The Astrophysical Journal, 720:1351–1369, 2010 September 10 doi:10.1088/0004-637X/720/2/1351

A GENERAL STUDY OF THE INFLUENCE OF CATASTROPHIC PHOTOMETRIC REDSHIFT ERRORS ON COSMOLOGY WITH COSMIC SHEAR TOMOGRAPHY

Andrew P. Hearin[1], Andrew R. Zentner[1], Zhaoming Ma[2,3], and Dragan Huterer[4]
[1] Department of Physics & Astronomy, The University of Pittsburgh, Pittsburgh, PA 15260, USA
[2] Brookhaven National Laboratory, Upton, NY 11973, USA
[3] Department of Physics & Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA
[4] Physics Department, University of Michigan, Ann Arbor, MI 48109, USA
Received 2010 February 18; accepted 2010 July 2; published 2010 August 20

ABSTRACT

A goal of forthcoming imaging surveys is to use weak gravitational lensing shear measurements to constrain dark energy. A challenge to this program is that redshifts to the lensed, source galaxies must be determined using photometric, rather than spectroscopic, information. We quantify the importance of uncalibrated photometric redshift outliers to the dark energy goals of forthcoming imaging surveys in a manner that does not assume any particular photometric redshift technique or template. In so doing, we provide an approximate blueprint for computing the influence of specific outlier populations on dark energy constraints. We find that outlier populations whose photo-z distributions are tightly localized about a significantly biased redshift must be controlled to a per-galaxy rate of $(1\text{–}3) \times 10^{-3}$ to insure that systematic errors on dark energy parameters are rendered negligible. In the complementary limit, a subset of imaged galaxies with uncalibrated photometric redshifts distributed over a broad range must be limited to fewer than a per-galaxy error rate of $F_{\rm cat} \lesssim (2\text{–}4) \times 10^{-4}$. Additionally, we explore the relative importance of calibrating the photo-z's of a *core* set of relatively well-understood galaxies as compared to the need to identify potential catastrophic photo-z outliers. We discuss the degradation of the statistical constraints on dark energy parameters induced by excising source galaxies at high- and low-photometric redshifts, concluding that removing galaxies with photometric redshifts $z^{\rm ph} \gtrsim 2.4$ and $z^{\rm ph} \lesssim 0.3$ may mitigate damaging catastrophic redshift outliers at a relatively small ($\lesssim$20%) cost in statistical error. In an Appendix, we show that forecasts for the degradation in dark energy parameter constraints due to uncertain photometric redshifts depend sensitively on the treatment of the nonlinear matter power spectrum. In particular, previous work using Peacock & Dodds may have overestimated the photo-z calibration requirements of future surveys.

Key words: cosmology: theory – dark energy – galaxies: distances and redshifts – galaxies: photometry – gravitational lensing: weak

Online-only material: color figures

1. INTRODUCTION

Weak gravitational lensing of galaxies by large-scale structure is developing into a powerful cosmological probe (e.g., Hoekstra et al. 2002; Pen et al. 2003; Jarvis et al. 2003, 2006; Van Waerbeke et al. 2005; Semboloni et al. 2006; Kitching et al. 2007; Benjamin et al. 2007; Doré et al. 2007; Fu et al. 2008). Forthcoming imaging surveys such as the Dark Energy Survey (DES), the Large Synoptic Survey Telescope (LSST), the European Space Agency's Euclid, and the Joint Dark Energy Mission (JDEM) expect to exploit measurements of weak gravitational lensing of distant source galaxies as one of the most effective means to constrain the properties of the dark energy (e.g., Hu & Tegmark 1999; Hu 1999; Huterer 2002; Heavens 2003; Refregier 2003; Refregier et al. 2004; Song & Knox 2004; Takada & Jain 2004; Takada & White 2004; Dodelson & Zhang 2005; Ishak 2005; Albrecht et al. 2006; Zhan 2006; Munshi et al. 2008; Hoekstra & Jain 2008; Zentner et al. 2008; Zhao et al. 2009). The most stringent dark energy constraints can be achieved when source galaxies can be binned according to their redshifts, yielding a tomographic view of the lensing signal. Among the contributions to the dark energy error budget will be the error induced by the need to use approximate redshifts determined from photometric data (Bolzonella et al. 2000; Collister & Lahav 2004; Feldmann et al. 2006; Banerji et al. 2008; Brammer et al. 2008; Oyaizu et al. 2008; Lima et al. 2008; Dahlen et al. 2008; Abdalla et al. 2008; Newman 2008; Ilbert et al. 2009; Coupon et al. 2009; Cunha et al. 2009; Schulz 2009) because it is not possible to obtain spectroscopic redshifts for the large numbers of source galaxies needed to trace cosmic shear. Photometric redshifts are and will be calibrated by smaller samples of galaxies with spectroscopic redshifts. In this paper, we study the influence of poorly calibrated photometric redshifts for small subsets of the galaxies within imaging samples on dark energy constraints.

The influence of uncertain photometric redshifts (photo-z hereafter) on the dark energy program has been studied by a number of authors (Ma et al. 2006; Huterer et al. 2006; Lima & Hu 2007; Kitching et al. 2008; Ma & Bernstein 2008; Sun et al. 2009; Zentner & Bhattacharya 2009; Bernstein & Huterer 2010; Zhang et al. 2009). Studies of the requirements for photo-z accuracy have assumed relatively simple forms for the relationship between the inferred photo-z of a galaxy and its spectroscopic redshift, in particular that this is a Gaussian distribution with a redshift-dependent bias and scatter. The underlying assumption is that this distribution can be calibrated with an appropriate spectroscopic sample over the range of redshifts of interest. These studies indicate that roughly $N_{\rm spec} \sim 10^5$ spectroscopic redshifts are needed to render photo-z uncertainty a small contributor to the dark energy error budget, but any particular number depends upon the many details of each study. Broadening the description of the photo-z distribution to a multi-component

Gaussian leads to slightly more demanding requirements on the spectroscopic calibration sample (Ma & Bernstein 2008). However, complexity or multi-modality of the photo-z distribution will not induce large systematic errors on dark energy parameters, provided that this complexity is known and that we have some ability to calibrate complex photo-z features using spectroscopic galaxy samples (a non-trivial assumption). That is not to say that dark energy constraints are insensitive to such complexity. Broad or multi-modal photo-z distributions will provide an effective limit to the redshift resolution of tomographic weak lensing and will degrade dark energy constraints. If this complexity can be diagnosed in spectroscopic samples, it may be treated by generalizing the modeling in Ma et al. (2006) and Ma & Bernstein (2008). In approximate accordance with the prevailing nomenclature, we refer to the galaxies for which spectroscopic calibration of the photo-z distribution will be possible as the *core* photo-z distribution. Ma & Bernstein (2008) studied multi-modal core photo-z distributions in some detail.

These studies assume that the spectroscopic samples that will be obtained will suffice to calibrate the photo-z's of all galaxies utilized in the weak-lensing analysis. However, spectroscopic calibration samples may well be deficient in spectra of some subset of galaxies that otherwise may not be easily identified and removed from the imaging sample (for example, see Newman 2008, for a discussion). Consequently, some fraction of galaxies in forthcoming imaging samples may not have photo-z's that are well calibrated spectroscopically and may have photo-z's that differ markedly from their true redshifts. Including such galaxies in weak-lensing analyses would lead one to infer biased estimators of dark energy parameters. These systematic offsets in dark energy parameters may be considerable compared to statistical errors.

We refer to such subsets of galaxies that are not well calibrated by spectroscopic samples and which have photo-z distributions that differ markedly from the photo-z distributions of the core galaxy samples as *catastrophic* photo-z outliers. Our chief aim in this study is to estimate the biases induced on dark energy estimators by catastrophic photo-z outliers for a variety of possible manifestations of catastrophic outliers, and to estimate the level at which such outliers must be controlled in order to mitigate dark energy biases.

We consider two broad classes of catastrophic outliers, differentiated by the breadth of their photo-z distributions. We emphasize that our definition of a catastrophic outlier is more inclusive than previous usage (compare to Bernstein & Huterer 2010). Outlier populations with photo-z's that are confined to a small range of highly biased redshifts make up the class we refer to as *localized catastrophes*. As an example, such outliers may correspond to galaxy populations in which spectral features have been misidentified in broadband photometric observations; the prevailing usage of the term *catastrophic error* closely resembles our usage of the term *localized catastrophe*. The second class of outliers we consider, which we refer to as *uniform catastrophes*, has photometric redshifts that are relatively unconstrained. This class may more naturally be associated with a level of spectroscopic incompleteness yielding a population of imaged galaxies with little information on the reliability of their photometric redshifts.

We describe our modeling techniques in Section 2. We detail our results on the potential importance of catastrophic photo-z outliers in Section 3. This section includes a brief discussion of mitigation strategies in which we explore the possibility of eliminating subsets of galaxies in order to reduce biases at the cost of increased statistical errors. We discuss the implications of our results in Section 4 and summarize our work in Section 5.

We include in this study an Appendix that may help in comparing published results on photometric redshift calibration requirements. All treatments of dark energy constraints from weak lensing rely on some approximate treatment of the growth of structure in the nonlinear regime. Several approaches are in common use (Scherrer & Bertschinger 1991; Peacock & Dodds 1996; Seljak 2000; Ma & Fry 2000; Scoccimarro et al. 2001; Cooray & Sheth 2002; Smith et al. 2003) and additional parameters have been introduced to model baryonic processes (Rudd et al. 2007; Zentner et al. 2008; Guillet et al. 2009). In the main body of our paper, we use the fitting form provided by Smith et al. (2003). We demonstrate in the Appendix that estimates of photo-z calibration requirements depends upon the modeling of nonlinear power. Implementing the Smith et al. (2003) relation for nonlinear power results in significantly reduced photo-z calibration requirements as compared to previous results (e.g., Ma et al. 2006) that employed the Peacock & Dodds (1996) approximation.

2. METHODS

In this section, we describe the methods used in our analysis. We begin in Section 2.1 with a discussion of our treatment of photometric redshifts, including both the *core* photometric redshift distributions as well as *catastrophic* outliers. In Section 2.2, we describe our weak-lensing power spectrum observables. We describe cosmological parameter forecasting in Section 2.3 and conclude with a description of our fiducial cosmology and representative surveys in Section 2.4.

2.1. Photometric Redshift Distributions of Source Galaxies

We characterize the distribution of photometric redshifts through the probability of obtaining a photometric redshift $z^{\rm ph}$, given a galaxy with spectroscopic (or "true") redshift z, $P(z^{\rm ph}|z)$. The distribution of true redshifts of galaxies in a photometric bin labeled with index i is

$$n_i(z) = n(z) \int_{z_i^{\rm low}}^{z_i^{\rm high}} dz^{ph}\ P(z^{\rm ph}|z), \quad (1)$$

where $n(z)$ is the number density of source galaxies per unit redshift z, $n_i(z)$ is the number density of sources per unit redshift that are assigned to the ith photo-z bin, and $z_i^{\rm low}$ and $z_i^{\rm high}$ delineate the boundaries of the ith tomographic bin.

We model the overall galaxy distribution via

$$n(z) \propto z^2\ \exp[-(z/z_0)^{1.2}], \quad (2)$$

where z_0 is determined by specifying the median redshift of the survey and the powers of redshift are representative of the distributions of observed high-redshift galaxies (J. A. Newman et al. 2010, in preparation). The normalization of the overall galaxy distribution is determined by the total number of galaxies per unit solid angle,

$$N^A = \int_0^\infty d\ z\ n(z),$$

and we designate the number of galaxies per solid angle in any photo-z bin as

$$N_i^A = \int_0^\infty d\ z\ n_i(z).$$

2.1.1. The Core Photometric Redshift Distribution

For the purposes of our study, we consider the *core* galaxy distribution to be comprised of galaxies with a *photometric* redshift distribution that will be well characterized through calibration with spectroscopically observed galaxy samples. Studies using existing spectroscopic galaxy samples to predict the photo-z distributions of galaxies in future large-scale image surveys indicate that the core distributions may be complicated (e.g., Jouvel et al. 2009; Ilbert et al. 2009; Coupon et al. 2009). A common simplifying assumption in the literature is that the photometric redshifts of galaxies in the core are distributed according to a Gaussian distribution with a redshift-dependent mean and variance (e.g., Ma et al. 2006; Ma & Bernstein 2008),

$$P_{\rm core}(z^{\rm ph}|z) = \frac{1}{\sqrt{2\pi}\sigma_z}\exp\left[-\frac{(z-z^{\rm ph}-z^{\rm bias})^2}{2\sigma_z^2}\right], \quad (3)$$

where both $\sigma_z(z)$ and $z^{\rm bias}(z)$ are functions of true redshift, z. The redshift-dependent mean and variance endow this form with sufficient flexibility to treat a wide variety of redshift distributions; however, this simple model does neglect complex features that may be present in the realized photometric redshift distributions of future surveys. We adopt this model because it is a published standard against which our results can be compared, and because the complexity of calibrating the core sample of photometric redshifts is not the primary aim of our work.

We compute the functions $\sigma_z(z)$ and $z^{\rm bias}(z)$ by linear interpolation between values tabulated at 31 redshift points spaced evenly between $z = 0$ and $z = 3$. This choice of binning allows for maximal degradation in dark energy constraints absent prior information about the photometric redshift distribution of source galaxies. We treat the bias and dispersion at each of these redshifts as free parameters in our forecasts, so that there are $2 \times 31 = 62$ free parameters describing the core photometric redshift distribution. For our fiducial model, we take $\sigma_z(z) = 0.05(1+z)$ and $z^{\rm bias}(z) = 0$.

2.1.2. Catastrophic Photometric Redshift Outliers

Forthcoming large imaging surveys will observe a tremendous number of galaxies. It is unlikely that accurate calibration of every class of photometric redshift distribution will be made, at least in part due to the limitations of obtaining reliable spectroscopic redshifts (e.g., Newman 2008) and observations of relatively rare objects. If either the uncalibrated objects follow the redshift distributions of the sample of calibrated photometric redshifts, or the uncalibrated objects can be identified from imaging data and removed from the sample, they will have a relatively benign impact on the dark energy aims of these surveys. In the former case, they present no systematic error because they follow the redshift distribution of the majority of galaxies, and in the latter case they can be removed from the imaging sample at a small cost in statistical uncertainty. Conversely, if a sample of uncalibrated source galaxies that does not follow the redshift distribution of the calibrated sources remains in the imaging data used for dark energy constraints, this could represent a significant additional systematic error. In approximate accordance with established nomenclature, we refer to subsets of galaxies that do not follow calibrated photometric redshift distributions *and* cannot be removed from imaging data as *catastrophic* photometric redshift outliers.

In practice, it is expected that catastrophic photometric redshift outliers will be present at some level in forthcoming imaging surveys. The prevalence of multi-modal features in the photo-z distributions of existing calibration samples is a clear illustration of the difficulty of determining galaxy redshifts from photometric colors (Oyaizu et al. 2008; Cunha et al. 2009; Ilbert et al. 2009; Coupon et al. 2009). When a population of galaxies responsible for a non-trivial photometric redshift determination appears sufficiently often in spectroscopic samples, its associated photo-z error can be calibrated, perhaps leading to multi-modal features in the core distribution. However, there will inevitably be populations of galaxies with photo-z degeneracies that are sufficiently rare so as to evade spectroscopic sampling, the spectroscopic calibration of a truly representative sample will not be complete, and the removal of galaxies with troublesome redshifts from the imaging data will be imperfect. Each of these difficulties leads to a population of outlier galaxies, with distributions not described by the core photometric redshift model, that contributes a systematic error to dark energy parameter estimators.

