Universiteit
Leiden
The Netherlands

Cosmic tomography with weak gravitational lensing

Li, S.S.

Citation
Li, S. S. (2023, October 25). *Cosmic tomography with weak gravitational lensing*. Retrieved from https://hdl.handle.net/1887/3645974

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3645974

Note: To cite this publication please use the final published version (if applicable).

CHAPTER 2

KiDS+VIKING-450: An internal-consistency test for cosmic shear tomography with a colour-based split of source galaxies

Abstract

We performed an internal consistency test for the KiDS+VIKING-450 (KV450) cosmic shear analysis, focusing on the colour-based split of source galaxies. Using the same measurements and calibrations across both sub-samples, we evaluated the properties of the shear measurements and the efficiency of the calibration pipelines. On the modelling side, we explored the observational nuisance parameters, particularly those associated with redshift calibration and intrinsic alignments, using a Bayesian analysis equipped with specific test parameters. Our study confirmed that the current nuisance parameters adequately capture residual systematic deviations in the KV450 data, albeit with minor discrepancies observed in the second and third redshift bins. Our results additionally revealed the degeneracy between the apparent amplitude of intrinsic alignments and redshift uncertainties in low redshift bins. Given the relative insensitivity of our test to the assumed cosmological model, it can be implemented in the cosmic shear analysis prior to drawing any cosmological conclusions.

S.-S. Li, K. Kuijken, H. Hoekstra,
H. Hildebrandt, B. Joachimi, and A. Kannawadi
Astronomy & Astrophysics, 646, A175

2.1 Introduction

Cosmic shear, a coherent distortion of distant galaxy shapes due to weak gravitational lensing by large-scale structures, is sensitive to the amplitude of matter density fluctuations (quantified by σ_8, which is the standard deviation of linear-theory density fluctuations in a sphere of radius $8h^{-1}$ Mpc, where $H_0 = 100h$ km s^{-1} Mpc^{-1}) and the mean matter density (Ω_m). As a result, cosmic shear surveys commonly report a derived parameter $S_8 \equiv \sigma_8(\Omega_\mathrm{m}/0.3)^{0.5}$. Alternatively, cosmic microwave background (CMB) measurements can infer local density fluctuations by extrapolating the measured amplitude of temperature fluctuations at recombination, based on a cosmological model. By comparing results from these two distinct probes, we can test cosmological models.

The latest *Planck* Legacy analysis of CMB measurements predicts an $S_8 = 0.832 \pm 0.013$ (68% credible region), assuming the standard Λ cold dark matter (ΛCDM) model (Planck Collaboration et al. 2020). This results is slightly higher than the results from the recent cosmic shear surveys, such as the Dark Energy Survey (DES; Troxel et al. 2018, $S_8 = 0.782^{+0.027}_{-0.027}$), the Hyper Suprime-Cam Subaru Strategic Program (HSC; Hikage et al. 2019, $S_8 = 0.780^{+0.030}_{-0.033}$), and especially the Kilo-Degree Survey (KiDS; Hildebrandt et al. 2020, hereafter H20, $S_8 = 0.737^{+0.040}_{-0.036}$).

It is crucial to consider potential systematic effects associated with observations when interpreting results from different surveys. Internal consistency tests are thus a standard part of any cosmological probe. Cosmic shear studies often base these checks on a split of the estimated two-point shear correlations (Köhlinger et al. 2019; or Sect. 7.4 of H20). By assigning duplicate model parameters to each subset, one can perform theoretical modelling of the reconstructed data vector and assess data consistency by comparing these duplicate parameters. This approach aids in verifying the consistency of a specific sample of source galaxies, but it only tests at a late stage in the analysis. Moreover, doubling cosmological parameters incurs a significant computational cost, which hinders further splitting of the source sample in practice. However, additional splits could be particularly interesting, as systematics may vary among them.

Source galaxy properties present two main challenges to calibration pipelines: shape measurements and redshift estimates. First, different galaxy samples typically exhibit varying ellipticity distributions, with red, early-type galaxies tending to have rounder shapes than their blue, late-type counterparts (Hill et al. 2019; Kannawadi et al. 2019, hereafter K19). This leads to a correlation between shear bias and the underlying galaxy sample, primarily because shape measurements are sensitive to ellipticity distributions. For instance, the *lens*fit algorithm used in the KiDS survey assigns weights to measured ellipticities, resulting in a bias towards intermediate ellipticity values (Fenech Conti et al. 2017). Second, both the accuracy and precision of photometric redshift estimates depend on a galaxy's broad spectral features, such as the Balmer break below 4000 (Salvato et al. 2019). The prominence of these broad spectral features varies by galaxy spectral type. Generally, galaxies with an older stellar population appear red at rest-frame optical wavelengths and exhibit a pronounced 4000 break. The more young stars a galaxy contains, the bluer it appears, causing the Balmer break and other broad spectral features

to be washed out. Consequently, errors in photometric redshifts correlate with galaxy spectral type (Mo et al. 2010).

In this chapter, we consider these sample-related systematic effects, specifically focusing on photometric redshift uncertainty. We divided the KiDS source galaxies into two mutually exclusive sub-samples based on their spectral types and applied the same measurement and calibration pipelines to each sub-sample. By doing this, we explored how sample-related systematics can affect measurements and assessed the effectiveness of calibration pipelines in mitigating these effects. This split also has implications for modelling intrinsic alignments, which must be explicitly accounted for. To quantify consistency, we performed a Bayesian analysis using dedicated test parameters that describe the relative deviations of nuisance parameters between the two sub-samples. By examining their posterior distributions, we can determine if the original settings are sufficient to capture residual biases. The analysis code is publicly available[1].

Our approach complements other studies that assess the consistency of inferred cosmological parameters by removing tomographic bins or by splitting the sample by galaxy type, while marginalising over the corresponding nuisance parameters (Köhlinger et al. 2019; Samuroff et al. 2019). Instead, we focused on a different aspect: keeping cosmological parameters fixed while examining changes in the nuisance parameters. We found that our method can effectively test for inconsistencies in redshift distributions and highlight the degeneracy between redshift uncertainties and apparent intrinsic alignment signals, all while remaining insensitive to cosmological assumptions.

