost perfectly on a straight line with little scatters. The results from the full $\chi^2$ calculations follow the same line when the biases are relatively small, and are bent downward when the biases are large. Noting that the biases depend on three quantities ($f_{\text{cata}}$, $z_m$, $\bar{\sigma}$), the lining up of the results indicate that there is a dominant factor among the three. Our detailed analyses, presented later in Figure 6, show that the biases are mainly determined by $(f_{\text{cata}} - f_{\text{cata}})$, and the shape of the redshift distribution in Bin-cata (described by the parameters $z_m$, $\bar{\sigma}$) has insignificant effects.

The straight line behavior presents us a quantitative way to describe the joint bias of $w_0$ and $w_a$. We define the joint bias of $w_0$ and $w_a$ for a realization, denoted by $\delta(w_0, w_a)$, as the distance in the $(w_0, w_a)$ plane, between the fiducial point and the best-fit point of this realization, with a sign equal to sign[$w_a$(fit) $- w_a$(fiducial)]. Furthermore, to quantify the relative bias with respect to the statistical errors, we define the corresponding joint statistical error $\sigma(w_0, w_a)$ as the distance in the $(w_0, w_a)$ plane between the fiducial point and the intersection point of the straight line with a given statistical error contour. This definition accounts correctly for the correlation between $w_0$ and $w_a$.

To see the dependence of the bias on calibrations, in Figure 6, we show the correlation of the joint bias $\delta(w_0, w_a)$ defined above and the residual $f_{\text{cata}}$ (corresponding to $f_{\text{cata}} - f_{\text{cata}}$) from each individual realization for $N_{\text{spec}} = 500$. It clearly demonstrates that the bias scales approximately linearly with the fraction of catastrophic failures. The small dispersions of $\delta(w_0, w_a)$ at a fixed residual $f_{\text{cata}}$ correspond to results with different calibrated $z_m$ and $\bar{\sigma}$ for Bin-cata. This reveals quantitatively that the effect of the specific distribution within Bin-cata is subdominant, compared to the impact of the overall fraction of Bin-cata, namely, $f_{\text{cata}}$. This fact actually allows us to obtain an approximate scaling relation for the dependence of the bias on the survey area $f_{\text{sky}}$ and on the catastrophic fraction $f_{\text{cata}}$, which will be discussed shortly.

With the biases obtained from individual realizations, we can statistically quantify the typical level of bias for a given $N_{\text{spec}}$. From Figures 5 and 6, we see that the average bias from all the realizations of a given $N_{\text{spec}}$ is expected to be close to zero. Hence, we use the rms bias calculated from the bias distribution to characterize the bias level for that particular $N_{\text{spec}}$. To illustrate it more clearly, in Figure 7, we show the distribution of $\delta(w_0, w_a)$ from 500 realizations for $N_{\text{spec}} = 500$. The bias for each individual realization is calculated by employing the Fisher matrix formalism Equation (10) since it can give good approximation when the residual $f_{\text{cata}}$ is reasonably small. It can be seen that the distribution is very close to a zero-mean Gaussian distribution. The rms bias is calculated to be $\sigma[\delta(w_0, w_a)] \approx 0.61$ for $N_{\text{spec}} = 500$. We perform such a calculation on $\sigma[\delta(w_0, w_a)]$ for each considered $N_{\text{spec}}$, and the results are shown in Figure 8. We can see that $\sigma[\delta(w_0, w_a)] \approx 1/\sqrt{N_{\text{spec}}}$ to a very good approximation, and can be written as $\sigma[\delta(w_0, w_a)] = 0.61/\sqrt{N_{\text{spec}}}/500$. For $f_{\text{sky}} = 0.025$ and $f_{\text{cata}} = 0.01$, the $1\sigma$ and $2\sigma w_0$ and $w_a$ joint statistical errors defined above are $\sim 0.465$ and 0.762, respectively. Then, the needed $N_{\text{spec}} > 860$ and 320 so that the rms bias $\sigma[\delta(w_0, w_a)]$ can be smaller than the $1\sigma$ and $2\sigma$ statistical errors, respectively.