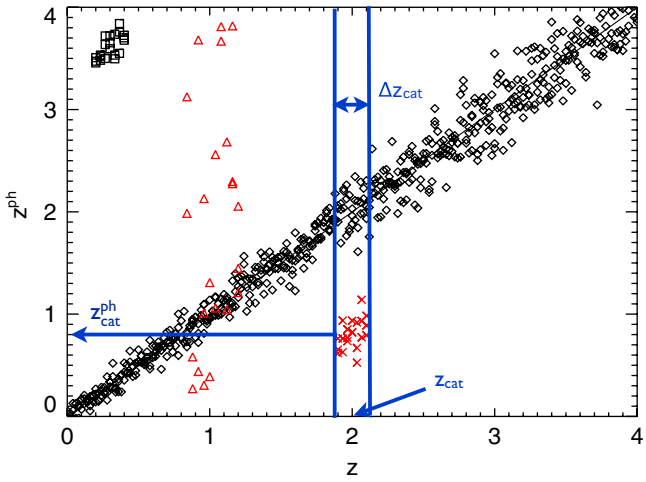

Figure 1. Toy illustration of a multi-component photometric redshift distribution. The aim of this figure is to provide a convenient, schematic representation of the photometric redshift distributions we explore. Black diamonds are galaxies in the primary peak of a Gaussian core population of photometric redshifts specified by Equation (3). Black squares are galaxies in a secondary peak in a multi-modal core distribution. These photometric redshifts are offset from the line $z = z^{\rm ph}$, but they are a known component of the photometric redshift distribution, and if they are represented adequately in spectroscopic data they can be calibrated out. Red crosses are galaxies that reside in a catastrophic outlier population with significantly biased, but relatively localized, photometric redshifts. In our nomenclature, this population is not represented in spectroscopic calibration samples and contributes a systematic error to dark energy parameters. The red triangles represent galaxies that comprise a uniform catastrophic outlier population, where photometric redshifts are relatively unconstrained. The labels $z_{\rm cat}$, $\Delta z_{\rm cat}$, and $z^{\rm ph}_{\rm cat}$ designate the parameters of our catastrophic photometric redshift models.

(A color version of this figure is available in the online journal.)

To illustrate the distinction between catastrophic outliers and multi-modal features in the core, consider the photo-z distribution illustrated in Figure 1. The bulk of the galaxies in this distribution (black diamonds) are scattered about the line $z = z^{\rm ph}$. This is a population of 400 galaxies drawn from the Gaussian distribution of Equation (3). There are also two "islands" in the distribution. The appearance of these island contributions to the photo-z distribution is quite similar, but they are intended to represent photo-z errors of a qualitatively different nature, as discussed below. One island has $(z, z^{\rm ph})$ coordinates (0.3, 3.7) and the other has (2.0, 0.8). The island at $(z, z^{\rm ph}) = (0.3, 3.7)$, consisting of black squares, is a

schematic representation of some subset of galaxies that give a known, calibrated, small probability of yielding a highly biased photometric redshift. This is a component of a multi-modal core distribution and may either be calibrated with spectroscopy or removed from the sample. The island at $(z, z^{\rm ph}) = (2.0, 0.8)$, consisting of the red crosses, is a schematic representation of a catastrophic outlier population. These are a small subset of galaxies with true redshifts near $z \approx 2$ that yield strongly biased, but localized, photometric redshifts. Moreover, this is a population that is either not identified and calibrated in spectroscopic samples, or is incompletely removed from imaging data, so that this outlier contributes a systematic error to the dark energy error budget. This is the type of error that is our focus in this manuscript. Finally, there is a population of galaxies that is localized near $z \approx 1$ and spread uniformly across $z^{\rm ph}$. These galaxies represent another extreme of catastrophic photo-z errors in that the redshifts may not be strongly biased, but they are poorly constrained and will contribute systematic errors for dark energy.

We emphasize the distinction between $P(z^{\rm ph}|z^{\rm sp})$ and the posterior redshift distribution for an individual galaxy resulting from a photometric redshift estimation algorithm, often denoted as $p(z^{\rm ph})$. Often, a single redshift estimate is assigned to a galaxy. In this case, each point in Figure 1 may correspond to the true redshift and the estimated redshift of a galaxy or population of galaxies. One may utilize more of the information in the $p(z^{\rm ph})$, in which case the local density of points in Figure 1 may correspond to regions in which the posterior has non-negligible support. Our aim is to outline a set of general impacts induced by making large, uncalibrated photometric redshift errors. We use $P(z^{\rm ph}|z^{\rm sp})$ to quantify these effects because this allows for a very general characterization of the influences of photometric redshift errors (or equally, errors in compressing the information contained in the posteriors). This is sensible because $P(z^{\rm ph}|z^{\rm sp})$ can be constructed from the posteriors of a calibration set in a straightforward manner, but this relationship is not invertible so that general statements are difficult or impossible to make.

2.1.3. Localized Catastrophic Outliers

One cause for localized catastrophic redshift errors (such as the red crosses in Figure 1) is the misidentification of a spectral feature in broadband photometric observations of galaxies over some range of true redshift. A specific example of this occurs when the Lyman break is confused with the 4000 Å break. The effect on the photo-z distribution of a small portion of errors due to Lyman–4000 Å confusion would look something like the small island of squares at $(z, z^{\rm ph}) = (0.3, 3.7)$ in Figure 1 (Bernstein & Huterer 2010). Confusion between the Lyman and 4000 Å breaks may occur often enough in spectroscopic samples to be calibrated and thus included as a secondary peak in the core distribution, but it is possible that there will be other small redshift windows where tertiary islands remain uncalibrated.

Throughout this paper, we adopt a simple model for the photo-z distributions of *localized catastrophes* as Gaussians with spreads $\sigma_{\rm cat}$ centered away from the core at $z^{\rm ph}_{\rm cat}$,

$$P_{\rm cat}(z^{\rm ph}|z) = \frac{1}{\sqrt{2\pi}\,\sigma_{\rm cat}} \exp\left[-\frac{\left(z^{\rm ph} - z^{\rm ph}_{\rm cat}\right)^2}{2\sigma^2_{\rm cat}}\right]. \quad (4)$$

The parameter $z^{\rm ph}_{\rm cat}$ specifies the location of the island in photometric redshift, and $\sigma_{\rm cat}$ gives the spread of the catastrophe in $z^{\rm ph}$. In the presence of a localized catastrophe the total photometric redshift distribution is

$$P_{\rm tot}(z^{\rm ph}|z) = [1 - \Xi(z)F_{\rm cat}]P_{\rm core}(z^{\rm ph}|z) + \Xi(z)F_{\rm cat}P_{\rm cat}(z^{\rm ph}|z). \quad (5)$$

The catastrophic error occurs over only a specified range of true redshifts, $z_{\rm cat} - \Delta z_{\rm cat}/2 < z < z_{\rm cat} + \Delta z_{\rm cat}/2$, as enforced by the function

$$\Xi(z) \equiv \Theta\left(\frac{\Delta z_{\rm cat}}{2} - |z - z_{\rm cat}|\right), \quad (6)$$

where $\Theta(x)$ is the Heaviside step function. The quantities $z^{\rm ph}_{\rm cat}$ (location of the local catastrophe in photometric redshift), $z_{\rm cat}$ (central value of the range of true redshifts over which the catastrophe occurs), $\Delta z_{\rm cat}$ (width of the range of true redshifts over which the catastrophic error is made), and $\sigma_{\rm cat}$ (spread in $z^{\rm ph}$ of the catastrophe) are four of the five parameters that specify the local catastrophe model. The fifth parameter, $F_{\rm cat}$, is the fraction of galaxies in the true redshift window set by $z_{\rm cat}$ and $\Delta z_{\rm cat}$ for which the catastrophic error occurs.

The term $\Xi(z)F_{\rm cat}$ removes the appropriate fraction of galaxies from the core distribution and ensures that $\int_0^\infty dz^{\rm ph}\ P_{\rm tot}(z^{\rm ph}|z) = 1$. As a concrete example, the catastrophic outliers represented by the red crosses in Figure 1 are galaxies drawn from our model with $F_{\rm cat} = 0.03$, $z^{\rm ph}_{\rm cat} = 0.8$, $z_{\rm cat} = 2.0$, $\sigma_{\rm cat} = 0.1$, and $\Delta z_{\rm cat} = 0.1$. For the sake of pragmatism, we present results for localized catastrophes in interesting limits of this five-dimensional parameterization rather than an exhaustive exploration of these parameters.

2.1.4. Uniform Catastrophic Outliers

Empirically, photometric redshift determination algorithms applied to extant calibration samples yield photometric redshift estimates that are relatively unconstrained on some subsets of galaxies. For example, the photo-z distribution of galaxies in both the Canada–France–Hawaii Telescope Legacy Survey (CFHTLS; Coupon et al. 2009) and the Cosmological Evolution Survey (COSMOS; Ilbert et al. 2009) possess such a feature within the range of error rates we explore in this work. Unconstrained photometric redshifts represent a regime complementary to localized catastrophes. In this case, photometric redshifts may be obtained with nearly equal probability over a significant range of redshift. Such broad errors may occur when light from one galaxy is contaminated by light from another source nearby in angular separation but at a different redshift. It is natural to expect that such errors will occur most frequently near the peak of the observed galaxy number density $n(z)$.

Motivated by the presence of such errors, we also treat the extreme case of relatively unconstrained photometric redshifts by using a uniform distribution for $z^{\rm ph}$, over a symmetric window in true redshift centered on $z_{\rm cat}$ and spanning a width of $\Delta z_{\rm cat}$. We refer to this kind of error as a *uniform* catastrophe for simplicity. In the presence of a uniform catastrophe, the total photometric redshift distribution is

$$P_{\rm tot}(z^{\rm ph}|z) = [1 - \Xi(z)F_{\rm cat}]P_{\rm core}(z^{\rm ph}|z) + \Xi(z)F_{\rm cat}/(z_{\rm max} - z_{\rm min}), \quad (7)$$

where $z_{\rm min}$ and $z_{\rm max}$ delineate the photometric redshift range of the survey. In analogy to localized catastrophes, the function $\Xi(z)$ restricts the true redshift range over which flat catastrophes occur and $F_{\rm cat}$ specifies the fraction of galaxies in this true redshift window whose redshifts are catastrophically in error.

Therefore, three parameters specify this simple model, namely, $F_{\rm cat}$, $z_{\rm cat}$, and $\Delta z_{\rm cat}$. The uniform catastrophe represented by the red triangles in Figure 1 is drawn from a model with $F_{\rm cat} = 0.05$, $z_{\rm cat} = 1.0$, and $\Delta z_{\rm cat} = 0.2$.

2.2. Cosmic Shear Tomography

In this study, we consider constraints from weak gravitational lensing observables only. We split source galaxies into $N_{\rm TOM}$ photometric redshift bins and consider as our observables the $N_{\rm TOM}(N_{\rm TOM}+1)/2$ distinct number-weighted auto- and cross-power spectra of convergence among the source redshift bins. Unless otherwise stated, we bin source galaxies in equal intervals of redshift between $z^{\rm ph} = 0$ and $z^{\rm ph} = 3$ and take $N_{\rm TOM} = 5$, resulting in 15 distinct observables. For this redshift range, five-bin tomography is a useful standard because this binning scheme suffices to saturate dark energy constraints (Ma et al. 2006; we have verified that this remains so within the parameters of our study as well).

The galaxy number count in each tomographic bin N_i^A, the cross-spectra between bins i and j, $P_\kappa^{\rm ij}(\ell)$, and the number-weighed spectra $\mathcal{P}_\kappa^{\rm ij}(\ell)$ are related by

$$\mathcal{P}_\kappa^{\rm ij}(\ell) = N_i^A N_j^A P_\kappa^{\rm ij}(\ell) = \int_0^\infty dz \frac{W_i(z)W_j(z)}{H(z)D_A^2(z)} P_\delta(k=\ell/D_A, z). \quad (8)$$

In Equation (8), ℓ is the multipole number, $H(z)$ is the Hubble expansion parameter, $D_A(z)$ is the angular diameter distance to redshift z, and $P_\delta(k, z)$ is the three-dimensional matter power spectrum. The lensing weight functions, $W_i(z)$, weight the cosmic shear signal according to the redshift distributions of galaxies within each tomographic bin and are defined as

$$W_i(z) = \frac{3}{2}\Omega_{\rm M} H_0^2 (1+z) D_A(z) \int_z^\infty dz' \frac{D_A(z,z')}{D_A(z')} n_i(z'), \quad (9)$$

where $D_A(z, z')$ is the angular diameter distance between redshifts z and z'.

2.3. Parameter Forecasting

We use the Fisher matrix formalism to study the constraining power of our weak-lensing observables on dark energy parameters as well as to quantify the systematic errors on dark energy parameters that result from catastrophic photometric redshift errors. The Fisher matrix formalism is ubiquitous in cosmological parameter forecasting (useful references related to the present application include Jungman et al. 1996; Tegmark et al. 1997; Seljak 1997; Kosowsky et al. 2002; Huterer & Takada 2005; Albrecht et al. 2006; Bernstein & Huterer 2010), so we simply quote relevant results here. The particular implementation we use closely mirrors that in Zentner et al. (2008) and Hearin & Zentner (2009), to which we refer the reader for details.

The Fisher matrix is given by a sum over the observables. In the particular case of weak-lensing power spectra, the spectra at different multipoles can be treated as independent and this sum can be written as

$$F_{\alpha\beta} = \sum_{\ell_{\rm min}}^{\ell_{\rm max}} (2\ell+1) f_{\rm sky} \sum_{\rm A,B} \frac{\partial \mathcal{P}_{\rm A}}{\partial p_\alpha} [C^{-1}]_{\rm AB} \frac{\partial \mathcal{P}_{\rm B}}{\partial p_\beta} + F_{\alpha\beta}^{\rm P}, \quad (10)$$

where the $\mathcal{P}_{\rm A}$ are the set of observables indexed by a single label, $\mathbf{C}^{-1}$ is the inverse covariance matrix of these observables at fixed multipole, $[C^{-1}]_{\rm AB}$ are the components of the inverse of the covariance matrix (we include the brackets for clarity), and p_α are the theoretical model parameters. We choose an indexing scheme in which lowercase Greek letters designate model parameters, upper-case Latin letters designate observables, and lowercase Latin letters designate photometric redshift bins, and take the mapping between observable number and tomographic bin number to be $A = i(i-1)/2 + j$. Throughout this paper, we use $\ell_{\rm min} = 2 f_{\rm sky}^{-1/2}$, where $f_{\rm sky}$ is the fractional sky coverage of the weak-lensing survey, and $\ell_{\rm max} = 3000$ as a rough indication of the scale beyond which a number of weak-lensing approximations break down (White & Hu 2000; Cooray & Hu 2001; Vale & White 2003; Dodelson et al. 2006; Semboloni et al. 2006; Rudd et al. 2007).

The covariance matrix of observables at each multipole is

$$C_{\rm AB}(\ell) = \bar{\mathcal{P}}_\kappa^{\rm ik}(\ell)\bar{\mathcal{P}}_\kappa^{\rm jl}(\ell) + \bar{\mathcal{P}}_\kappa^{\rm il}(\ell)\bar{\mathcal{P}}_\kappa^{\rm jk}(\ell), \quad (11)$$

where the indices i and j map onto A and k and l map onto B. The observed number-weighted power spectra, $\bar{\mathcal{P}}_\kappa^{\rm ij}(\ell)$, have contributions from signal and shot noise,

$$\bar{\mathcal{P}}_\kappa^{\rm ij}(\ell) = \mathcal{P}_\kappa^{\rm ij}(\ell) + N_i^A \delta_{ij} \langle \gamma_{\rm i}^2 \rangle, \quad (12)$$

where the quantity $\langle \gamma_{\rm i}^2 \rangle$ is the intrinsic source galaxy shape noise. We conform to recent convention and fix $\sqrt{\langle \gamma_{\rm i}^2 \rangle} = 0.2$, so that all deviations from this noise level are incorporated into an effective galaxy number density.