The remainder of this chapter is organised as follows. In Sect. 2.2, we provide a brief overview of the cosmic shear catalogues under consideration. We present the redshift calibration in Sect. 2.3 and the shear bias calibration in Sect. 2.4. Next, we introduce the measurement and modelling of the shear signal in Sect. 2.5. We discuss the covariance matrix and consistency tests in Sect. 2.6. Our main results are presented in Sect. 2.7, and we conclude with a summary in Sect. 2.8.

2.2 Data

Our test is based on the first release of optical and infrared KiDS cosmic shear data, known as KiDS+VIKING-450 (KV450; Wright et al. 2019, hereafter W19). This data set includes four-band optical photometry ($ugri$) from the first three data releases of KiDS (de Jong et al. 2015, 2017) and five-band near-infrared photometry ($ZYJHK_s$) from the overlapping VISTA Kilo-Degree Infrared Galaxy Survey (VIKING, Edge et al. 2013).

Details on the derivation and verification of this cosmic shear catalogue can be found in the main KiDS cosmic shear papers (Hildebrandt et al. 2017; H20) and their companion papers (Fenech Conti et al. 2017; W19). The public catalogue provides all the necessary information for conducting a tomographic cosmic shear analysis. Among the most important columns are the photometric redshifts (photo-zs, or z_B as in the catalogues) and the galaxy shapes (described by two ellipticity components, ϵ_1 and ϵ_2).

[1] https://github.com/lshuns/CosmicShearRB

The z_B values are estimated using the Bayesian photometric redshift code (BPZ, Benítez 2000; Coe et al. 2006) with an improved redshift prior from Raichoor et al. (2014) and the nine-band photometry from W19. The galaxy shapes are measured from the r-band images (median seeing of 0.″7) using the *lens*fit algorithm (Miller et al. 2007; Kitching et al. 2008; Miller et al. 2013) with a self-calibration for noise bias (Fenech Conti et al. 2017).

Throughout this study, we only use sources with valid nine-band photometry (`GAAP_Flag_ugriZYJHKs==0`). This mask reduces the original area by approximately 5% and retains around 13 million objects, which is identical to the choice made by the main KV450 cosmic shear analysis. Following H20, we bin source galaxies into five tomographic bins defined as $0.1 < z_B \leq 0.3$, $0.3 < z_B \leq 0.5$, $0.5 < z_B \leq 0.7$, $0.7 < z_B \leq 0.9$, $0.9 < z_B \leq 1.2$. To check for systematic effects caused by galaxy properties, we further split the whole sample into two sub-samples based on the spectral types of source galaxies. This is achieved by using the T_B values reported by the BPZ code during the photo-z estimation procedure (see Benítez 2000, for a detailed discussion). Briefly, the T_B value is calculated within a Bayesian framework using six templates of galaxy spectra (Coleman et al. 1980; Kinney et al. 1996). We define our two sub-samples as $T_B \leq 3$ (a combination of E1, Sbc, Scd types, labelled as 'red') and $T_B > 3$ (a combination of Im and two starburst types, labelled as 'blue'). This cut is chosen to ensure similar statistical power in the two sub-samples (see Fig. 2.1). Source properties of these two sub-samples are summarised in Table 2.1.

2.3 Calibration of redshift distributions

One of the most challenging tasks in a tomographic cosmic shear study is estimating the source redshift distribution for each tomographic bin. Since these intrinsic redshift distributions vary with galaxy samples, we need to calibrate the photo-z estimates for the two sub-samples separately. We employed the technique from the fiducial KV450 cosmic shear analysis, known as DIR in H20, for this task. This method directly estimates the underlying redshift distributions of a photometric sample using deep spectroscopic redshift (spec-z) catalogues that overlap with the photometric survey. In this section, we briefly discuss our implementation of this method and refer interested readers to the original papers for more details (Lima et al. 2008; Hildebrandt et al. 2017, 2020).

The DIR method necessitates that the calibration sample (the spec-z sample) spans, at least sparsely, the full extent of the multi-band magnitude space covered by the target sample (the photo-z sample), and that the mapping from magnitude space to redshift space is unique. Consequently, the coverage of the spec-z sample is essential for this method's accuracy. We used the same set of spec-z catalogues as employed in the fiducial KV450 cosmic shear analysis. This includes the zCOSMOS survey (Lilly et al. 2009), the DEEP2 survey (Newman et al. 2013), the VIMOS VLT Deep survey (Le Fèvre et al. 2013), the GAMA-G15Deep survey (Kafle et al. 2018), and a combined catalogue provided by ESO in the *Chandra* Deep Field South area. These independent spec-z surveys, with different lines-of-sight and depths, minimise shot noise and sample

Table 2.1: Source information in the two sub-samples.

Sample	Bin	Photo-z range	Total *lens*fit weights	$n_{\rm eff}$ $\left[\mathrm{arcmin}^{-2}\right]$	$\sigma_{\epsilon,i}$	m-bias	Mean($z_{\rm DIR}$)	Median($z_{\rm DIR}$)
$T_{\rm B} \leq 3$	1	$0.1 < z_{\rm B} \leq 0.3$	7 031 963	0.38	0.279	-0.029 ± 0.010	0.351	0.282
(red)	2	$0.3 < z_{\rm B} \leq 0.5$	10 404 223	0.59	0.252	-0.009 ± 0.007	0.430	0.396
	3	$0.5 < z_{\rm B} \leq 0.7$	15 508 696	0.90	0.276	-0.010 ± 0.007	0.546	0.531
	4	$0.7 < z_{\rm B} \leq 0.9$	9 837 460	0.64	0.250	0.008 ± 0.006	0.744	0.732
	5	$0.9 < z_{\rm B} \leq 1.2$	8 466 542	0.59	0.275	0.006 ± 0.008	0.909	0.894
$T_{\rm B} > 3$	1	$0.1 < z_{\rm B} \leq 0.3$	7 269 125	0.42	0.270	-0.004 ± 0.008	0.437	0.244
(blue)	2	$0.3 < z_{\rm B} \leq 0.5$	12 200 673	0.75	0.277	-0.007 ± 0.006	0.573	0.431
	3	$0.5 < z_{\rm B} \leq 0.7$	21 116 034	1.46	0.292	-0.002 ± 0.006	0.791	0.644
	4	$0.7 < z_{\rm B} \leq 0.9$	12 134 896	0.92	0.286	0.026 ± 0.006	0.914	0.842
	5	$0.9 < z_{\rm B} \leq 1.2$	10 207 426	0.87	0.293	0.036 ± 0.009	1.081	1.022