The specific results on the required $N_{\text{spec}}$ shown above are calculated with $f_{\text{sky}} = 0.025$ and $f_{\text{cata}} = 1\%$. However, based
on our analyses, we can derive a scaling relation which permits us to estimate the needed $N_{\text{spec}}$ for different $f_{\text{sky}}$ and $f_{\text{cata}}$. As illustrated in Figure 6, the bias is linearly proportional to $f_{\text{cata}} - f_{\text{cata}}$. Thus, the rms bias $\sigma(f_{\text{cata}} - f_{\text{cata}})$ should be approximately proportional to the r.m.s. of $f_{\text{cata}}$, namely, $\sigma(f_{\text{cata}} - f_{\text{cata}})$. Furthermore, for a set of realizations, the number of labeling galaxies that are identified to be in Bin-cata follows the binomial distribution with a mean value of $N_{\text{spec}} p$ and the variance of $N_{\text{spec}} p(1 - p)$, where $p = f_{\text{cata}} \times N_{\text{tot}} / \Delta N_5$, with $N_{\text{tot}}$ and $\Delta N_5$ representing the total number of galaxies and the number of galaxies with $3 \leq z_{\text{phot}} \leq 4$, respectively. In our consideration, we have $p \sim 0.15 \ll 1$, and the distribution is close to the Poisson distribution. Then, $\sigma(f_{\text{cata}} - f_{\text{cata}}) \propto \sqrt{f_{\text{cata}} / N_{\text{spec}}}$. On the other hand, as we discussed in Section 3, the absolute bias should not depend on the survey area. The statistical error $\sigma(w_0, w_a)$ scales as $1/\sqrt{f_{\text{sky}}}$, therefore, the relative bias with respect to the statistical error scales as $\sqrt{f_{\text{cata}} \times f_{\text{sky}} / N_{\text{spec}}}$. This in turn gives rise to the scaling relation for the needed $N_{\text{spec}} \propto f_{\text{cata}} / f_{\text{sky}}$. We then have

$$N_{\text{spec}} > 860 \left( \frac{f_{\text{cata}}}{0.01} \right) \left( \frac{f_{\text{sky}}}{0.025} \right)$$

(13)

for keeping the bias smaller than the 1σ statistical error. Thus, for $f_{\text{cata}} = 1\%$ and $f_{\text{sky}} = 0.25$ ($A = 10,000 \text{deg}^2$), we would need $N_{\text{spec}} > 8600$. We emphasize that the number 860 in the scaling relation above depends on the specific detail of the considered catastrophic errors and the overall galaxy redshift distribution $n(z)$ (see Equation (10)). Our analyses are particularly for the island located at $(z_{\text{spec}}, z_{\text{phot}}) \sim (0.4, 3.5)$, which is the dominant catastrophic error occurrence seen from our photo-$z$ simulations. It is known that the pattern of catastrophic errors depends on specific photo-$z$ measurements, noticeably on the wavelength coverage and the filter set used. Thus, different sets of filters would result in different catastrophic errors, and consequently different calibration requirements. It is noted that the above linear scaling relation for $f_{\text{cata}}$ is an approximate one as the normalization (860 here) has a mild dependence on the fiducial value of $f_{\text{cata}}$. It should also be pointed out that the calibration requirement we present does not contain the part needed internally for the photo-$z$ measurements, which we assume have been done with considerable accuracy, prior to yielding the $f_{\text{cata}} = 1\%$ catastrophic outliers.