The Fisher matrix formalism provides an estimate of the parameter covariance near a fiducial point in the parameter space. One chooses fiducial values for the model parameters and estimates the error on parameter α from the inverse of the Fisher matrix at this point, $\sigma(p_\alpha) = [F^{-1}]_{\alpha\alpha}$. Within this formalism, statistically independent prior information about the parameters is easily incorporated by simple matrix addition. The second term in Equation (10) is the prior matrix. In our analysis, we assume *independent* prior constraints on cosmological parameters, so that the prior matrix reduces to a simple diagonal matrix, $F_{\alpha\beta}^{\rm P} = \delta_{\alpha\beta}/(\sigma_\alpha^{\rm P})^2$, where $\delta_{\alpha\beta}$ is the Kronecker-δ symbol and $\sigma_\alpha^{\rm P}$ is the prior 1σ, Gaussian constraint on parameter p_α. We itemize our fiducial model and priors in the following subsection.

Given a systematic error that induces a specific shift in the observables, one can use the Fisher matrix to estimate the ensuing systematic error in model parameters. Using $\Delta\mathcal{P}_{\rm A}$ to denote the difference between the fiducial observables and the observables perturbed by the presence of the systematic error, one will infer a set of parameters that is systematically offset from the true parameters by

$$\delta p_\alpha = \sum_\beta [F^{-1}]_{\alpha\beta} \sum_\ell (2\ell+1) f_{\rm sky} \sum_{\rm A,B} \Delta\mathcal{P}_{\rm A} [C^{-1}]_{\rm AB} \frac{\partial \mathcal{P}_{\rm B}}{\partial p_\beta}. \quad (13)$$

The primary results of our work are estimates of the systematic errors in dark energy parameters induced by catastrophic photometric redshift outliers. In related literature, the δp_α are often referred to as *biases*; however, we refer to them as *systematic errors* in order to avoid potential confusion with the biases in photometric redshifts.

2.4. Cosmological Model and Survey Characteristics

We assume a cosmological model specified by seven parameters. Three of these parameters describe the dark energy. These

three parameters are the present energy density in units of the critical density, $\Omega_{\rm DE} = 0.76$, and two parameters, $w_0 = -1$ and $w_{\rm a} = 0$, that describe a linearly evolving dark energy equation of state, $w(a) = w_0 + (1-a)w_{\rm a}$ (e.g., Linder 2003; Chevallier & Polarski 2001; Huterer & Turner 2001; Albrecht et al. 2006). The values specified for these parameters are those in our fiducial cosmological model. In models with a time-varying dark energy equation of state, it is interesting to present results for the constraint on $w(a)$ at the scale factor at which it is most well constrained. The scale factor at which $w(a)$ can be best constrained is the *pivot* scale factor $a_{\rm p}$, and is related to the Fisher matrix components as

$$a_{\rm p} = 1 + \frac{[F^{-1}]_{w_0 w_{\rm a}}}{[F^{-1}]_{w_{\rm a} w_{\rm a}}}. \quad (14)$$

The pivot equation of state parameter is

$$w_{\rm p} \equiv w(a_{\rm p}) = w_0 + (1-a_{\rm p})w_{\rm a} \quad (15)$$

and the error on $w_{\rm p}$ is

$$\sigma^2(w_{\rm p}) = [F^{-1}]_{w_0 w_0} - \frac{([F^{-1}]_{w_0 w_{\rm a}})^2}{[F^{-1}]_{w_{\rm a} w_{\rm a}}}. \quad (16)$$

The dark energy task force quantifies the constraining power of forthcoming surveys according to a figure of merit that reflects the areas of the confidence ellipses in the w_0–$w_{\rm a}$ plane. In particular, the task force quotes values for the combination $\mathcal{F} \equiv [\sigma(w_{\rm a}) \times \sigma(w_{\rm p})]^{-1}$ (Albrecht et al. 2006).

The other cosmological parameters we consider and the fiducial values they assume in our modeling are the nonrelativistic matter density $\omega_{\rm M} \equiv \Omega_{\rm M} h^2 = 0.13$, the baryon density $\omega_{\rm B} = \Omega_{\rm B} h^2 = 0.0223$, the amplitude of the primordial curvature fluctuations $\Delta^2_{\mathcal{R}} = 2.1 \times 10^{-9}$ (though in practice we vary $\ln \Delta^2_{\mathcal{R}}$ when computing derivatives of this parameter) evaluated at the pivot scale $k_{\rm p} = 0.05$ Mpc^{-1}, and the power-law index of the spectrum of primordial density fluctuations $n_{\rm s} = 0.96$. We adopt relatively conservative priors of $\sigma^{\rm P}(\omega_{\rm M}) = 0.007$, $\sigma^{\rm P}(\omega_{\rm B}) = 10^{-3}$, $\sigma^{\rm P}(\ln \Delta^2_{\mathcal{R}}) = 0.1$, and $\sigma^{\rm P}(n_{\rm s}) = 0.04$, each of which is comparable to contemporary, marginalized constraints on these parameters (Komatsu et al. 2008). Using marginalized, contemporary priors allows for somewhat more parameter degeneracy than may be possible with Planck data and leads to dark energy parameter forecasts that are relatively conservative.

In principle, it is relatively straightforward to scale parameter forecasts from one experiment to another (e.g., Ma et al. 2006; Bernstein & Huterer 2010); however, in the interest of simplicity, we present explicit results for three specific experimental configurations that span the range of observations expected of forthcoming instruments.

The DES is the most near-term survey that we consider.[5] We model a DES-like survey by assuming a fractional sky coverage of $f_{\rm sky} = 0.12$ and a surface density of imaged galaxies of $N^A = 15$ arcmin^{-2}. Second, we consider a narrow, deep imaging survey similar to a Supernova Acceleration Probe-like implementation of a JDEM.[6,7] We refer to this second type of survey as DEEP and model it with $f_{\rm sky} = 0.05$ and $N^A = 100$ arcmin^{-2}. Lastly, motivated by a future ground-based imaging survey such as may be carried out by the LSST[8] (Abell et al. 2009), or a space-based mission such as the European Space Agency's Euclid[9] (Refregier et al. 2010), we consider a survey with very wide sky coverage taking $f_{\rm sky} = 0.5$ and $N^A = 30$ arcmin^{-2}. We refer to this class of survey as WIDE. We assume that the median galaxy redshift in the WIDE and DEEP surveys is $z_{\rm med} = 1.0$ and that the median galaxy redshift in the DES-like survey is $z_{\rm med} = 0.7$. In all cases, we follow recent convention by taking the shape noise to be $\sqrt{\langle \gamma^2_{\rm int} \rangle} = 0.2$, subsuming additional noise contributions into an effective galaxy number density. Table 1 summarizes our assumed survey properties.

Table 1
Representative Surveys and Baseline Constraints

Survey	$f_{\rm sky}$	N^A (arcmin^{-2})	$z_{\rm med}$	$\sigma(w_0)$	$\sigma(w_{\rm a})$	$\sigma(w_{\rm p})$	$\mathcal{F}$
DES	0.12	15	0.7	0.25	0.77	0.07	18.6
WIDE	0.50	30	1.0	0.07	0.22	0.02	227.3
DEEP	0.05	100	1.0	0.10	0.33	0.04	75.6

Notes. Column 1 gives the survey that motivates the particular choice of parameters. Column 2 is the fractional sky coverage of the survey. Column 3 gives the effective galaxy number density N^A, in arcmin^{-2}. We have followed current convention and adopted a fixed shape noise of $\sqrt{\langle \gamma^2_{\rm int} \rangle} = 0.2$, assuming deviations from this assumption to be encapsulated in the effective galaxy number density. Column 4 gives the median redshift of galaxies in the survey. Columns 5–8 give dark energy equation of state constraints in the limit of perfect knowledge of the photometric redshift distribution of sources. These include the uncertainty on the pivot equation of state $\sigma(w_{\rm p})$ and the product $\mathcal{F} = [\sigma(w_{\rm a}) \times \sigma(w_{\rm p})]^{-1}$. Note that these constraints are from the weak-lensing components of these surveys only and account for statistical errors only.

3. RESULTS: SYSTEMATIC ERRORS ON THE DARK ENERGY EQUATION OF STATE

In this section, we present the results of our study of catastrophic photometric redshift outliers. We begin with the baseline constraints on the dark energy equation of state parameters in the limit of perfect knowledge of the source galaxy photometric redshift distribution in Section 3.1. We continue in a sequence of increasing complexity. We quantify the influence of catastrophic photometric redshift errors in the limit of perfect knowledge of the core photometric redshift distribution in Section 3.2. We present results on the influence of catastrophic photometric redshift errors in the more realistic case of imperfect knowledge of the core distribution in Section 3.3. We explore the prospect of excising galaxies based on their photometric redshifts as a simple, first-line defense against systematic errors induced by catastrophic photometric redshift errors in Section 3.4.

3.1. Baseline Constraints

We begin our results section by stating our forecasts for dark energy constraints in the limit of perfect knowledge of the photometric redshift distribution. With little uncertainty in photometric redshift distributions, the statistical limits of forthcoming survey instruments would allow for constraints on the dark energy equation of state at the level of a few percent, as summarized in Table 1. We emphasize here that the limit of perfect knowledge of the photo-z distributions

[5] http://www.darkenergysurvey.org
[6] http://jdem.gsfc.nasa.gov/
[7] http://snap.lbl.gov/
[8] http://www.lsst.org
[9] http://sci.esa.int/euclid

is not the assumption that photometric redshifts are precisely equal to the true redshifts of the source galaxies. Rather, the assumption is that there are no catastrophic errors, and that the photometric redshift distribution is described by the Gaussian in Section 2.1.1 such that all 62 parameters used to specify the Gaussian distribution are known precisely.

3.2. Systematic Errors in The Limit of Perfect Core Knowledge

In this section, we present results for systematic photometric redshift errors in the limit of perfect knowledge of the *core* distribution of photometric redshifts. This amounts to the assumption of prior knowledge of the 31 dispersion [$\sigma_z(z)$] and 31 bias [$z_{\rm bias}(z)$] parameters defined in Section 2.1.1 to a level of $\lesssim 10^{-3}$, which could be achieved with a *representative* sample of $\gtrsim 4 \times 10^5$ spectroscopic redshifts distributed in redshift in a manner similar to those in the imaging survey (see Ma et al. 2006; Ma & Bernstein 2008, and the discussion in the Appendix of this manuscript). This is a simple case to begin with as it allows exploration of the influence of catastrophic redshift errors over a range of the catastrophic photo-z parameter space without the additional complications associated with redshift-dependent priors on the core photo-z distribution. This is the limit explored by Bernstein & Huterer (2010).

3.2.1. Uniform Catastrophes

First, we address systematic errors induced on dark energy parameters by a small population of *uniform* catastrophes. Uniform catastrophes are cases in which some small population of galaxies with true redshifts in the range ($z_{\rm cat} - \Delta z_{\rm cat}/2 < z < z_{\rm cat} + \Delta z_{\rm cat}/2$) yield photometric redshift estimates that are distributed broadly in $z^{\rm ph}$. This class of error differs from the conventional use of the term *catastrophic error* and may more naturally be interpreted as a tolerance on spectroscopic incompleteness.

For simplicity, we take the central redshift of the uniform catastrophe to be $z_{\rm cat} = z_{\rm med}$, and determine systematic errors as a function of $\Delta z_{\rm cat}$, the width of the range of redshifts over which such errors occur, and $F_{\rm cat}$, the fraction of galaxies in this range of true redshift that correspond to this type of catastrophic error. We refer the reader to Equations (6) and (7) for the expressions that formally define these parameters. While we vary these parameters independently, they are both related to the total number density of sources with redshifts that are catastrophically in error,

$$N^A_{\rm cat} = F_{\rm cat} \int_{z_{\rm cat} - \frac{\Delta z_{\rm cat}}{2}}^{z_{\rm cat} + \frac{\Delta z_{\rm cat}}{2}} dz'\, n(z'), \qquad (17)$$

where $n(z)$ is the overall, true redshift distribution of galaxies. We should expect systematic errors to increase with both $F_{\rm cat}$ and $\Delta z_{\rm cat}$ because higher values of either parameter result in a greater total number of catastrophic errors in the outlier population.

In Figure 2, we have quantified the systematic errors induced by uniform catastrophic errors as a function of the parameters of our simple model. The curves in Figure 2 are contours of constant systematic error on dark energy parameters (for example, $|\delta(w_0)|$ for w_0) expressed in units of the statistical error ($\sigma(w_0)$ for w_0) at points in the $\Delta z_{\rm cat}$–$F_{\rm cat}$ plane. For each of the DES, Wide, and Deep surveys, the solid curves trace systematic errors in dark energy that are three times the statistical errors, while the dashed curves trace systematic errors that are 1/3 of the statistical error. For each of the surveys depicted in

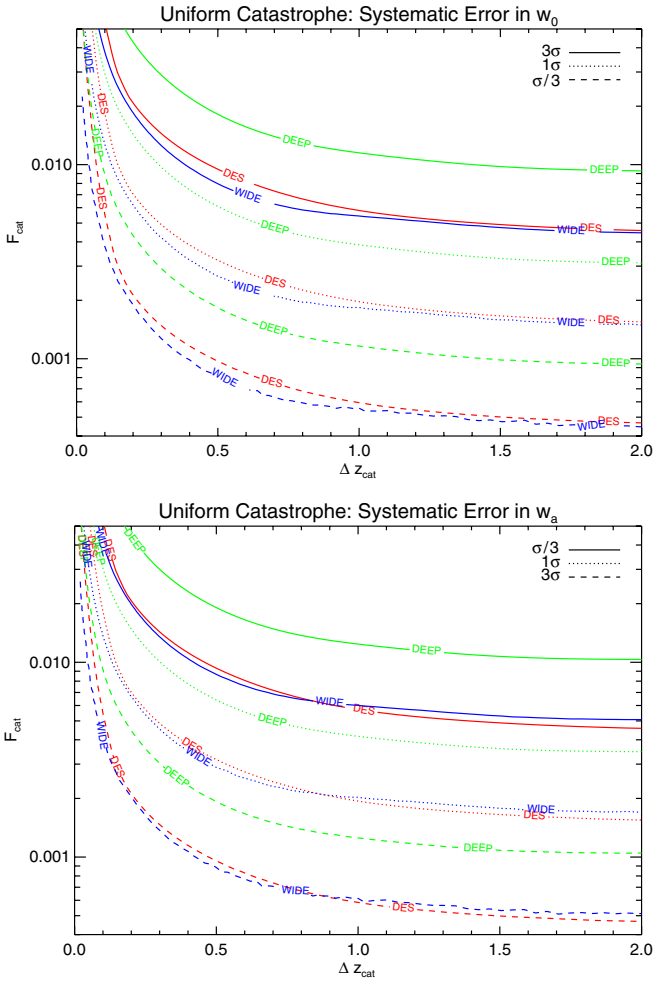

Figure 2. Systematic errors on dark energy parameters in the case of a *uniform* photometric redshift catastrophe. The horizontal axis is the width of the range in true redshift over which the uniform catastrophe is realized, $\Delta z_{\rm cat}$. This range in true redshift is centered at $z_{\rm cat} = z_{\rm med}$ for each experiment ($z_{\rm med} = 0.7$ for DES and $z_{\rm med} = 1$ for DEEP and WIDE). The vertical axis is the catastrophic error rate per galaxy within this true redshift range, $F_{\rm cat}$. The solid (dashed) lines show contours of constant systematic error equal to three times (one third) the statistical errors on each of the dark energy parameters. The top panel shows contours for w_0 and the bottom for w_a. Chance cancellations in the biases induced by high- and low-redshift galaxies cause the DES to be sensitive to catastrophic errors at similar levels to the WIDE survey and more sensitive than the DEEP survey.