The effective number density $n_{\rm eff}$ is calculated using Eq. (1) of Heymans et al. (2012). The reported ellipticity dispersion is defined as $\sigma_{\epsilon,i} = (\sigma_{\epsilon 1} + \sigma_{\epsilon 2})/2$. The m-bias is defined in Eq. (2.1) and detailed in Sect. 2.4. Uncertainties reported were computed from the dispersion of 50 bootstrap samples. The mean and median of the redshift distributions were obtained from the DIR calibration, detailed in Sect.2.3.

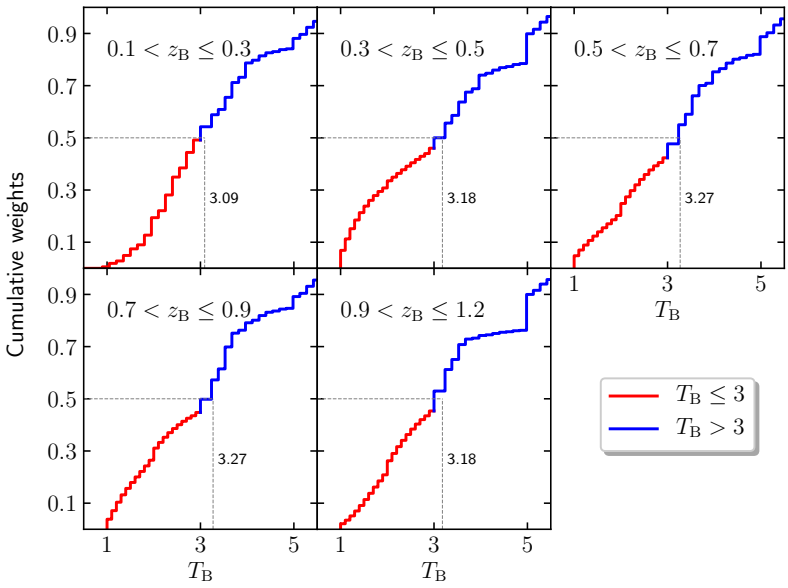

Figure 2.1: Cumulative *lens*fit-weighted distributions of T_B values. The dashed line indicates the ideal equal split in each tomographic bin, which is closely approximated by our split at $T_B = 3$.

variance in the calibration sample.

Since the spec-z catalogues cannot fully represent the photometric sample, it is necessary to weight spec-z objects to ensure an appropriate match between the spectroscopic and photometric distributions. The method, based on a kth nearest neighbour (kNN) approach, is detailed in Sect. 3 of Hildebrandt et al. (2017). In brief, it assigns weights to the spec-z objects by comparing the volume densities of the spec-z and photometric objects in the nine-band magnitude space ($ugriZYJHK_s$). As a result, KiDS+VIKING-like observations are required in the same areas as the aforementioned spec-z surveys. H20 have constructed these photometric observations from multiple sources, depending on the availability of specific data sets in those spec-z survey fields. We adopted the same sample and split it with the same criterion as used for the main KV450 sample to build two representatives of our two sub-samples.

The resulting redshift distributions of the two sub-samples are shown in Fig. 2.2. Also presented are the mean and median differences between these two redshift distributions (see Table 2.1 for separate values). The importance of photo-z calibration is demonstrated by the tails of the DIR redshift distributions compared to the ranges selected by the photo-z cuts (shaded regions). These differences between the DIR results and photo-z estimates are more significant in the red sub-sample, where an overall bias towards overestimating photo-z is shown. This may seem counter-intuitive at first, given the

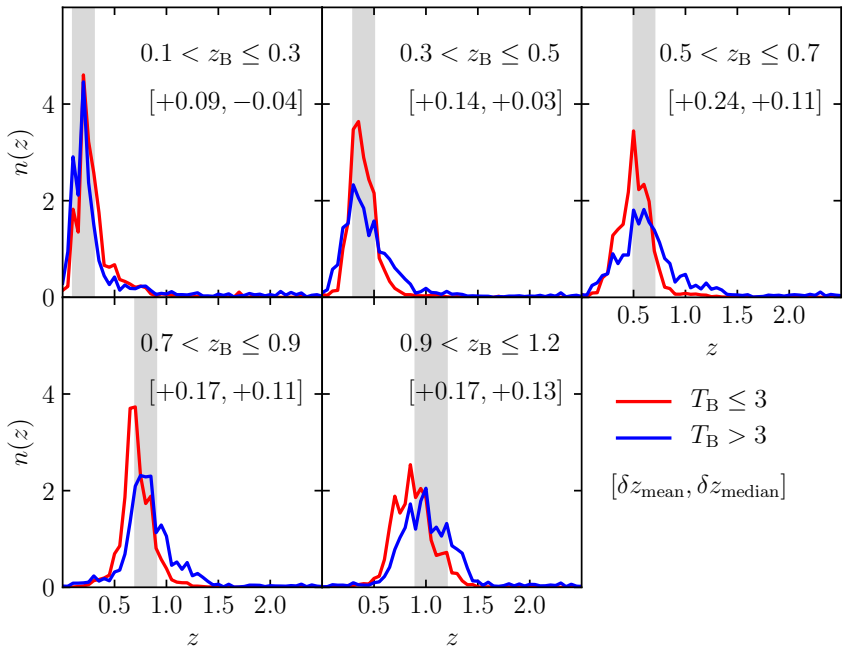

Figure 2.2: Redshift distributions for the two sub-samples, estimated using the DIR technique. Shaded regions correspond to photo-z cuts for the tomographic binning. Mean and median differences were calculated as $\delta z_{\mathrm{mean/median}} = z_{\mathrm{mean/median,blue}} - z_{\mathrm{mean/median,red}}$.