5. SUMMARY AND DISCUSSION

In this paper, we concentrate on the island part of the galaxies in the $(z_{\text{spec}}, z_{\text{phot}})$ plot where $z_{\text{phot}}$ is completely misestimated. This type of large catastrophic failures in the photometric redshift measurements arise from the confusion of different characteristics of galaxies’ SED due to the limited wavelength coverage. The most noticeable features used in the photo-$z$ measurement of galaxies are the Lyman and 4000 Å breaks. The misidentification of the two results in the islands at $(z_{\text{spec}}, z_{\text{phot}}) \sim (0.4, 3.5)$ and $(z_{\text{spec}}, z_{\text{phot}}) \sim (3, 0.2)$ seen in Figure 1. We particularly analyze the systematic bias induced by such islands on cosmological parameter estimations. We find that a fraction of $f_{\text{cata}} = 1\%$ catastrophic failures due to galaxies at $z_{\text{spec}} \sim 0.4$ being misidentified to be at $z_{\text{phot}} \sim 3.5$, which is the dominant case of catastrophic failure occurrence in our photo-$z$ simulations, can bring systematic biases on $w_0$ and $w_a$ at a level far exceeding the statistical errors in five-tomographic-bin analysis.

It has been realized that spectroscopic redshift calibration may be necessary for future weak lensing surveys in order to reduce the uncertainty of photometric redshift errors. The large bias shown in our analyses reinforces such an argument. We investigate the requirements of the spectroscopic calibration to retrieve the unbiased parameter estimates. We find that for $f_{\text{cata}} = 1\%$, and $f_{\text{sky}} = 0.025$, about 400–1000 spectral redshift measurements for galaxies with $z_{\text{phot}} = 3, 4$ are needed. We further present a scaling relation of $N_{\text{spec}} \propto f_{\text{cata}} \times f_{\text{sky}}$. Thus, for a future survey with a survey area of 10,000 deg$^2$, the size of the calibration sample in the redshift range $z_{\text{phot}} = 3, 4$ should be on the order of $10^4$ if $f_{\text{cata}}$ remains to be $\sim 1\%$. In analyzing the error degradation from general photometric errors, people have argued that the inaccuracy of the high-redshift photometric measurements can be compensated by increasing the measurement accuracy for galaxies at relatively low redshifts. Therefore, much less spectroscopic calibration samples at high redshifts are needed (e.g., Ma et al. 2006). Concerning the catastrophic-error-induced bias discussed in this paper, however, there is no such a compensation. Thus, the required $\sim 10^4$ spectral calibrations in $z_{\text{phot}} = 3, 4$ may be regarded as challenging tasks.

It is also argued that the inclusion of both $U$ and near-IR filters is essential for the accuracy of photo-$z$ and the minimization of the catastrophic outliers (e.g., Dahlen et al. 2008; Abdalla et al. 2008), as the Lyman and 4000 Å breaks
can be effectively distinguished. Given that our simulations with nine-filter set already include the near-IR bands, we consider the improvements from \( U \)-band coverage. Extending the wavelength coverage to 3200 Å, our simulations show that the catastrophic failure fraction decreases dramatically to the level around 0.1\% (S. Jouvel et al. 2009, in preparation). In this case, the bias should no longer be significant compared to the statistical errors. It is thus important to investigate the possibility to include < 3500 Å sensitivity for the next generation space dark energy mission.

It has also been proposed to filter out those problematic galaxies, such as late-type galaxies, low signal-to-noise galaxies, or very low/high redshift (e.g., \( z < 0.5/z > 2.5 \)) galaxies whose photo-\( z \)'s are most suspicious, in the shear analysis to reduce the impact of catastrophic failures (Abdalla et al. 2008; Margoniner & Wittman 2008; Heavens et al. 2006). In such cases, the statistical power of a survey is inevitably lowered to certain extents depending on the fraction of galaxies discarded. In Figure 9, we show the degradation of the constraints if those galaxies at \( z < 0.5 \) and \( z > 2.5 \) (thus three bins left) are excluded in the analyses, compared to the case with calibrated (known) \( f_{\text{data}} \). We see that the F.o.M\(^{10} \) decreases from 41 to 24, i.e., almost a factor of 2 degradation. If we redefine the galaxies with 0.5 \( \leq z \leq 2.5 \) into five finer bins (the right panel of Figure 9), the F.o.M = 29, or a degradation factor \( \sim 1.5 \) (see also Table 1 for a summary). All these calculations are done with our fiducial survey conditions and Gaussian priors \( \sigma (\Omega_b) = 0.01 \) for \( \Omega_b \) and \( \sigma (p_j) = 0.05 \) are applied upon all other parameters.