(A color version of this figure is available in the online journal.)

Figure 2, the region of catastrophic parameter space that is bracketed by the solid and dashed curves labeled with the corresponding survey name corresponds to outliers that produce systematic errors which are comparable to statistical errors. Systematic errors are relatively small compared to statistical errors in the regions below the dashed curves. Each curve plotted in Figure 2 has been generated with a fixed value of $z_{\rm cat} \equiv z_{\rm med}$. For the Wide and Deep surveys $z_{\rm med} = 1$, so when $\Delta z_{\rm cat} = 2$ the uniform catastrophes are made over the true redshift range $0 < z < 2$. For DES $z_{\rm med} = 0.7$, so once $\Delta z_{\rm cat} > 1.4$ the true redshift window over which catastrophes are made only increases at the high-redshift boundary.

For each of the contours of constant systematic error in Figure 2, $F_{\rm cat}$ decreases with increasing $\Delta z_{\rm cat}$. This is simply because increasing the redshift range over which the catastrophic errors are being made ($\Delta z_{\rm cat}$) leads to an increased total number of catastrophic errors, resulting in a decreased tolerance to

the error rate, ($F_{\rm cat}$). Alternatively, the total number of catastrophic errors in an outlier population is given by the integral in Equation (17), and the contours of constant systematic error roughly trace constant values of $N^A_{\rm cat}$. The contours flatten considerably for errors that occur over a redshift range $\Delta z_{\rm cat} \gtrsim 0.4$ because there will be comparably few imaged sources with true redshifts near $z \sim 0$ or with $z \gtrsim 1.5$.

This treatment of a uniform $z^{\rm ph}$ catastrophe may appear somewhat contrived but it gives insight into a few basic results that are important to recognize. It is clear that the utility of forthcoming shear surveys to constrain dark energy is sensitive to a fractionally small population of galaxies that may yield poorly determined photo-z estimates. If the error is only relevant to galaxies that are relatively isolated in narrow regions of true redshift, for example, with $\Delta z_{\rm cat} \lesssim 0.1$, then error rates as high as $F_{\rm cat} \sim 1\%$ in this region of true redshift are tolerable. This is simply because errors that occur with a fixed rate over a small redshift range result in a small total number of catastrophic outliers to corrupt the weak-lensing tomography. In contrast, if such an error occurs for a subset of galaxies with true redshifts in an interval of width $\Delta z_{\rm cat} \gtrsim 0.1$, then the error rate per galaxy must be significantly lower than $F_{\rm cat} \lesssim 0.01$ in order to render the systematic errors on dark energy equation of state parameters small.

The limit of $\Delta z_{\rm cat} \gg 0.1$ is interesting to consider. This may correspond to the case of a small fraction of galaxies that yield very poorly constrained photometric redshifts over a broad range of true redshifts and that otherwise cannot be identified and removed from the imaging survey. In this case, the systematic error from catastrophic photometric redshifts becomes a considerable portion of the dark energy error budget at a rate of only $F_{\rm cat} \sim 10^{-3}$. Reducing the systematic error due to such an outlier population to a negligible level requires reducing the occurrence of such an outlier population to $F_{\rm cat} \lesssim 4 \times 10^{-4}$. Strictly speaking, Figure 2 corresponds to errors that occur when the true redshift band over which the uniform catastrophe occurs is centered on $z_{\rm cat} = z_{\rm med}$, but for the $\Delta z_{\rm cat} \gtrsim 1$ limit, similar results hold for a wide range of $z_{\rm cat}$ near unity, so this result is of some general relevance to photometric redshift calibration studies.

We conclude this section with a discussion of cancellations that may occur among systematic errors. It may seem somewhat surprising that DES exhibits comparable sensitivity to uniform errors as WIDE and is more sensitive than DEEP as shown in Figure 2. In Section 3.2.2, we will discuss systematic errors from local catastrophes. In particular, we will show that large biases occur for low $z_{\rm cat}$ and for higher $z_{\rm cat}$ just over the median redshifts of the surveys (see Figure 3). These biases have opposite signs and partially cancel in our forecasts for both the DEEP and WIDE surveys. DES is less sensitive to biases from galaxies in the low-redshift range $0.4 < z_{\rm cat} < 0.6$ that get misplaced to higher redshifts because these shifts must compete with the larger shot noise of DES. The degree of cancellation depends upon modeling choices, such as fiducial model and cosmological parameters, but the occurrence of this cancellation is robust.

3.2.2. Localized Catastrophes: Details

Localized catastrophes correspond to the case where a small fraction of galaxies near some true redshift $z_{\rm cat}$ yield photometric redshifts that are narrowly distributed about a biased value $z^{\rm ph}_{\rm cat}$ that is very different from the true redshift $z_{\rm cat}$. Such errors could arise due to incomplete calibration by spectroscopic surveys or from difficulty in removing troublesome galaxies from the imaged galaxy sample. A known example of such an error occurs when photo-z algorithms confuse the 4000 Å break with the Lyman break, but other isolated islands of biased $z^{\rm ph}$ persist in contemporary photo-z algorithms (see, e.g., Coupon et al. 2009, Ilbert et al. 2009) and may be relevant to forthcoming imaging surveys.

The class of localized photo-z catastrophes is more complex than the uniform case because there are more relevant parameters needed to specify the manner in which a localized outlier population is distributed in $z^{\rm ph}$. Our toy model requires five parameters (see Section 2.1.2 and Figure 1 for an illustration). Two are the central value of the true redshift over which this error is operative ($z_{\rm cat}$) and the width of the true redshift range over which this error is operative ($\Delta z_{\rm cat}$). As in the uniform case, some fraction $F_{\rm cat}$ of galaxies with true redshifts in the interval $z_{\rm cat} - \Delta z_{\rm cat}/2 < z < z_{\rm cat} + \Delta z_{\rm cat}/2$ are catastrophically in error. The final two parameters specify the biased distribution of photometric redshifts that these galaxies are assigned. These are the (systematically erroneous) value of the photometric redshift $z^{\rm ph}_{\rm cat}$, and the dispersion in the catastrophic photometric redshift distribution $\sigma_{\rm cat}$ about $z^{\rm ph}_{\rm cat}$.

We make an effort to remain agnostic about the classes of photo-z errors that may be realized in future imaging data. However, a complete mapping of even the simple parameter space we have specified for catastrophic photo-z's would require a lengthy discussion, so we explore useful limits of the model parameters in order to distill our results into a small number of points. We are particularly interested in the limit where the source galaxies are placed in a narrow range of biased photo-z ($\sigma_{\rm cat} \ll 0.3$ or so) because the limit of large dispersion in the catastrophic photometric redshift population is similar to the *uniform* catastrophe of the previous section.

We first isolate the sinister regions in the space of $z_{\rm cat}$–$z^{\rm ph}_{\rm cat}$ that lead to the most destructive systematic errors in dark energy parameters. At a set of points in the parameter space of ($z_{\rm cat}$, $z^{\rm ph}_{\rm cat}$), we have calculated the systematic error induced in w_0 and $w_{\rm a}$ by distributing some fraction $F_{\rm cat}$ of the galaxies with true redshifts near $z_{\rm cat}$ in photometric redshifts centered around some $z^{\rm ph}_{\rm cat}$ that is generally very different from $z_{\rm cat}$. We sample a range of values of true redshifts from $z_{\rm cat} = 0.05$ to $z_{\rm cat} = 2.95$, evenly spaced in redshift intervals of $\delta z = 0.1$ and likewise for the photometric redshifts, $z^{\rm ph}_{\rm cat}$. In the interest of simplicity, we fix the remaining parameters of our catastrophic photo-z model to $F_{\rm cat} = 0.05$, $\Delta z_{\rm cat} = 0.05$, and $\sigma_{\rm cat} = 0.01$ to isolate the dependence of the parameter bias upon the location of the catastrophe.

It is important to note explicitly that we present results here at a fixed error fraction $F_{\rm cat}$, and a fixed true redshift window width and $\Delta z_{\rm cat}$. However, even with these parameters fixed the absolute number of errors varies with $z_{\rm cat}$ according to Equation (17), which is roughly $N^A_{\rm cat} \sim n(z_{\rm cat})\Delta z_{\rm cat} F_{\rm cat}$ for sufficiently small $\Delta z_{\rm cat}$, along the lines of the analogous discussion for the uniform catastrophe in Section 3.2.1. The aim of this calculation is to map out the relative importance of making errors at a *fixed rate per galaxy* as a function of the true and photometric redshifts of the outliers.

The results of this exercise are depicted in Figure 3. In each column of Figure 3 there are two panels, corresponding to the systematic errors in w_0 and $w_{\rm a}$, for each representative experiment. The horizontal axes show values of $z_{\rm cat}$ and the vertical axes show values of $z^{\rm ph}_{\rm cat}$. The systematic error is

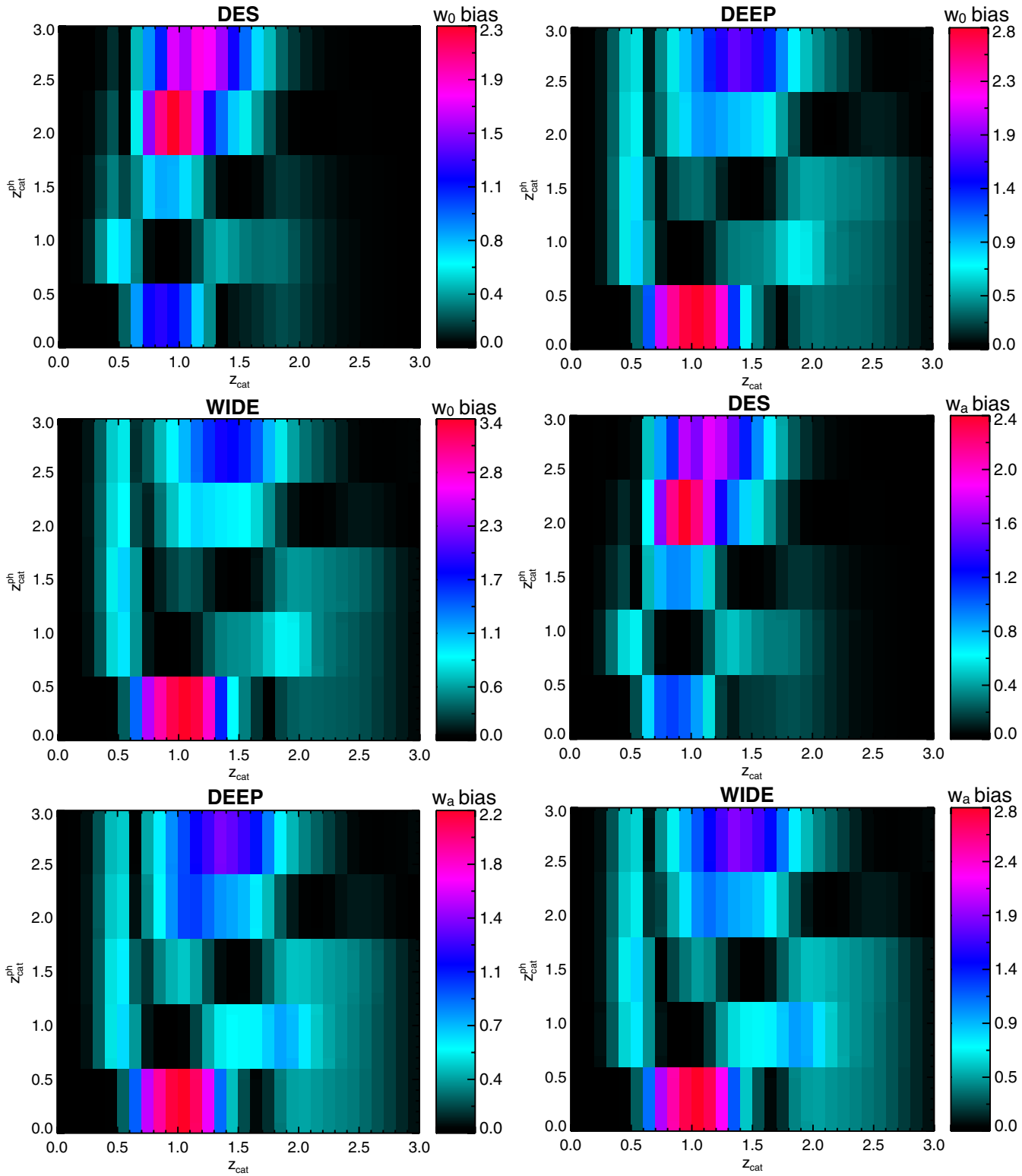

Figure 3. Severity of localized catastrophic errors as a function of the values of source $z_{\rm cat}$, and target $z_{\rm cat}^{\rm ph}$, of the catastrophic errors. Along the horizontal axes are the values of $z_{\rm cat}$ while the vertical axes show $z_{\rm cat}^{\rm ph}$, just as in Figure 1. Each point on this grid corresponds to a localized catastrophe with a fixed per-galaxy error rate of $F_{\rm cat} = 0.05$, and fixed values of both the photo-z spread $\sigma_{\rm cat} = 0.01$ as well as the width of the true redshift range over which the catastrophic error is made, $\Delta z_{\rm cat} = 0.05$. The effect of these catastrophes on w_0 is shown in the top row of panels, while the systematic error on $w_{\rm a}$ is shown in the bottom row. The absolute value of the induced systematic error is color coded; the numerical values labeling the color table to the right of each panel indicate the systematic error in units of the statistical uncertainty in the limit of perfect core calibration.

(A color version of this figure is available in the online journal.)

represented on the grid of $(z_{\rm cat}, z^{\rm ph}_{\rm cat})$ by the color in each of the cells. In discussing the results of this exercise, we find the terminology of Bernstein & Huterer (2010) to be a useful, descriptive shorthand. We will refer to the tomographic bin that contains the $z_{\rm cat}$ value of an outlier as the *Source Bin* of that catastrophic photo-z population. We call the bin containing $z^{\rm ph}_{\rm cat}$ its *Target Bin*. This is because galaxies with true redshifts near $z_{\rm cat}$ are erroneously placed in the Target Bin containing the redshift $z^{\rm ph}_{\rm cat}$. Our sampling guarantees that no localized outlier straddles a tomographic bin boundary so there are always unique Source and Target Bins. Outlier populations that straddle a boundary dividing two tomographic bins can be substantially more severe than those that do not because such an outlier simultaneously contaminates multiple Target Bins. We have chosen to ignore such outlier populations for simplicity, but such outliers can be modeled by two catastrophic outlier populations, one for each affected Target Bin. We will return to the issue of tomographic binning and straddling outlier populations below.

The prominent block-like features in Figure 3 reflect the tomographic redshift bins used in our analysis. The tomographic bins of the source and target galaxies largely determine both the magnitude and sign of the induced systematic error in dark energy parameters. This gives rise to features that reflect the structure of the photometric redshift binning in the $(z_{\rm cat},z^{\rm ph}_{\rm cat})$ plane. Indeed, for fixed Target and Source Bins, the specific value of the target redshift, $z^{\rm ph}_{\rm cat}$, within the target photometric redshift bin has little influence on the severity of the systematic error. However, small steps in $z^{\rm ph}_{\rm cat}$ can lead to large changes in systematic error when the boundary dividing two tomographic bins is crossed.