discussion presented in Sect. 2.1, which states that young stars can wash out spectral features for photo-z estimation, resulting in larger errors in bluer galaxies. However, we emphasise that the 'red' sub-sample defined in Sect. 2.2 is not purely red, but also includes Sbc and Scd types (see Sect. 2.2), which could worsen the photo-z estimates. For our purposes, we are interested in the redshift difference between the two sub-samples. As can be seen, the differences are significant, with the median differences as high as ∼0.13 and the mean differences ∼0.24 in certain bins. This level of difference will result in considerably different cosmic shear signals for the two sub-samples (see Sect. 2.5).

In practice, the DIR method is susceptible to various systematic effects, primarily induced by the incompleteness of the spec-z sample, due to selection effects and sample variance in the different spectroscopic surveys that comprise the spec-z catalogue (see Wright et al. 2020a for an updated method that is more robust against such incompleteness). To account for these potential systematic effects, H20 introduced five nuisance parameters δ_{z_i} in their model to allow for linear shifts of the redshift distributions $n_i(z) \rightarrow n_i(z + \delta_{z_i})$ (see Table 2.2). Priors for these parameters are obtained using a spatial bootstrapping approach. In our consistency tests described below, we

focus on an extension of these nuisance parameters to the colour-split sub-samples (see Sect. 2.6).

2.4 Calibration of shape measurements

The shape measurements are susceptible to various biases due to the noise of galaxy images, the complexity of galaxy shapes, selection effects, and so on (see Sect. 2 of K19 for a theoretical discussion). The weak lensing community has conducted several blind challenges to test the performance of shape measurement pipelines (see, e.g. Heymans et al. 2006; Massey et al. 2007; Bridle et al. 2010; Kitching et al. 2012; Mandelbaum et al. 2015). These tests, based on simplified image simulations, are useful for understanding common sources of shear bias but cannot eliminate biases in a specific survey. In particular, differences in selection criteria between surveys affect the shear bias. These residual biases need to be calibrated with dedicated, tailor-made image simulations (see, e.g. Hoekstra et al. 2015). Following Heymans et al. (2006), we quantify these residual biases using a linear parameterisation:

$$g_i^{\rm obs} = (1 + m_i) g_i^{\rm true} + c_i \ , \tag{2.1}$$

where $g_i^{\rm obs}$ and $g_i^{\rm true}$ are the observed and true gravitational shears, respectively, with $i = 1, 2$ referring to the two different components. In practice, we found isotropy in the m results, meaning that $m_1 \approx m_2$, so we simply adopt $m = (m_1 + m_2)/2$.

The two types of biases, m (the multiplicative bias) and c (the additive bias or c-term), have distinct sources and properties. The former is typically determined from image simulations, while the latter can be inferred directly from the data. As demonstrated in K19, shear biases depend not only on the selection function but also on the overall population of galaxies. Consequently, shear calibrations should be performed separately for samples containing different galaxy populations. This was the case for the various tomographic bins in the KV450 analysis and is even more relevant for our split analysis.

We therefore re-estimated the multiplicative biases in the two sub-samples using the COllege simulations (COSMOS-like lensing emulation of ground experiments, K19), which were also employed in the fiducial KV450 cosmic shear analysis. The main features of the COllege simulations are the observation-based input catalogue and the assignment of photometric redshifts. The input catalogue contains information on galaxy morphology and position from *Hubble* Space Telescope observations (Griffith et al. 2012) of the COSMOS field (Scoville et al. 2007). The photometric redshifts of simulated galaxies are assigned by cross-matching the input catalogue to the KiDS catalogue. This setup ensures a high level of realism in the simulated catalogue and allows us to analyse the simulated data using the same pipelines as for the real data. K19 have demonstrated that the simulated catalogue faithfully matches the full KV450 catalogue in all crucial properties, including galaxy shapes, sizes, and positions.

As anticipated, we observed significant differences in galaxy properties between the two sub-samples. We showcase one of these comparisons in Fig. 2.3, which compares

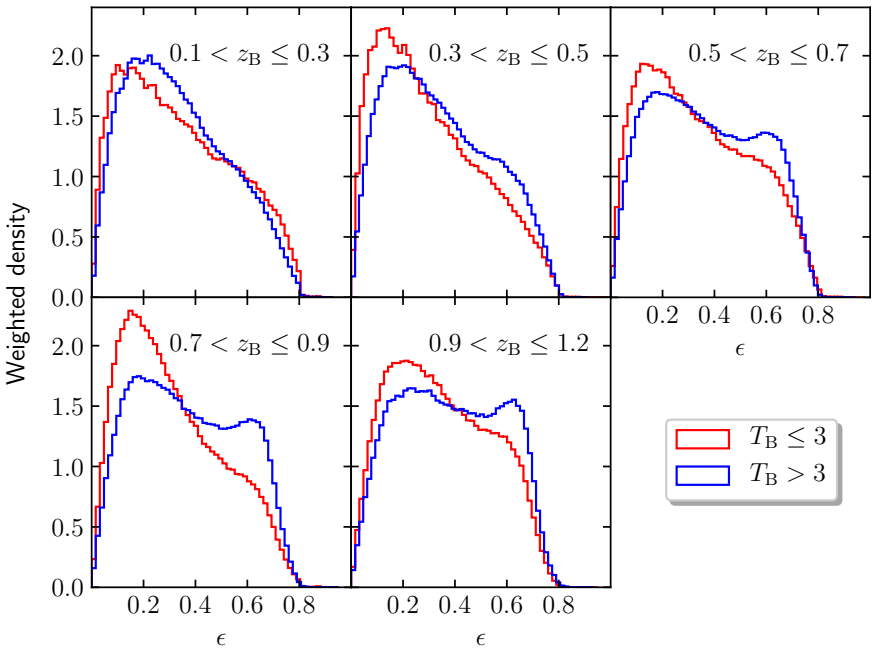

Figure 2.3: Normalised *lens*fit-weighted distributions of ellipticities for galaxies in the two sub-samples. The ellipticity is defined as $\epsilon = \sqrt{\epsilon_1^2 + \epsilon_2^2}$. We note that the differing distributions reflect distinct galaxy populations and suggest different shear biases in the two sub-samples.