In our current analyses, we isolate the island galaxies appearing around \( (z_{\text{spec}}, z_{\text{photo}}) = (0.4, 3.5) \) to see clearly their effects on cosmic shear analysis. On the other hand, the scatter and small bias of \( z_{\text{photo}} \) around \( z_{\text{spec}} \) have been modeled and studied in, for example, Ma et al. (2006). With \( \sigma_z \) and \( \Delta z_{\text{bias}} \) as well as their uncertainties \( \Delta \sigma_z \) and \( \Delta \Delta z_{\text{bias}} \) taken into account, the statistical errors of the constrained cosmological parameters will be degraded depending on the values of \( (\sigma_z, \Delta z_{\text{bias}}, \Delta \sigma_z, \Delta \Delta z_{\text{bias}}) \). For instance, considering \( (\sigma_z, \Delta z_{\text{bias}}, \Delta \sigma_z, \Delta \Delta z_{\text{bias}}) = (0.05(1+z), 0.003, 0.003) \), the degradation factor for \( w_0 \) and \( w_a \) is about 1.5 (Ma et al. 2006). Then, a crude estimate based on our analyses gives the needed \( N_{\text{spec}} \) at \( z \sim [3, 4] \) to be about 8600/(1.5)^2 \( \sim 4000 \) to control the bias down to 1σ level for \( f_{\text{data}} = 1\% \) and \( f_{\text{sky}} = 0.25 \). More complete analyses should be done for both the degradation and the bias simultaneously taking into account both the small and the large photo-\( z \) errors.

We thank the referee for the very constructive comments and suggestions, which help to improve the paper greatly. We are grateful to Dragan Huterer, Gary Bernstein, Anne Ealet, Alexie Leauthaud, Adam Amara, Hu Zhan, Liping Fu, Vera Margoniner, Sarah Bridle, and Ludovic Van Waerbeke for their helpful discussions. We thank Craig Markwardt for his released chi-square fitting software “MPFIT” and his very useful suggestions on some technical aspects. This research is supported in part by the NSFC of China under grants 10373001, 10533010, and 10773001, and the 973 program No. 2007CB815401. L.S. also benefits from a scholarship from the “Ministère des Affaires Étrangères” of the French government. S.J. and J.P.K. acknowledge support from CNRS and CNES. C.T., L.S., and A.T. acknowledge support from CNRS.

### Table 1

| Fiducial Five-Bins | With Cut and 0.5 < \( z < 2.5 \) | Three-Bins Left | Five-Bins Redefined |
|-------------------|---------------------------------|-----------------|---------------------|
| F.o.M             | 41                              | 24              | 29                  |

Notes. The fiducial F.o.M is given under a calibrated (known) \( f_{\text{data}} \). Gaussian priors \( \sigma (\Omega_b) = 0.01 \) for \( \Omega_b \) and \( \sigma (p_j) = 0.05 \) are applied upon all other parameters.

---

\(^{10} \text{F.o.M} = \sqrt{\text{det} (\mathbf{F})} \), where \( \mathbf{F} \) is the reduced Fisher matrix for \( w_0 \) and \( w_a \) after marginalization over other parameters.
Stabenau, H. F., Connolly, A., & Jain, B. 2008, MNRAS, 387, 1215
Tereno, I., Schimd, C., Uzan, U. Z., Kilbinger, M., Vincent, F. H., & Fu, L.  
2008, arXiv:0810.0555
van Waerbeke, L., Mellier, Y., Pello, R., Pen, U.-L., McCracken, H. J., & Jain,  
B. 2002, A&A, 393, 369

Wittman, D. M., Riechers, P., & Margoniner, V. E. 2007, ApJ, 671, 109
White, M. 2004, Astropart. Phys., 22, 211
White, M., & Hu, W. 2000, ApJ, 537, 1
Zhan, H., & Knox, L. 2004, ApJ, 616, L75