Varying the location in true redshift, $z_{\rm cat}$, leads to somewhat more significant changes in dark energy systematic error. Changing $z_{\rm cat}$ within fixed Source and Target Bins can result in up to a factor of 2 difference in systematic errors. Two factors primarily determine the severity of the systematic error as a function of the true redshift of the galaxies, $z_{\rm cat}$. The primary factor stems from the fact that a fixed fractional error rate ($F_{\rm cat}$) corresponds to a different absolute number of errors $N^A_{\rm cat}$ as a function of redshift, $z_{\rm cat}$. This is reflected in Equation (17). The number of errors $N^A_{\rm cat}$ will be relatively large in a region near the median redshift of the survey, where the number of source galaxies per unit redshift, $n(z)$, is largest. There are relatively few galaxies at low and high redshifts, so for a fixed error rate, outlier populations with low or high true redshifts contribute a relatively small absolute number of galaxies with highly biased redshifts.

Second, an outlier population naturally results in a more severe systematic error the more the photometric redshift is biased away from the true galaxy redshift. Consider the region of catastrophic parameter space near $(z_{\rm cat}, z^{\rm ph}_{\rm cat}) = (1.5, 2.7)$ in either color plot for DES. Catastrophes in this region of parameter space correspond to outlier populations whose Source Bin is the third tomographic bin and Target Bin is the fifth tomographic bin. Outliers in this region of parameter space are assigned photo-z's that are significantly too high. Decreasing the value of $z_{\rm cat}$ (the "source" of the error) increases the distance between core and outlier populations, thereby increasing the systematic error. This behavior contributes significantly to the systematic error gradient near $(z_{\rm cat}, z^{\rm ph}_{\rm cat}) = (1.5, 2.7)$ for DES in Figure 3.

Each of the three representative experiments that we consider has two distinct "hot spots" in Figure 3 that correspond to the most severe types of error given a fixed error rate per galaxy, $F_{\rm cat}$. A common feature of all these hot spots is their $z_{\rm cat}$ location. Each of the hot spots lies at a $z_{\rm cat}$ slightly beyond the median survey redshift. This is sensible because for a fixed error rate, the absolute number of catastrophic errors is greatest when they are made at the peak in the overall galaxy distribution, that is near $z_{\rm cat} = z_{\rm med}$. The most damaging systematic errors occur when the galaxies are shifted to either very low or very high photometric redshifts, when the target redshift, $z^{\rm ph}_{\rm cat}$, is very different from the source redshift, $z_{\rm cat}$, because the galaxies in error are then placed at distances significantly different from their true redshifts. For our WIDE and DEEP surveys, the largest systematic errors tend to occur for galaxies shifted from a source redshift $z_{\rm cat}$ near $z_{\rm med} = 1$ to very low photometric redshifts.

The pattern of the DES catastrophic photo-z "hot spots" differs from that of the WIDE or DEEP surveys. Outliers with large values of $z^{\rm ph}_{\rm cat}$, that is those with a target in the fourth or fifth tomographic bin, are relatively more severe for DES. This is driven by the (assumed) comparably low-redshift extent of imaged sources for a DES-like survey (with median redshift $z_{\rm med} = 0.7$) This renders a contamination that extends to high redshift more disruptive due to the small population of galaxies with truly high redshifts. Though less striking, it is also evident in Figure 3 that the DEEP survey is somewhat more sensitive than the WIDE survey to contamination of its fourth and fifth tomographic bins. The differences here are likewise driven by different survey depths and sky coverages. A deeper, but narrower survey (a JDEM perhaps) is relatively more sensitive to small-scale fluctuations induced by structure at high redshift, so disruptions to the higher tomographic bins are more statistically significant for DEEP than for WIDE.

Finally, we return to the issue of tomographic binning with respect to the systematic errors in Figure 3. We noted above that systematic errors in dark energy parameters can become markedly worse when the biased photometric redshifts ($z^{\rm ph}_{\rm cat}$) distribute galaxies across the boundary of a photometric redshift bin. The reason is because two sets of observables, namely, the auto- and cross-spectra associated with the two target photometric redshift bins, become corrupted by the catastrophic photometric redshift error. The implication is that the level of systematic error induced by a localized catastrophic error is quite sensitive to photometric redshift binning. This is contrary to the statistical errors, which are insensitive to binning more finely than $N_{\rm TOM} \approx 5$ over the range $0 < z < 3$ (Ma et al. 2006).

Indeed this is the case. The general pattern shown in Figure 3 is physically quite sensible and is robust to binning. However, in the case of localized catastrophes, binning more finely may reduce the absolute amplitude of systematic errors if the catastrophes do not occur near the edge of a photometric redshift bin. This is because smaller tomographic bins result in a smaller fraction of source galaxies that belong to a contaminated bin. This may be useful because even in the absence of significant prior indications of a localized catastrophe, re-analyzing the data with different photometric redshift binning schemes may reveal potential local catastrophes. At the least, it should be a useful strategy to choose photometric redshift bins such that suspect regions of $z^{\rm ph}$, where localized catastrophes may be anticipated, are contained in individual bins.

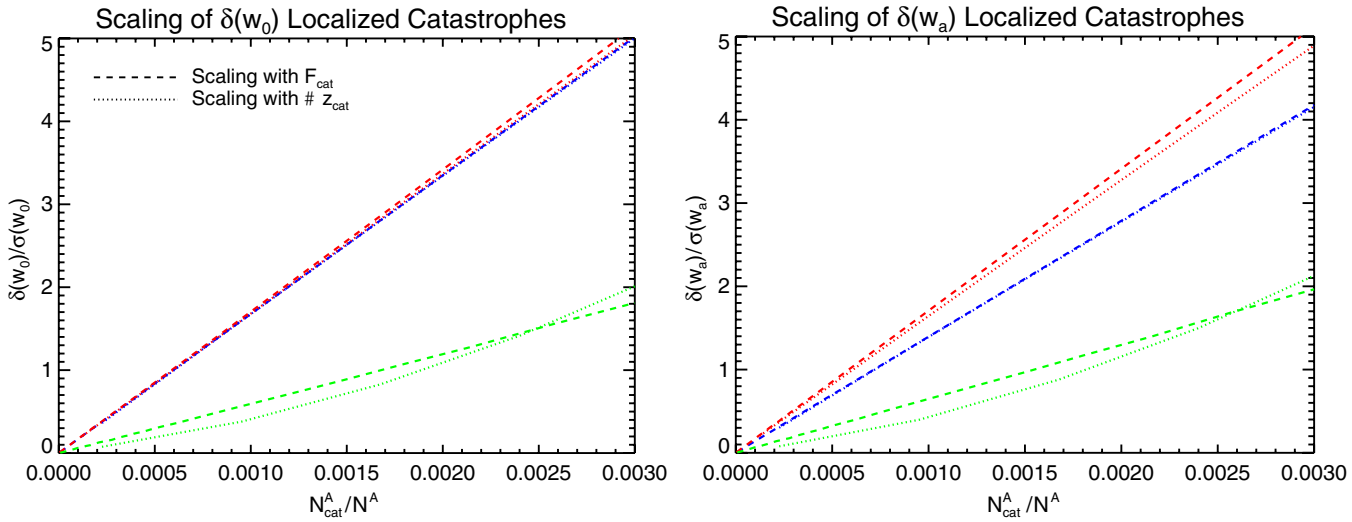

Figure 4. Scaling of systematic error in dark energy parameters with the fractional number density of sources whose photometric redshifts are catastrophically in error for our WIDE survey. On the vertical axes are the absolute value of the systematic error in w_0 (left panel) and w_a (right panel) in units of statistical uncertainty. On the horizontal axis is $N^A_{\rm cat}/N^A \approx F_{\rm cat} n(z_{\rm cat}) \Delta z_{\rm cat}$, where $n(z_{\rm cat})$ is the overall redshift distribution of sources. For the dashed curves, $\Delta z_{\rm cat} \equiv 0.1$ and we increase $N^A_{\rm cat}$ by increasing $F_{\rm cat}$. These curves are all linear, as they should be. For the dotted curves $F_{\rm cat} \equiv 0.03$ and we increase $N^A_{\rm cat}$ by increasing $\Delta z_{\rm cat}$. These curves grow approximately as the linear, dashed curves. Three different catastrophic error localizations are color coded as $(z_{\rm cat}, z^{\rm ph}_{\rm cat}) = (0.9, 0.3)$ in red, $(0.9, 2.7)$ in blue, and $(1.5, 2.7)$ in green. The difference in intrinsic severity between these outlier populations is reflected by the slope of the corresponding curves, with the steeper lines corresponding to the more severe systematic errors. The agreement between dashed and dotted lines for each outlier demonstrates that the systematic errors induced by sufficiently well-localized catastrophes ($\Delta z_{\rm cat} \lesssim 0.3$) scale approximately linearly with $\Delta z_{\rm cat}$ over an interesting range.

(A color version of this figure is available in the online journal.)

3.2.3. Localized Catastrophes: Summary

A succinct distillation of the dominant effects that determine the structure of Figure 3 is as follows. The systematic error induced by a localized catastrophe will be most severe when

1. $z_{\rm cat} \approx z_{\rm med}$, which maximizes the total number of outliers;
2. the distance between $z_{\rm cat}$ and $z^{\rm ph}_{\rm cat}$ is significant;
3. and when $z^{\rm ph}_{\rm cat}$ is such that the photometric redshift bins contain a fractionally large contaminant (in practice, high- and low-redshift extremes).

The details governing the magnitude of systematic errors generated by different regions of catastrophic error parameter space can be complicated. In general, these details depend on the relative statistical weights of the affected redshift bins, as well as the characteristics of the survey.

In isolation, Figure 3 is useful in identifying the redshift errors that most seriously compromise dark energy constraints. A shortcoming of Figure 3 is that we have assumed catastrophic errors that occur at a fixed rate of $F_{\rm cat} = 0.05$ and are active only over a range $\Delta z_{\rm cat} = 0.05$. The systematic errors induced on cosmological parameters scale approximately with the total number of catastrophes, $N^A_{\rm cat}$, in Equation (17). In practice, scaling the systematic errors to new values of $\Delta z_{\rm cat}$ can be enacted over an interesting range of the parameter space by approximating $N^A_{\rm cat} \approx F_{\rm cat}\, n(z_{\rm cat})\, \Delta z_{\rm cat}$.

Figure 4 demonstrates the validity of scaling systematic error by the total number of errors, $N^A_{\rm cat}$, for three example localized catastrophes. Together, Figures 3 and 4 provide a blueprint for estimating the systematic error induced by a wide range of localized catastrophes. One first reads off the systematic error level from Figure 3 for the grid point of interest. For definiteness, suppose this systematic error in either of w_0 or w_a is δ. Provided that $\Delta z_{\rm cat}$ is small, one can approximate the systematic error induced by a different effective value of $\Delta z_{\rm cat}$ or $F_{\rm cat}$ (call it δ') by scaling δ in proportion to $N^A_{\rm cat}$ (Equation (17)),

$$\delta' \approx \delta \times (F'_{\rm cat}/0.05) \times (\Delta z'_{\rm cat}/0.05). \quad (18)$$

In Section 3.2, we presented results on the influence of catastrophic, uncalibrated photometric redshift errors on the systematic error budget for dark energy parameters w_0 and w_a. In that section, we assumed that the bulk of photometric redshifts had been well characterized by spectroscopy. In the nomenclature of this and other papers, we assumed the limit in which the *core* of the photometric redshift distribution is calibrated so that its uncertainty does not contribute to the dark energy error budget. We developed guidance on how to optimally focus photo-z calibration efforts and identified the most severe types of catastrophes. In this section, we drop the assumption of arbitrarily precise calibration of the core populations of photometric redshifts. Our goal is to assess the relative importance of calibrating the core photometric redshift distribution compared to eliminating catastrophic errors.

We assume that the core photometric redshift distribution is specified by a Gaussian with redshift-dependent mean and dispersion. Following Ma et al. (2006), we specify the unknown mean and dispersion at 31 points spaced evenly in redshift from $z = 0$ to $z = 3$ and allow for uncertainty in these parameters. In the interest of simplicity, we consider a one-parameter family for the prior knowledge about the core photometric redshifts that may be provided by a spectroscopic calibration sample. We do this by assuming a representative population of $N_{\rm spec}$ galaxies with spectroscopic redshifts, distributed evenly in redshift from $z = 0$ to $z = 3$, which can be used to calibrate the core photometric redshift distribution.

We implement core calibration by introducing priors on the values of the dispersion and bias at the ith point in redshift.

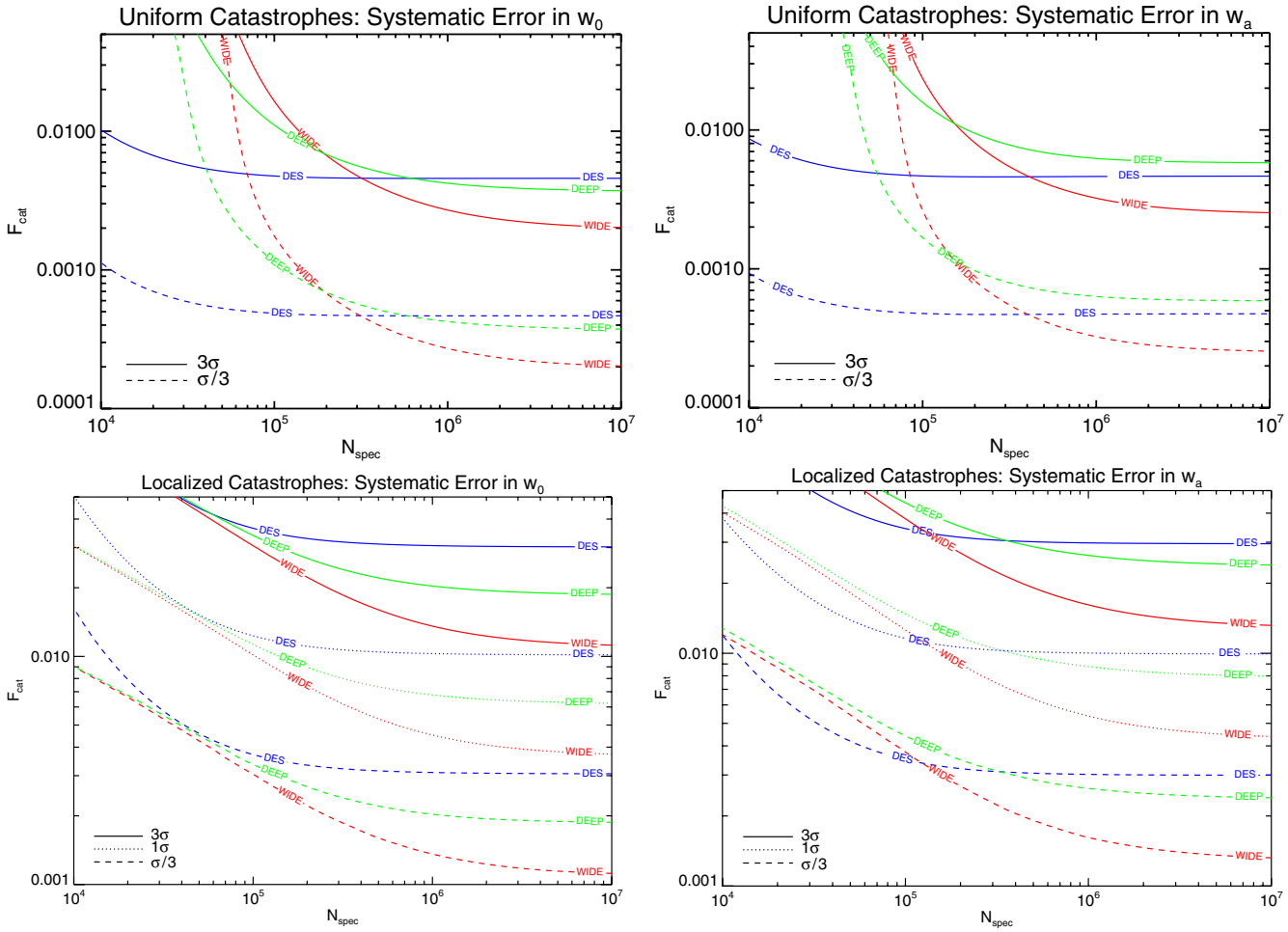

Figure 5. Contours of constant w_0 and w_a bias from the worst-case catastrophe in units of the statistical uncertainty of the survey. Systematic errors in w_0 appear in the left panels and w_a in the right panels. Results for the worst-case uniform catastrophe appear in the upper panels, and were generated with $\Delta z_{cat} = 1.5$ and $z_{cat} = 1.5$. Contours of systematic error produced by localized catastrophes appear in the bottom panels. Each of the localized contours have been calculated with $\Delta z_{cat} = 0.1$ and $\sigma_{cat} = 0.03$. For DEEP and WIDE $z_{cat} = 1.15$. and $z_{cat}^{ph} = 0.15$, and for DES $z_{cat} = 0.85$, and $z_{cat}^{ph} = 0.15$, in accordance with the results illustrated in Figure 3. (A color version of this figure is available in the online journal.)