the distributions of galaxy ellipticities. As previously mentioned in Sect. 2.1, ellipticity variance is one of the main sources of shape measurement biases (see also Viola et al. 2014) and thus serves as an indicator of the variance in shear biases between the two sub-samples.

Our calibration approach is identical to that used in the fiducial KV450 cosmic shear analysis. It adopts a re-weighting scheme referred to as 'Method C' in Fenech Conti et al. (2017) to account for slight differences between the observations and the simulations. The m value is reported per tomographic bin, using a weighted average of individual galaxies belonging to the corresponding tomographic bin. We direct readers to Sect. 6 of K19 for further details.

We present our estimates of multiplicative biases for the two sub-samples in Fig. 2.4, comparing them with the results from the whole sample. The five sections from top to bottom correspond to the five tomographic bins, ranging from lower to higher redshifts. We noticed some significant differences in the m values, particularly for the higher tomographic bins; these differences mainly arise from the ellipticity distributions shown in Fig. 2.3. However, when considering their impact on the cosmic shear signals, the

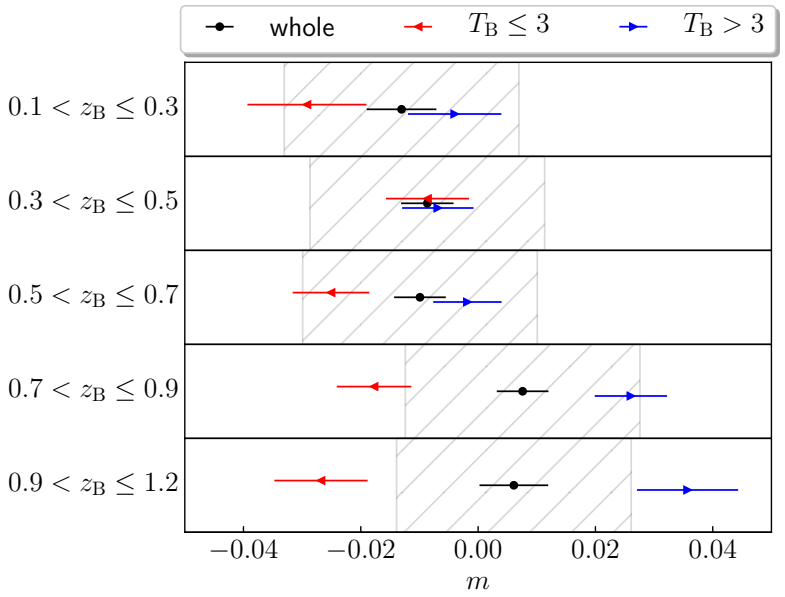

Figure 2.4: Multiplicative biases for the two sub-samples and the whole sample in each tomographic bin. Errors shown were estimated from bootstrapping. The hatched regions indicate the 0.02 error budget adopted by H20.

adjustments induced by these m-value differences are much smaller than those caused by the redshift differences, as demonstrated in Fig. 2.5. Therefore, we assume that residual systematics from the shear calibration are secondary and focus our consistency tests on the redshift calibration.

The treatment of additive bias in the fiducial KV450 cosmic shear analysis is sophisticated (see Sect. 4 of H20 for details). Briefly, the treatment can be summarised in three aspects: First, the value of c_i in each tomographic bin and in each patch is estimated by averaging over the measured galaxy ellipticities. These c_i values are then subtracted from the galaxy ellipticities before the shear correlation functions are calculated (Eq. 2.2). Second, a nuisance parameter δ_c is introduced into the model to account for a potential offset of the empirically determined c_i values. The result from forward-modelling suggests that δ_c is very close to 0 (see Table 2.2). Third, a position-dependent additive bias pattern in the ϵ_1 ellipticity component is introduced to account for an imperfection in the OmegaCAM detector chain. This pattern is publicly available as a supplementary file along with the main cosmic shear catalogues. Furthermore, another nuisance parameter A_c is introduced to allow an overall scaling of this 2D pattern (see Table 2.2).

We primarily adhered to this strategy for the additive bias calibration. We corrected the c-term per tomographic bin and per patch using the same empirical approach mentioned above. We also incorporated the 2D c-term pattern into our models. However, we excluded the two nuisance parameters δ_c and A_c from our model, as they do not have a significant impact on the fit.

2.5 Cosmic shear signal

The cosmic shear signal is encoded in the measured shapes of source galaxies as minor coherent distortions. Therefore, proper statistical measures and models are essential for a cosmic shear study. We delve into these processes in this section: Section 2.5.1 presents the construction of the joint data vector for the two sub-samples using the shear correlation functions; while Section 2.5.2 discusses the modelling of the measured data vector while taking into account various astrophysical and cosmological effects. Our analysis builds on the fiducial cosmic shear analysis of H20 but includes adjustments to accommodate our testing objectives.