These priors are

$$\Delta\sigma_z^i = \sigma_z^i \sqrt{\frac{1}{2N_{spec}^i}}, \tag{19}$$

$$\Delta z_{bias}^i = \frac{\sigma_z^i}{\sqrt{N_{spec}^i}}, \tag{20}$$

where z_{bias}^i is the bias at the ith point in the tabulated core distribution, σ_z^i is the dispersion at this redshift, and N_{spec}^i is the number of spectroscopic galaxies in each of the 31 bins of width $\delta z = 0.1$ used to calibrate the core photo-z redshift distribution. This prior model is certainly simplistic. For example, in our analysis we have chosen for the sake of simplicity to set all of the N_{spec}^i equal to each other, so that our implementation assumes that calibrating spectra are sampled equally in redshift, whereas in reality we will have much looser constraints on sources at high redshift than those at low redshift. Moreover, both core calibration *and* the ability to identify catastrophic outliers improve with larger spectroscopic samples. However, we consider these issues independently in the interest of completeness because the details of how a realistic calibration program may proceed remain uncertain.

3.3. Catastrophic Redshift Errors with Core Uncertainty

Figure 5 is a contour plot depicting the systematic errors in w_0 and w_a induced by the worst-case-scenario catastrophes determined in Section 3.2. The prior core knowledge is specified by N_{spec}, which runs along the horizontal axis. The error rate, F_{cat}, runs along the vertical axis. For uniform catastrophes the worst-case outliers span the true redshift range of the survey. For localized catastrophes the most sinister outliers lie at the points of maximum systematic error in Figure 3. The dashed (solid) curves are lines of constant systematic error at a level of one-third (three times) the statistical error on each parameter. Clearly, then, systematic errors are dominant above the solid curves and become unimportant well below the dashed curves. In the bottom panels, we have included dotted curves to emphasize the region of parameter space where systematic errors are equal to statistical errors. These 1σ contours are omitted in the upper panels to avoid clutter, but the linear dependence of the induced systematic error on the catastrophic error rate ensures that the 1σ can be estimated by scaling the 3σ or $\sigma/3$ contours by a factor of 3.

Several aspects of Figure 5 are worthy of note. The contours all become very flat at large N_{spec}. This is the limit in which the core photo-z distribution is calibrated sufficiently well that

it no longer contributes to the error budget of w_0 and w_a (e.g., Ma et al. 2006; Ma & Bernstein 2008). This corresponds to the limit of perfect knowledge of the core photo-z distribution, and accordingly, the systematic errors asymptote to those quoted in Section 3.2 at large $N_{\rm spec}$.

For a fixed level of systematic error, experiments generally become less tolerant of catastrophic outliers as $N_{\rm spec}$ increases. This behavior is reflected in the negative slope at the low $N_{\rm spec}$-end of the contours of constant systematic error in Figure 5. This is an explicit manifestation of the competition between calibration of the "core" population of photometric redshifts and the ability to diagnose and eliminate a sub-dominant, poorly understood "catastrophic" outlier population. The reason for this is simply that *systematic* errors must be better understood for samples with smaller *statistical* uncertainty. If the statistical errors in the measurement are intrinsically large, as they would be in the limit of poorly calibrated photo-z's for the majority of the imaging sample, then high rates of catastrophic outliers are tolerable because the systematic they contribute is not large compared to the statistical error induced by a poorly constrained core distribution.

When the core distribution is very well calibrated, most obviously at $N_{\rm spec} \gtrsim 10^5$ for uniform catastrophes in DES, the contours of constant systematic error transition to slightly positive slope. This occurs when the core distribution has been sufficiently well calibrated that degeneracies between the photometric redshift parameters of the core distribution and cosmological parameters are no longer significant. Calibrating beyond the level required to break degeneracies between cosmology and the core photo-z parameters results in a slight reduction in systematic errors on cosmological parameters. This is a specific manifestation of the general result that improving priors can only lead to a net reduction in the systematic errors of inferred parameters, a result discussed in considerable detail in Bernstein & Huterer (2010). Clearly, the reduction in systematic error at very large $N_{\rm spec}$ is not significant in the cases of interest here.

To illustrate the competition between core calibration and the removal of outliers, consider some explicit examples. In the case of the uniform catastrophe, our DEEP (WIDE) survey can tolerate catastrophic errors at a rate $F_{\rm cat} > 1\%$ if the core calibration is worse than the statistical equivalent of $N_{\rm spec} \lesssim 3 \times 10^4$ ($N_{\rm spec} \lesssim 6 \times 10^4$). For both surveys, even the worst-case, localized catastrophes can occur at a rate of $F_{\rm cat} > 1\%$ if $N_{\rm spec} \lesssim 10^4$. Of course, the worst-case, localized systematic errors are more subtle to interpret, as we have assumed they are only actively affecting galaxies over a range of true redshifts with width $\Delta z_{\rm cat} = 0.1$; however, the magnitude of the induced systematic errors produced by localized catastrophes active over different redshift ranges scales in proportion to $N^A_{\rm cat} \sim n(z_{\rm cat})\Delta z_{\rm cat} F_{\rm cat}$, as illustrated in Figure 4. Detailed results are complex, but two simple conclusions are clear.

1. Limiting uniform catastrophic error rates to $F_{\rm cat} \lesssim 4 \times 10^{-4}$ ($F_{\rm cat} \lesssim 2 \times 10^{-4}$) for DES and DEEP (WIDE) will render them unimportant.
2. Limiting individual localized catastrophic error rates to $F_{\rm cat}(\Delta z_{\rm cat}/0.1) \lesssim 10^{-3}$ will render them unimportant for each experiment.

In practice, some amount of uncertainty in the calibration of the core distribution is inevitable, so error rates higher by a factor of a few may be tolerable, but in detail this will depend upon the nature of the error and the properties of the core sample of well-calibrated photometric redshifts. Figures 3 and 5 contain the information necessary to diagnose the systematic error for a variety of idealized, but interesting cases.

3.4. Mitigating Systematic Errors by Sacrificing Statistics

In Sections 3.2 and 3.3, we estimated the systematic errors that could be induced by two broad families of catastrophic photometric redshift error, remaining relatively agnostic about the source of the error. We found generally that error rates must be kept to levels below $F_{\rm cat} \sim 10^{-3}$, or one of a thousand imaged galaxies with large, uncalibrated redshift errors in order for systematic errors not to contribute to the dark energy error budget (though specific tolerances depend upon several details). This will be a relatively challenging goal for a photometric redshift calibration program to attain. DES, JDEM, EUCLID, and LSST will all require calibration of very faint galaxies, where precise photo-z's are difficult to obtain. Moreover, the types of galaxies imaged, and for which spectra may be available, varies as a function of redshift, so some understanding of the details of galaxy evolution will be needed in order to achieve calibration goals.

It is natural to explore simple methods to sacrifice some of the statistical power of imaging surveys in order to mitigate larger systematic errors. One of the simplest techniques we can employ to limit the effect of catastrophic outliers is to place cuts on the range of photometric redshifts utilized to infer cosmological parameters (Bernstein & Huterer 2010 have explored such cuts for a particular model of photo-z outliers).[10] The most damaging catastrophic errors are those that take galaxies near the median redshift of the survey and scatter them to significantly lower or higher redshifts, so it is sensible to explore the losses in statistical power incurred by excising galaxies at the low- and high-redshift ends of surveys.

We demonstrate the utility of photometric redshift excision in this section by exploring a class of simple excision algorithms. In particular, we cut out all galaxies with *photometric* redshifts greater than some value, $z^{\rm cut}_{\rm max}$, and smaller than some value $z^{\rm cut}_{\rm min}$. Figure 6 shows the statistical errors on w_0 and w_a as a function of $z^{\rm cut}_{\rm max}$ and $z^{\rm cut}_{\rm min}$ for our Wide survey, whose characteristics are similar to those expected from an LSST- or Euclid-like survey. The relative costs depend mildly upon survey parameters.

Excising galaxies with photometric redshifts lower than $z^{\rm ph} \sim 0.3$ results in only a $\sim$7% increase in the statistical errors on dark energy parameters. Likewise, excising galaxies with $z^{\rm ph} \gtrsim 2.4$ results in only a $\sim$10% degradation in w_0 and w_a constraints. Excising both of these regions of photometric redshift leads to a reduction in constraining power of $\lesssim$20%. Figure 6 is a valuable itemization of the statistical losses incurred by redshift cuts and indicates that excising low- and high-redshift portions of the imaging surveys may be an effective method to mitigate the influence of catastrophic photometric redshift errors at little cost in statistical error.

While Figure 6 quantifies the cost of excising regions of photometric redshift, the parametric complexity of catastrophic photo-z errors makes specific statements about the benefit of such cuts more difficult. In the case of a *localized* catastrophe that places galaxies erroneously in the excised high- or low-redshift ends of the survey, the induced bias can be nearly completely removed at the cost of the statistical degradation in Figure 6. We have begun a preliminary study of the benefits of

[10] Nishizawa et al. (2010) also study the ability to employ photometric redshift cuts to mitigate the effects of catastrophic outliers, which became available on the arXiv while we were submitting this manuscript for publication.

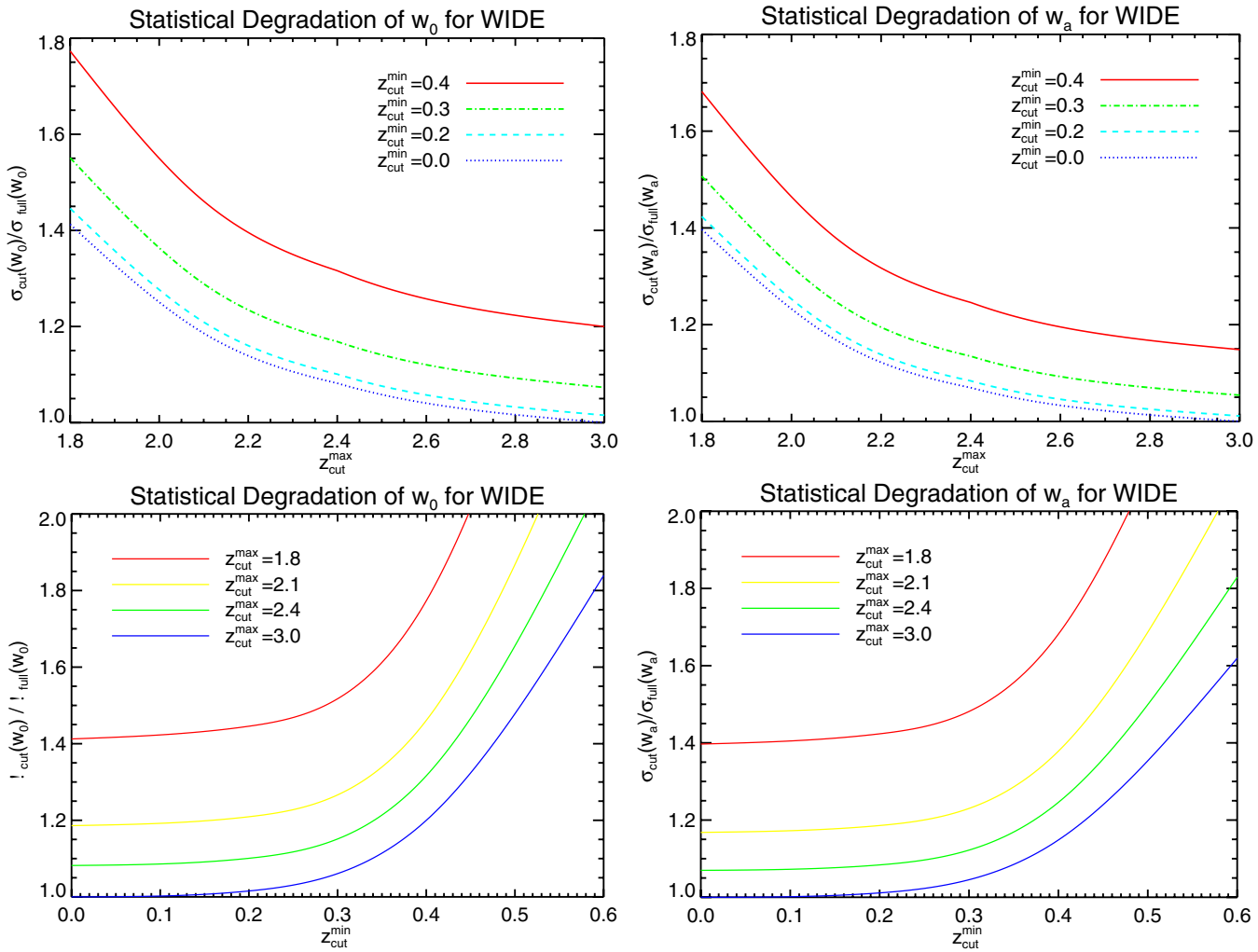

Figure 6. Statistical cost of excising low- and high-redshift shear information on constraints of w_0 (left panels) and w_a (right panels) for our WIDE survey. In the top row, the value of the maximum photometric redshift of the survey appears along the horizontal axis while the different lines show different choices of the minimum photometric redshift as indicated. Along the vertical axis is the fractional increase in dark energy parameter constraints relative to the constraints provided by a survey with our standard tomography. In the bottom row, the value of the minimum photometric redshift of the survey runs along the horizontal axis while the different lines show different choices of the maximum photometric redshift as indicated.

(A color version of this figure is available in the online journal.)

redshift excision, including the case of *uniform* catastrophes. In the case of our WIDE survey, excision can considerably reduce systematic errors induced even by the *uniform* catastrophe when the core is not well calibrated ($N_{\rm spec} \lesssim 10^5$), but this strategy is only of marginal value in the limit of a well-calibrated core. We limit the present discussion to the itemization in Figure 6 and relegate further study of redshift cuts and possible self-calibration of specific types of catastrophic error to a follow-up study.