2.5.1 Statistical measures

The shear signal is quantified by two-point shear correlation functions. These can be calculated from two tomographic bins i and j as follows

$$\xi_{\pm}^{ij}(\theta) = \frac{\sum_{ab} w_a w_b \left[\epsilon_t^i(\boldsymbol{x}_a)\epsilon_t^j(\boldsymbol{y}_b) \pm \epsilon_\times^i(\boldsymbol{x}_a)\epsilon_\times^j(\boldsymbol{y}_b)\right]}{(1+m^i)(1+m^j)\sum_{ab} w_a w_b}\,, \tag{2.2}$$

where $\epsilon_{t,\times}$ represent the tangential and cross ellipticities concerning the vector $\boldsymbol{x}_a - \boldsymbol{y}_b$ between a pair of galaxies (a, b), and w is the *lens*fit weight. The summation runs over all galaxy pairs within a designated spatial bin $\Delta\theta$ for each $\theta = |\boldsymbol{\theta}_b - \boldsymbol{\theta}_a|$. The multiplicative biases m^i, obtained in Sect. 2.4 for each tomographic bin i, are factored in.

We computed Eq. (2.2) separately for the two sub-samples using the publicly available TreeCorr code[2] (Jarvis et al. 2004). The spatial binning scheme is identical to the one used in H20, featuring nine logarithmically spaced bins within the $[0.''5, 300'']$ interval. We used the first seven bins for ξ_+ and the last six bins for ξ_-. These selection criteria help mitigate the impact of baryonic feedback on small scales and additive shear biases on large scales (for details, see H20). The joint data vector $(\xi_\pm^{\rm blue},\ \xi_\pm^{\rm red})$ that we created through these measurements comprises $(7+6)\times 15\times 2 = 390$ data points.

Our estimates of the data vector are presented in Fig. 2.5, with differences defined as $\Delta\xi_\pm = \xi_\pm^{\rm blue} - \xi_\pm^{\rm red}$. The accompanying error bars were derived from the analytical covariance matrix, as discussed in Sect. 2.6.1. The two sets of data vectors correspond to the results obtained with and without the multiplicative shear calibration. The discrepancy between them is minor, as anticipated given the relatively small m values (refer to Table 2.1). Some bins show non-zero trends, which can be attributed to the differing redshift distributions of the two sub-samples, as depicted in Fig. 2.2. The influence of the redshift distributions in explaining these measurements is further elaborated in the subsequent section.

[2] https://github.com/rmjarvis/TreeCorr

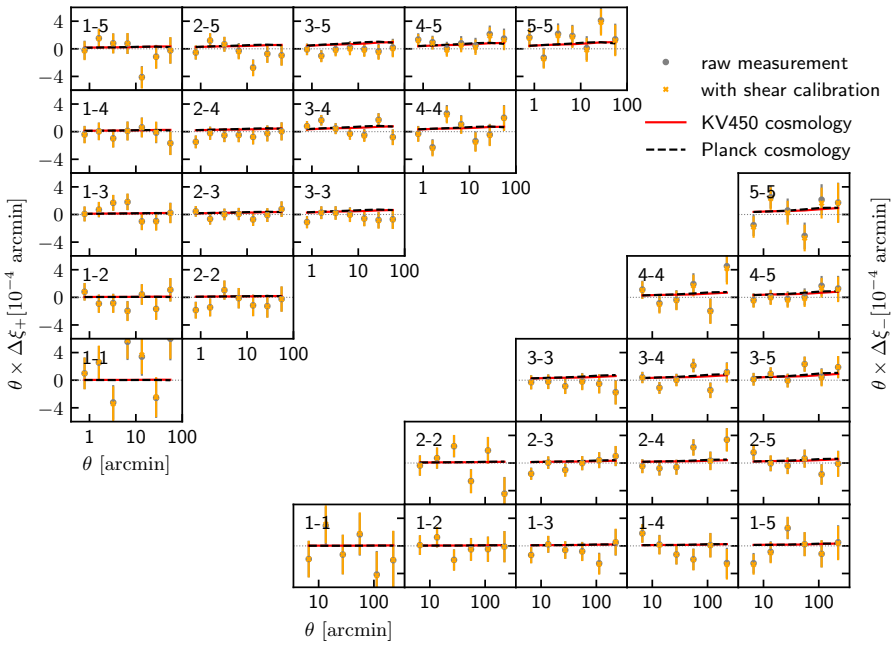

Figure 2.5: Differences between two-point shear correlation functions for the two sub-samples ($\Delta\xi_\pm = \xi_\pm^{\rm blue} - \xi_\pm^{\rm red}$). The error bars are defined as $\sigma_C = \sqrt{C_{\rm b,D} + C_{\rm r,D} - 2C_{\rm br,D}}$, where the subscript 'D' refers to the diagonal of a matrix, and the three unique parts of the complete covariance matrix are denoted as $C_{\rm b}$ for the blue sub-sample, $C_{\rm r}$ for the red sub-sample, and $C_{\rm br}$ for their cross-covariance. These errors closely match the measurement errors reported by the TreeCorr code ($\sigma_{\rm measure}/\sigma_C \gtrsim 0.8$), suggesting that the diagonal elements of the covariance matrix are predominantly influenced by measurement noise. The overall consistency between the two sets of data vectors—with and without shear calibration (represented by orange crosses and black dots, respectively)—implies that the multiplicative bias exerts a minimal effect in this study.