4. CONCLUSIONS AND DISCUSSION

We have studied the potential systematic errors that may be induced in dark energy parameters inferred from forthcoming weak-lensing surveys as a result of a population of source galaxies with photometric redshifts that deviate significantly from their true redshifts. We used a particular operational definition of catastrophic photo-z errors that is subtly distinct from the use of this term in some of the existing literature. Throughout this work, the term *catastrophic photometric redshift error* refers to cases in which photo-z estimates differ significantly from true redshifts, the nature of the error has not been identified or calibrated with an accompanying spectroscopic data set, *and* the outlier population has not been removed reliably from the imaging data prior to the construction of shear correlation statistics. One way to interpret our results is as requirements for spectroscopic calibration of outliers and the completeness with which outlier galaxies must be culled from the data set in order to render systematic errors in dark energy parameters small.

In order to provide relatively general guidelines on the fidelity with which outlier photo-z's must be understood, we have taken an agnostic position on the nature of what types of catastrophic photometric redshift outliers may be realized in forthcoming imaging data. This eliminates the need to anticipate what types of photo-z errors may occur at very small fractional rates in order to assess their general influence on dark energy parameters. To be sure, there are reasonable guesses that can be made regarding the nature of photometric redshift errors and many algorithms exist that estimate redshifts from photometric data and refine estimates based upon comparisons with large, spectroscopic data sets (e.g., Bolzonella et al. 2000; Collister & Lahav 2004; Oyaizu et al. 2008; Feldmann et al. 2006; Brammer et al. 2008; Margoniner & Wittman 2008; Cunha et al. 2009; Ilbert et al.

2009; Coupon et al. 2009). However, we have not adopted any particular template for photometric redshift outliers. Instead, we have studied two extreme limiting cases of catastrophic photometric redshift error.

In the first class of photometric redshift error, which we dubbed the *uniform* catastrophe, photometric redshifts are poorly constrained and scattered over a broad range (see, e.g., Ilbert et al. 2009; Coupon et al. 2009, for examples of such features). Photo-z errors resembling our uniform type must be well controlled. If such errors occur even for a relatively small fraction of galaxies near the median redshift of a given survey, the systematic errors induced on dark energy parameter estimators will be significant. Roughly speaking, we find that the error rate per galaxy must be maintained at $F_{\rm cat} \lesssim$ a few $\times\ 10^{-4}$. However, the uniform catastrophic error is a relatively simple variety so that self-calibration may well be feasible. One could resign oneself to the fact that such an error will occur and add the error rate $F_{\rm cat}$ (and perhaps other parameters such as $\Delta z_{\rm cat}$) to the set of nuisance parameters to be marginalized over. This self-calibration could eliminate the systematic error, but will broaden statistical errors. We explore self-calibration of particular catastrophic photo-z errors in a forthcoming paper.

The second class of errors, which we refer to as *localized* catastrophes, takes source galaxies with particular true redshifts and assigns them photometric redshifts with a large bias but small scatter. Localized catastrophes have a broader range of possibilities and are more difficult to deal with. Figures 3 and 5 and Equation (4) constitute a blueprint for estimating the severity of a broad range of possible localized photometric redshift catastrophes. Quite generally, we find that the systematic errors they induce are sensitive to the scheme used to bin the source galaxies in photometric redshift. This suggests that an iterative scheme of re-binning may be an effective strategy for identifying and mitigating the influence of localized catastrophic photo-z errors.

In Section 3.4, we studied a simple strategy to limit the systematics induced by catastrophic photo-z outliers. First, we showed that the statistical leverage of the highest redshift ($z \gtrsim 2.4$) and lowest redshift ($z \lesssim 0.3$) source galaxies on dark energy constraints is minimal. Eliminating all such galaxies from consideration in inferring dark energy parameters results in only a small increase in the statistical errors of dark energy equation of state constraints, but may eliminate some of the most severe systematic errors induced by localized catastrophic photo-z outliers. This implies that well-designed cuts on $z^{\rm ph}$ will likely be a powerful and general means to mitigate systematics associated with photo-z determination at a relatively small cost in statistical error.

The published work that is most closely related to the present work is Bernstein & Huterer (2010). Our work is an extension and generalization of their study. Overall, we reach the same broad conclusions where the two studies are commensurable. In particular, we find that catastrophic errors of the localized variety must be controlled such that the rate of errors per galaxy is $F_{\rm cat} \lesssim 10^{-3}$ if they are to induce tolerable systematic errors on dark energy parameters.

Our work differs from and complements Bernstein & Huterer (2010) in several important ways. First, we have relaxed the assumption that the true redshift distribution of the outlier population perfectly traces that of the core population within individual source photometric redshift bins (see Equation (4)). Our treatment of photometric redshift errors is independent of the photometric redshift binning (as such errors would be in practice), while the approach of Bernstein & Huterer (2010) is limited to cases in which photometric errors both trace the galaxy distributions within the Source Bin and span the redshift range of the Source Bin. While contamination of the target redshift bin is typically the larger source of induced systematic error, our generalization illustrates that the effects of modifications to the Source Bin are non-negligible and in some cases these offsets contribute significantly to the systematic errors on dark energy parameters. Second, we have studied catastrophic errors in cases where the *core* photometric redshift distribution is not perfectly calibrated. Accounting for uncertainty in the core distribution turns out to be quite important: for a fixed catastrophe the magnitude of the induced systematic errors can vary by several orders of magnitude over a reasonable range of priors on the core distribution. Third, we have explored cases of correlated shifts in photo-z errors that span multiple tomographic redshift bins (which will occur in practice), the extreme example being the *uniform* error.

We conclude our discussion section by referring to interesting, tangential results given in the Appendix. In the Appendix, we discuss the effect of different models of the nonlinear evolution of cosmological density perturbations on photometric redshift calibration requirements. Weak-lensing measurements take significant advantage of measurements on nonlinear scales in order to constrain cosmology. Previous work on the calibration of photometric redshifts has utilized the Peacock & Dodds (1996) formula (e.g., Ma et al. 2006; Ma & Bernstein 2008); however, we find that using the more recent and more accurate fit of Smith et al. (2003) significantly reduces the need for independent calibration of photometric redshifts. We have used the Smith et al. (2003) formula in the main body of this paper. We refer the reader to the Appendix for further details.

5. SUMMARY

We have adopted a simple, agnostic approach to estimate the levels at which uncalibrated photometric redshift outliers must be controlled to maximize the dark energy constraints from the weak-lensing components of forthcoming imaging surveys such as DES, LSST, EUCLID, and JDEM. We present results for three fiducial imaging surveys: a relatively near-term DES-like survey, a future survey with a high surface density of galaxies but a relatively small fractional sky coverage (DEEP), and a future survey with half-sky coverage and a lower galaxy surface density (WIDE). We considered two extreme cases of large, uncalibrated errors. In the case of a *uniform* photo-z catastrophe, we considered galaxies erroneously assigned photometric redshifts that are unrelated to their true redshifts. In the case of a *localized* photo-z catastrophe, we considered the erroneous placement of a small fraction of galaxies in some range of true redshifts at significantly different *photometric* redshifts. To be specific, we assigned galaxies in some range of true redshifts of width Δz centered on a true redshift $z_{\rm cat}$ to photometric redshifts near $z^{\rm ph}_{\rm cat}$ that differ significantly from $z_{\rm cat}$. For each type of error and survey, we assessed the severity of the systematic errors on dark energy parameters that would be induced by catastrophic photometric redshift errors. Our primary results are as follows.

1. A photometric redshift error of the *uniform* variety that is relevant for galaxies near the median redshift of the imaging survey, must be limited to a fraction of galaxies

$F_{\rm cat} \lesssim 5 \times 10^{-4}$ for DES or DEEP and $F_{\rm cat} \lesssim 2 \times 10^{-4}$ for WIDE, in order to induce systematic errors that are small compared to the statistical errors on w_0 and $w_{\rm a}$.

2. *Localized* catastrophic errors are most severe when they take some fraction of galaxies with true redshifts near the median survey redshift and assign them significantly higher or lower photo-z's. For DES, assignments to higher photo-z's are more severe than assignments to lower photo-z's while the opposite is true for WIDE and DEEP. However, the systematic errors induced by these two extremes differ by less than a factor of 2 in all cases.
3. Limiting the fraction of galaxies exhibiting *localized* catastrophes at all redshifts to $F_{\rm cat} \lesssim 3 \times 10^{-3}$ for DES or $F_{\rm cat} \lesssim 10^{-3}$ for WIDE or DEEP will render them unimportant. For localized catastrophes that occur over a range of true redshifts of width $\Delta z_{\rm cat}$ near the median survey redshift, the fractional error rate must be controlled such that $F_{\rm cat}(\Delta z_{\rm cat}/0.1) \lesssim 1\text{–}3 \times 10^{-3}$.
4. Imperfect knowledge of the photo-z distribution for the *core* sample of galaxies loosens these requirements for uncalibrated catastrophic outlier control as depicted in Figure 5. Roughly speaking, core calibration with spectroscopic samples smaller than the statistical equivalent of $N_{\rm spec} \lesssim 10^5$ leads to significantly reduced catastrophic error control requirements. Of course, in practice catastrophic error control and core calibration will both improve as $N_{\rm spec}$ increases.
5. The statistical leverage of the highest redshift ($z \gtrsim 2.4$) and lowest redshift ($z \lesssim 0.3$) source galaxies on dark energy constraints is small. Eliminating all such galaxies from consideration in inferring dark energy parameters results in a $\lesssim$20% increase in the statistical errors on dark energy, but may eliminate the most severe systematic errors induced by localized catastrophic photo-z outliers.
6. In the Appendix, we show that dark energy parameter forecasts that include photometric redshift uncertainty vary significantly depending upon the treatment of the nonlinearity in the matter power spectrum. In particular, using the Smith et al. (2003) fitting form (as we do in the main text) leads to weaker photo-z calibration requirements than does the Peacock & Dodds (1996) formula upon which the results of Ma et al. (2006) are based. The Smith et al. (2003) formula has been shown to be more accurate than Peacock & Dodds (1996) suggesting that degradation due to photo-z uncertainty may be less than Ma et al. (2006) forecast. Only a rigorous numerical study can determine this definitively.

This level of photometric redshift outlier control is challenging in comparison to the yields of contemporary methods and data. Existing spectroscopic samples are not representative of the galaxy populations that will be utilized to constrain dark energy in forthcoming imaging surveys. For example, the Deep Extragalactic Evolutionary Probe (DEEP2) has a 70% success rate for obtaining spectroscopic redshifts (Cooper et al. 2006), where star-forming galaxies with $z > 1.4$ constitute roughly half of the failed targets (Freeman et al. 2009). Moreover, spurious photo-z outliers persist even with techniques developed in conjunction with spectroscopic data that span the region of parameter space occupied by the photometric sample. As a nearly contemporaneous example, Nishizawa et al. (2010) construct galaxies using simple spectral templates assuming a number of particular stellar populations based on the COSMOS galaxy catalog. Applying LePhare[11] to their mock spectra gives offset islands in $z - z^{\rm ph}$ space containing more than 5% of the probability, which remains true even after refining their redshift estimator. In the analysis of the COSMOS data, Ilbert et al. (2009) achieve an outlier rate of 0.7% for a subsample of their brightest objects ($17.5 < i_{\rm AB} < 22.5$). However, their outlier rate dramatically increases to 15.3% when they apply their photo-z techniques to a subsample of faint objects ($22.5 < i_{\rm AB} < 24$). An alternative approach to photo-z calibration is adopted in Cunha et al. (2009), who applied their weighted training set method to both simulated and actual SDSS data. Their methods substantially improve upon the ability to directly reconstruct the redshift distribution of a photometric sample, but errors in the reconstructed $N(z)$ remain at the percent level. Thus, while contemporary photo-z codes do provide useful guidance, outlier fractions greater than $\sim 10^{-3}$ persist and can affect the dark energy program. The ability to either limit, or understand, such outlier populations significantly better than the current state-of-the-art will be necessary to exploit fully the promise of cosmic shear tomography.

[11] http://www.cfht.hawaii.edu/~arnouts/LEPHARE/cfht_lephare/lephare.html

We thank Carlos Cunha, Joe Davola, Scott Dodelson, Salman Habib, Wayne Hu, Zeljko Ivezic, Bhuvnesh Jain, Arthur Kosowsky, Dan Matthews, Jeff Newman, Hiro Oyaizu, Martin White, and Michael Wood-Vasey for useful discussions and email exchanges during the course of this work. We are particularly grateful to Gary Bernstein and Alexia Schulz for detailed comments on an early draft of this manuscript. A.P.H. and A.R.Z. are supported by the University of Pittsburgh and by the National Science Foundation (NSF) through grant NSF AST 0806367 and by the Department of Energy (DOE). Z.M. is supported by the DOE under contracts DOE-DE-AC02-98CH10886 and DOE-DE-FG02-95ER40893. D.H. is supported by the DOE OJI under contract DOE-DE-FG02-95ER40899, NSF under grant NSF AST 0807564, and the National Aeronautics and Space Administration under grant NNX09AC89G. A.R.Z. and A.P.H. thank the organizers of the 2009 Santa Fe Cosmology Workshop, where a significant portion of this work was completed. A.R.Z. thanks the Michigan Center for Theoretical Physics at the University of Michigan for hospitality and support while some of this work was performed. This research made use of the National Aeronautics and Space Administration Astrophysics Data System.

APPENDIX

THE NONLINEAR POWER SPECTRUM AND PHOTOMETRIC REDSHIFT CALIBRATION REQUIREMENTS

Much of the constraining power of weak-lensing surveys arises from measurements on scales where the structures causing the deflections are undergoing nonlinear gravitational evolution (e.g., Huterer & Takada 2005). Restricting consideration to large scales significantly degrades cosmological constraints (e.g., Huterer 2002; Huterer & Takada 2005; Zentner et al. 2008; Schmidt 2008; Hearin & Zentner 2009), so it is necessary to model nonlinear evolution in order to utilize weak lensing to constrain dark energy. At least three approximate and related techniques are in common use: (1) the fitting formula of Peacock & Dodds (1996), which is based on the HKLM method (Hamilton et al. 1991), (2) the halo model (Scherrer & Bertschinger 1991; Seljak 2000; Ma & Fry 2000; Scoccimarro et al. 2001; Cooray & Sheth 2002), and (3) the fitting formula of Smith et al. (2003). The works of Ma et al. (2006) and Ma &