2.5.2 Theoretical modelling

The measured correlation functions $\xi_\pm^{ij}(\theta)$ are connected to the lensing convergence power spectrum $P_\kappa^{ij}(\ell)$ through (see e.g. Bartelmann & Schneider 2001)

$$\xi_\pm^{ij}(\theta) = \frac{1}{2\pi} \int {\rm d}\ell\, \ell P_\kappa^{ij}(\ell) J_{0/4}(\ell\theta)\,, \tag{2.3}$$

where ℓ signifies the angular wavenumber in the Fourier domain. The terms $J_{0/4}(\ell\theta)$ represent Bessel functions of the first kind, where J_0 corresponds to the zeroth-order (applied in the case of ξ_+), and J_4 denotes the fourth-order (used for ξ_-). Using the

Kaiser-Limber approximation (Limber 1953; Kaiser 1992, 1998; Loverde & Afshordi 2008), $P_\kappa^{ij}(\ell)$ is in turn related to the physical matter power spectrum P_δ, via

$$P_\kappa^{ij}(\ell) = \int_0^{\chi_\mathrm{H}} \mathrm{d}\chi \, \frac{q_i(\chi) q_j(\chi)}{[f_\mathrm{K}(\chi)]^2} P_\delta \left(\frac{\ell + 1/2}{f_\mathrm{K}(\chi)}, \chi \right) , \tag{2.4}$$

where χ represents the comoving radial distance, and $f_\mathrm{K}(\chi)$ denotes the comoving angular distance. The integration is carried out up to χ_H, which is the comoving horizon distance. The lensing efficiency $q_i(\chi)$ for a given tomographic bin i can be defined as

$$q_i(\chi) = \frac{3H_0^2\Omega_\mathrm{m}}{2c^2} \frac{f_\mathrm{K}(\chi)}{a(\chi)} \int_\chi^{\chi_\mathrm{H}} \mathrm{d}\chi' \, n_i(\chi') \frac{f_\mathrm{K}(\chi' - \chi)}{f_\mathrm{K}(\chi')} , \tag{2.5}$$

which is dependent on the redshift distribution of galaxies, represented as $n_i(\chi)\mathrm{d}\chi = n_i(z)\mathrm{d}z$, as well as various other cosmological parameters. This implies that variances in the redshift distributions between two sub-samples can lead to differences in their shear signals.

We used the Boltzmann-code CLASS (Blas et al. 2011) to compute the matter power spectrum, incorporating non-linear corrections from HMCode (Mead et al. 2016). In line with H20, we adopted a ΛCDM model comprising five primary cosmological parameters and an additional parameter representing baryonic feedback processes on smaller scales. These parameters include the densities of cold dark matter and baryons (Ω_CDM and Ω_b), the amplitude and the index of the scalar power spectrum ($\ln(10^{10}A_\mathrm{s})$, n_s), the scaled Hubble parameter (h), and the amplitude of the halo mass-concentration relation (B).

For consistency tests, it is not necessary to traverse the entire cosmological parameter space since these parameters remain the same for both sub-samples. Therefore, we set the aforementioned cosmological parameters to two distinct sets of best-fit values derived from KV450 (Hildebrandt et al. 2020) and *Planck* (Planck Collaboration et al. 2020) (see Table 2.2). This approach allows us to simplify our theoretical models while concurrently investigating potential dependencies on cosmology.

The final piece of information necessary for modelling the observed correlation functions is the intrinsic alignment (IA) of galaxies (Troxel & Ishak 2015; Joachimi et al. 2015). In the KiDS analyses, this effect is accounted for by incorporating a 'non-linear linear' IA model into the measured shear signal (Hirata & Seljak 2004; Bridle & King 2007):

$$\hat{\xi}_\pm = \xi_\pm + \xi_\pm^\mathrm{II} + \xi_\pm^\mathrm{GI} , \tag{2.6}$$

where $\hat{\xi}_\pm$ and $\xi_\pm$ represent the measured shear signal and the pure cosmic shear signal, respectively. The IA signals are incorporated in the form of $\xi_\pm^\mathrm{II}$ and $\xi_\pm^\mathrm{GI}$. The term $\xi_\pm^\mathrm{II}$ represents the 'intrinsic-intrinsic' correlation, that is, the correlation between the intrinsic ellipticities of nearby galaxies. The term $\xi_\pm^\mathrm{GI}$ stands for the 'gravitational-intrinsic' correlation, which refers to the correlation between the intrinsic ellipticity of a foreground galaxy and the shear experienced by a background galaxy. These two IA

terms can be calculated using the same formula as given in Eq. (2.3), but with the power spectra defined as

$$P_{\rm II}^{ij}(\ell) = \int_0^{\chi_{\rm H}} d\chi\ F^2(z) \frac{n_i(\chi) n_j(\chi)}{[f_{\rm K}(\chi)]^2} P_\delta\left(\frac{\ell+1/2}{f_{\rm K}(\chi)}, \chi\right) , \tag{2.7}$$

$$P_{\rm GI}^{ij}(\ell) = \int_0^{\chi_{\rm H}} d\chi\ F(z) \frac{q_i(\chi) n_j(\chi) + q_j(\chi) n_i(\chi)}{[f_{\rm K}(\chi)]^2} P_\delta\left(\frac{\ell+1/2}{f_{\rm K}(\chi)}, \chi\right) , \tag{2.8}$$

where

$$F(z) = -A_{\rm IA} C \rho_{\rm crit,0} \frac{\Omega_{\rm m}}{D_+(z)} . \tag{2.9}$$

The normalisation constant is given by $C = 5 \times 10^{-14} h^{-1} M_\odot^{-1} {\rm Mpc}^3$, and $\rho_{\rm crit,0}$ is the critical density today. The linear growth factor $D_+(z)$ is normalised to unity today. Consistent with H20, we disregarded the redshift or luminosity dependence of intrinsic alignments, assigning one nuisance parameter $A_{\rm IA}$ to account for these effects (for a more detailed treatment, see Fortuna et al. 2021a).

Equipped with all the necessary data, we can now forward-model the shear correlation functions. For illustrative purposes, we first maintained all model parameters constant, using the redshift distributions estimated in Sect. 2.3 to predict the combined data vector of the two sub-samples. The outcomes are depicted in Fig. 2.5. Two distinct predictions arise from two separate sets of cosmological parameters: the red solid line represents the KV450 best-fit values, and the black dashed line symbolises the *Planck* best-fit values. All other nuisance parameters align with the best-fit KV450 results as indicated in Table 2.2. Even with this simplified approach, the predicted results generally mirror the trends observed in the data, underscoring the redshift difference as the primary factor for the dissimilar shear correlation functions in the two sub-samples. Another notable characteristic is the resemblance between the two predictions derived from two distinct sets of cosmological parameters. This suggests that our test model is not significantly influenced by the adopted cosmological parameters. However, to accurately assess the goodness of fit and verify the robustness of the pipelines, we require a more meticulous Bayesian analysis, implementing appropriate test models and taking into account correlations between measurements.