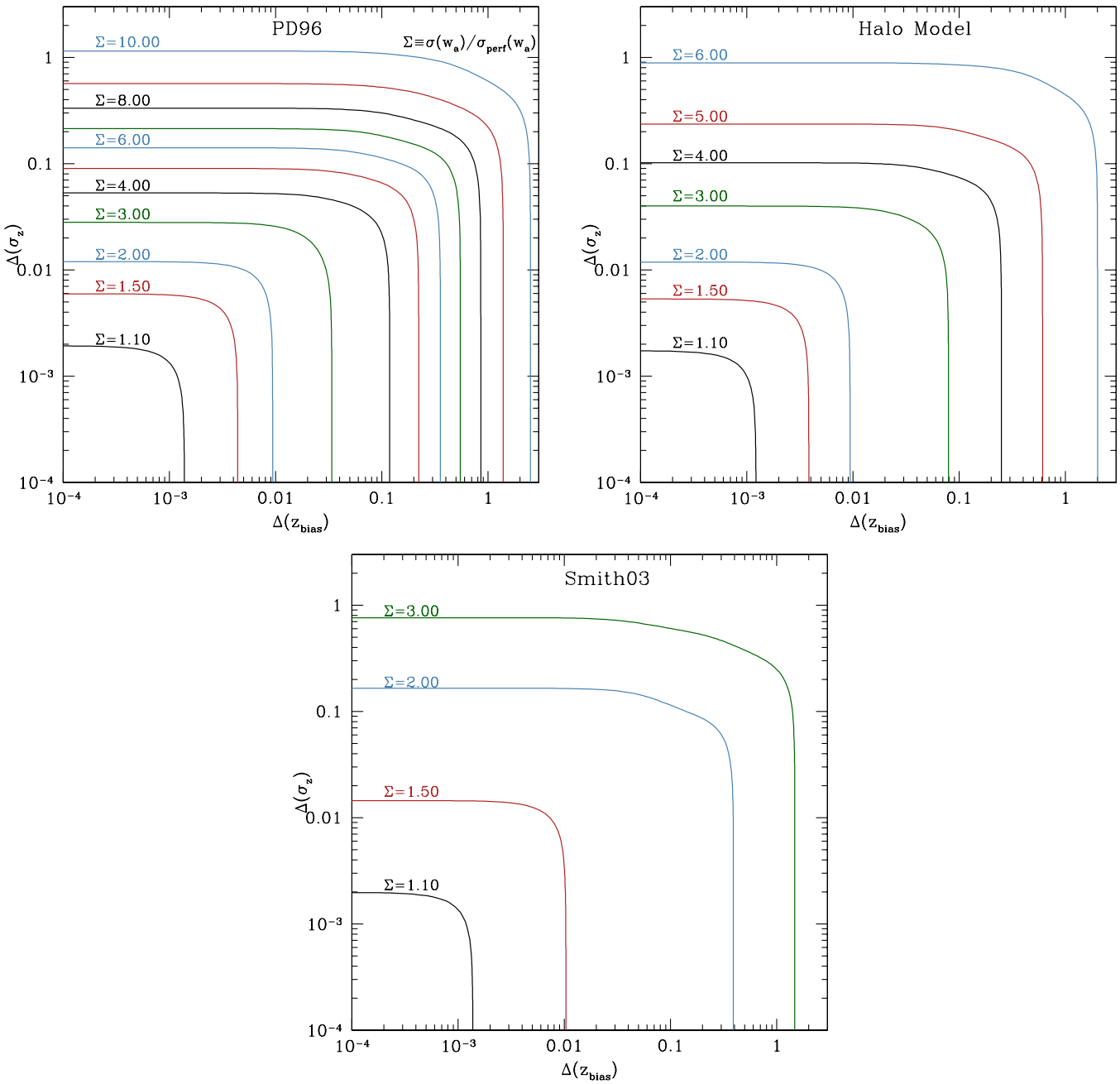

Figure 7. Contour plots for the level of w_a constraint degradation as a function of priors on the photometric redshift scatter σ_z and bias $z^{\rm bias}$. In this case, the priors are applied uniformly to the photometric redshift parameters are each redshift. The contours demarcate equal parameter degradation defined as the error on w_a after marginalizing over photometric redshift uncertainties. We show constraints in units of the equivalent constraint in the limit of perfect knowledge of photometric redshift parameters, $\Sigma \equiv \sigma(w_a)/\sigma_{\rm perf}(w_a)$. The upper, left panel was computed using the Peacock & Dodds (1996) fitting formula for the nonlinear power spectrum of density fluctuations and amounts to a near reproduction of the right panel of Figure 7 in Ma et al. (2006). The upper, right panel was computed using the halo model as described in Zentner et al. (2008). The bottom panel was computed using the Smith et al. (2003) relation for the nonlinear power spectrum of density fluctuations. Significant differences between the levels of degradation are evident. Note that in this figure, we use a different set of cosmological and experimental parameters so that this result is directly comparable to those in Figure 7 of Ma et al. (2006).

(A color version of this figure is available in the online journal.)

Bernstein (2008) specifying requirements for photometric redshift calibration employ the Peacock & Dodds (1996) relation.

In the course of our study, we have recomputed the photometric redshift calibration requirements using each of the three approximate techniques mentioned in the previous paragraph. In the limit of perfect knowledge of the photometric redshift distribution, each of these fitting formulas gives nearly identical dark energy constraints. However, we have found that the photometric redshift calibration requirements have a strong dependence upon the method used to model nonlinear structure. We summarize this finding in Figure 7 where we display contours of constant degradation in the statistical error on w_a as a function of both the prior on the bias $\Delta z^{\rm bias}$ and the prior on the dispersion σ_z. In other words, we show contours of $\sigma(w_a)$ in units of the statistical constraint on w_a in the limit that the photo-z distribution parameters are known perfectly prior to the weak-lensing analysis, $\sigma_{\rm perf}(w_a)$. We assume that the same priors are applied to all of our 31 dispersion parameters and

31 bias parameters at each redshift bin (see Section 2.1.1). We summarize our findings in this way so that these results can be compared directly to Figure 7 in Ma et al. (2006). To make the comparison as direct as possible, we have computed these forecasts using the fiducial cosmology and experimental setup of Ma et al. (2006), which differs slightly from those considered in the main text. In this Appendix only, our fiducial cosmology is $\omega_{\rm M} = 0.14$, $\omega_{\rm B} = 0.024$, $n_{\rm s} = 1.0$, $\Delta^2_{\mathcal{R}} = 2.4 \times 10^{-5}$ (giving $\sigma_8 \simeq 0.91$), and $\Omega_{\rm DE} = 0.73$ combined with experimental parameters of $f_{\rm sky} = 0.1$ and $N^A = 55$ arcmin^{-2}.

The upper left panel in Figure 7 shows photo-z calibration requirements estimated using the Peacock & Dodds (1996) treatment of nonlinear power. This panel shows nearly identical results to those in Figure 7 of Ma et al. (2006) so that this panel validates our methods and provides a useful baseline to compare with the other panels. According to this result, ensuring that constraints on $w_{\rm a}$ are not degraded by more than a factor of 2 requires knowing the photo-z dispersion and bias to roughly $\sim$1% prior to undertaking the weak-lensing analysis. The upper, right panel of Figure 7 shows the same requirements constructed using the halo model for nonlinear clustering. In the limit of restrictive prior knowledge of the photo-z distribution, the Peacock & Dodds (1996) and halo model results yield nearly the same constraints. When the photo-z distributions have relatively unrestrictive priors, the two techniques yield moderately different levels of projected degradation with, for example, uncertainty in the photo-z dispersion of $\Delta(\sigma_z) \approx 1$ corresponding to a factor of 10 degradation in the Peacock & Dodds (1996) case but a factor of 6 degradation in the halo model calculation.

The largest differences among the forecasts comes from comparing the requirements using Peacock & Dodds (1996) to those computed using the Smith et al. (2003) fit. As with the halo model comparison, the different techniques agree well when prior knowledge of the photo-z distribution is very restrictive; thus as long as degradations due to photo-z uncertainty are $\lesssim$10%–20% then it does not matter which technique one uses to predict the nonlinear evolution. It is interesting that the constraints in the case of the halo model treatment degrade significantly less rapidly as prior knowledge becomes less and less restrictive. Turning to the Smith et al. (2003) fit, one would conclude that ensuring less than a factor of 2 degradation on the $w_{\rm a}$ constraint requires $\sim$18% knowledge of the dispersion and $\sim$40% knowledge of the bias as compared to the $\sim$1% requirements that result from the Peacock & Dodds (1996) analysis.

Clearly, at most one of these treatments can represent the growth of cosmic structure faithfully. In the main text, we presented results using the Smith et al. (2003) formula because these authors perform a detailed numerical study that finds the Peacock & Dodds (1996) and simple implementations of the halo model to be imprecise on scales relevant for cosmic shear cosmology. In the context of these fitting formulae, we find that Smith et al. (2003) predicts greater power than Peacock & Dodds (1996) on scales most relevant to lensing ($0.1 \lesssim k/h$ Mpc$^{-1} \lesssim 10$), particularly at high redshift. At this point, it is not possible to make a firm statement as to which approach is correct, but an exhaustive simulation program similar to that being carried out by Heitmann et al. (2005, 2008, 2009) may be capable of providing a more definitive resolution in the case of dissipationless evolution. Additional effort will be needed to treat any modifications induced by the baryonic component of the universe (White 2004; Zhan & Knox 2004; Jing et al. 2006; Rudd et al. 2007; Zentner et al. 2008; Stanek et al. 2009; Guillet et al. 2009).

REFERENCES

Abdalla, F. B., Banerji, M., Lahav, O., & Rashkov, V. 2008, arXiv:0812.3831
Abell, P., et al. 2009, arXiv:0912.0201
Albrecht, A., et al. 2006, arXiv:astro-ph/0609591
Banerji, M., Abdalla, F. B., Lahav, O., & Lin, H. 2008, MNRAS, 386, 1219
Benjamin, J., et al. 2007, MNRAS, 381, 702
Bernstein, G., & Huterer, D. 2010, MNRAS, 401, 1399
Bolzonella, M., Miralles, J., & Pelló, R. 2000, A&A, 363, 476
Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
Chevallier, M., & Polarski, D. 2001, Int. J. Mod. Phys. D, 10, 213
Collister, A. A., & Lahav, O. 2004, PASP, 116, 345
Cooper, M. C., et al. 2006, MNRAS, 370, 198
Cooray, A., & Hu, W. 2001, ApJ, 554, 56
Cooray, A., & Sheth, R. 2002, Phys. Rep., 372, 1
Coupon, J., et al. 2009, A&A, 500, 981
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, AJ, 136, 1361
Dodelson, S., Shapiro, C., & White, M. 2006, Phys. Rev. D, 73, 023009
Dodelson, S., & Zhang, P. 2005, Phys. Rev. D, 72, 083001
Doré, O., et al. 2007, arXiv:0712.1599
Feldmann, R., et al. 2006, MNRAS, 372, 565
Freeman, P. E., Newman, J. A., Lee, A. B., Richards, J. W., & Schafer, C. M. 2009, MNRAS, 398, 2012
Fu, L., et al. 2008, A&A, 479, 9
Guillet, T., Teyssier, R., & Colombi, S. 2009, arXiv:0905.2615
Hamilton, A. J. S., Kumar, P., Lu, E., & Matthews, A. 1991, ApJ, 374, L1
Hearin, A. P., & Zentner, A. R. 2009, J. Cosmol. Astropart. Phys., JCAP04(2009)032
Heavens, A. 2003, MNRAS, 343, 1327
Heitmann, K., Higdon, D., White, M., Habib, S., Williams, B. J., & Wagner, C. 2009, arXiv:0902.0429
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, arXiv:0812.1052
Hoekstra, H., & Jain, B. 2008, Ann. Rev. Nucl. Part. Sci., 58, 99
Hoekstra, H., Yee, H. K. C., & Gladders, M. D. 2002, ApJ, 577, 595
Hu, W. 1999, ApJ, 522, L21
Hu, W., & Tegmark, M. 1999, ApJ, 514, L65
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
Ilbert, O., et al. 2009, ApJ, 690, 1236
Ishak, M. 2005, MNRAS, 363, 469
Jarvis, M., Bernstein, G. M., Fischer, P., Smith, D., Jain, B., Tyson, J. A., & Wittman, D. 2003, AJ, 125, 1014
Jarvis, M., Jain, B., Bernstein, G., & Dolney, D. 2006, ApJ, 644, 71
Jing, Y. P., Zhang, P., Lin, W. P., Gao, L., & Springel, V. 2006, ApJ, 640, L119
Jouvel, S., et al. 2009, A&A, 504, 359
Jungman, G., Kamionkowski, M., Kosowsky, A., & Spergel, D. N. 1996, Phys. Rev. D, 54, 1332
Kitching, T. D., Heavens, A. F., Taylor, A. N., Brown, M. L., Meisenheimer, K., Wolf, C., Gray, M. E., & Bacon, D. J. 2007, MNRAS, 376, 771
Kitching, T. D., Taylor, A. N., & Heavens, A. F. 2008, MNRAS, 389, 173
Komatsu, E., et al. 2008, arXiv:0803.0547
Kosowsky, A., Milosavljevic, M., & Jimenez, R. 2002, Phys. Rev. D, 66, 063007
Lima, M., Cunha, C. E., Oyaizu, H., Frieman, J., Lin, H., & Sheldon, E. S. 2008, MNRAS, 390, 118
Lima, M., & Hu, W. 2007, Phys. Rev. D, 76, 123013
Linder, E. V. 2003, Phys. Rev. Lett., 90, 091301
Ma, Z., & Bernstein, G. 2008, ApJ, 682, 39
Ma, C.-P., & Fry, J. N. 2000, ApJ, 543, 503
Ma, Z., Hu, W., & Huterer, D. 2006, ApJ, 636, 21
Margoniner, V. E., & Wittman, D. M. 2008, ApJ, 679, 31
Munshi, D., Valageas, P., van Waerbeke, L., & Heavens, A. 2008, Phys. Rep., 462, 67
Newman, J. A. 2008, ApJ, 684, 88
Nishizawa, A. J., Takada, M., Hamana, T., & Furusawa, H. 2010, arXiv:1002.2476

Oyaizu, H., Lima, M., Cunha, C. E., Lin, H., & Frieman, J. 2008, ApJ, 689, 709
Peacock, J. A., & Dodds, S. J. 1996, MNRAS, 280, L19
Pen, U.-L., Zhang, T., van Waerbeke, L., Mellier, Y., Zhang, P., & Dubinski, J. 2003, ApJ, 592, 664
Refregier, A. 2003, ARA&A, 41, 645
Refregier, A., Amara, A., Kitching, T. D., Rassat, A., Scaramella, R., Weller, J., & Euclid Imaging Consortium 2010, arXiv:1001.0061
Refregier, A., et al. 2004, AJ, 127, 3102
Rudd, D. H., Zentner, A. R., & Kravtsov, A. V. 2007, ApJ, 672, 19
Scherrer, R. J., & Bertschinger, E. 1991, ApJ, 381, 349
Schmidt, F. 2008, Phys. Rev. D, 78, 043002
Schulz, A. E. 2009, arXiv:0910.3683
Scoccimarro, R., Sheth, R. K., Hui, L., & Jain, B. 2001, ApJ, 546, 20
Seljak, U. 1997, ApJ, 482, 6
Seljak, U. 2000, MNRAS, 318, 203
Semboloni, E., et al. 2006, A&A, 452, 51
Smith, R. E., et al. 2003, MNRAS, 341, 1311
Song, Y.-S., & Knox, L. 2004, Phys. Rev. D, 70, 063510
Stanek, R., Rudd, D., & Evrard, A. E. 2009, MNRAS, 394, L11
Sun, L., Fan, Z., Tao, C., Kneib, J., Jouvel, S., & Tilquin, A. 2009, ApJ, 699, 958
Takada, M., & Jain, B. 2004, MNRAS, 348, 897
Takada, M., & White, M. 2004, ApJ, 601, L1
Tegmark, M., Taylor, A. N., & Heavens, A. F. 1997, ApJ, 480, 22
Vale, C., & White, M. 2003, ApJ, 592, 699
Van Waerbeke, L., Mellier, Y., & Hoekstra, H. 2005, A&A, 429, 75
White, M. 2004, Astropart. Phys., 22, 211
White, M., & Hu, W. 2000, ApJ, 537, 1
Zentner, A. R., & Bhattacharya, S. 2009, ApJ, 693, 1543
Zentner, A. R., Rudd, D. H., & Hu, W. 2008, Phys. Rev. D, 77, 043507
Zhan, H. 2006, J. Cosmol. Astropart. Phys., JCAP08(2006)008
Zhan, H., & Knox, L. 2004, ApJ, 616, L75
Zhang, P., Pen, U., & Bernstein, G. 2009, arXiv:0910.4181
Zhao, G.-B., Pogosian, L., Silvestri, A., & Zylberberg, J. 2009, arXiv:0905.1326