2.6 Consistency tests

Quantifying internal consistency is no trivial task, given the correlations between measurements and the challenge of comparing different models. On the one hand, overlooking intrinsic correlations between measurements can lead to unreliable conclusions. As demonstrated by Köhlinger et al. (2019), failing to consider these correlations can confound residual systematics with the overall goodness of fit. On the other hand, null tests based on global summary statistics, such as Bayesian evidence, are practically challenging for high-dimensional models (see e.g. Trotta 2008). Furthermore, varying prior choices between hypotheses can complicate the interpretation of the final

Table 2.2: Model parameters and their best-fit values from KV450 cosmic shear analysis (Hildebrandt et al. 2020) and the *Planck* CMB analysis (Planck Collaboration et al. 2020).

Parameter	KV450	*Planck*	Definition
$\Omega_{\rm CDM}h^2$	0.058	0.120	CDM density today
$\Omega_{\rm b}h^2$	0.022	0.022	Baryon density today
$\ln(10^{10}A_{\rm s})$	4.697	3.045	Scalar spectrum amplitude
$n_{\rm s}$	1.128	0.966	Scalar spectrum index
h	0.780	0.673	Hubble parameter
B	2.189	-	Baryon feedback amplitude
$A_{\rm IA}$	0.494	-	IA amplitude
$\delta_c \times 10^5$	2.576	-	c-term offset
A_c	1.143	-	2D c-term amplitude
δ_{z_1}	−0.006	-	Bin 1 offset
δ_{z_2}	0.001	-	Bin 2 offset
δ_{z_3}	0.026	-	Bin 3 offset
δ_{z_4}	−0.002	-	Bin 4 offset
δ_{z_5}	0.003	-	Bin 5 offset

The first five parameters are the standard cosmological parameters. The remaining parameters are nuisance parameters introduced by Hildebrandt et al. (2020) to account for various effects associated with cosmic shear analysis. The KV450 best-fit values are derived from the primary Monte Carlo Markov Chain, which is publicly available at `http://kids.strw.leidenuniv.nl/cosmicshear2018.php`. The *Planck* best-fit values correspond to the TT, TE, EE+lowE+lensing results with the `Plik` likelihood (Table 1 of Planck Collaboration et al. 2020).

results (Handley & Lemos 2019b; Lemos et al. 2019).

In this section, we address these challenges. First, we construct an analytical covariance matrix to account for all the correlations between measurements (Sect. 2.6.1). Then, we conduct a Bayesian analysis using dedicated test parameters to quantify potential discrepancies between measurements from the two sub-samples (Sect. 2.6.2). Our conclusions are based on the posterior distributions of these test parameters. This approach allows us to strike a balance between model accuracy and simplicity.

The modelling pipeline described below is publicly available[3]. It is a modified

[3]`https://github.com/lshuns/montepython_KV450`

version of the MontePython package (Audren et al. 2013; Brinckmann & Lesgourgues 2018) using the PyMultiNest algorithm (Buchner et al. 2014), which is a Python wrapper of the nested sampling algorithm MultiNest (Feroz et al. 2009). The original MontePython package was used for the KV450 cosmological analysis in H20 and for the consistency tests with a split of data vector (Köhlinger et al. 2019).

2.6.1 Covariance matrix

We estimated the covariance matrix for the joint data vector assembled in Sect. 2.5.1 using the analytical model developed in Hildebrandt et al. (2017), H20, and Joachimi et al. (2021). This analytical approach represents an advancement over traditional numerical or Jackknife methods, offering benefits in managing noise effects and finite survey areas. Here, we provide a brief overview of the key aspects of this analytical approach, while detailed discussions can be found in Sect. 5 of Hildebrandt et al. (2017) and Joachimi et al. (2021).

The analytical model consists of three components: a Gaussian term related to sample variance and shape noise, a non-Gaussian term originating from in-survey modes, and a third term, also non-Gaussian, derived from super-survey modes (known as super-sample covariance, or SSC). The Gaussian term is calculated following Joachimi et al. (2008), with a transfer function taken from Eisenstein & Hu (1998) and non-linear corrections from Takahashi et al. (2012). The information about the sources used is detailed in Table 2.1; it includes the effective galaxy number density ($n_{\rm eff}$) and the weighted ellipticity dispersion ($\sigma_{\epsilon,i}$). The second non-Gaussian term is derived using the formalism proposed by Takada & Hu (2013), incorporating the halo mass function and halo bias from Tinker et al. (2010). The halo profile is described using a Fourier-transform version of the NFW model (Navarro et al. 1996; Scoccimarro et al. 2001), with the concentration-mass relation taken from Duffy et al. (2008). The final SSC term is modelled once again using the formalism from Takada & Hu (2013). The survey footprint is modelled with a HEALPix map (Górski et al. 2005).

The shear calibration outlined in Sect. 2.4 also entails uncertainties. Following the approach used in H20 and Wright et al. (2020b), we considered a systematic uncertainty $\sigma_m = 0.02$ for the multiplicative biases as estimated by K19. This uncertainty is propagated into the covariance matrix using the relation $C_{ij}^{\rm cal} = 4\xi_i^{\rm T}\xi_j^{\rm T}\sigma^2 m + C_{ij}$, where $\xi^{\rm T}$ is the joint data vector predicted with the KV450 best-fit values and the DIR redshift distributions (refer to Sect. 2.3 for more details). The error associated with the additive biases was neglected due to its minimal impact. For a more detailed discussion on this topic, please refer to Appendix D4 of Hildebrandt et al. (2017).

The final correlation matrix for the joint data vector is presented in Fig. 2.6. There are noticeable contributions from the off-diagonal regions, which highlight the significant correlations within each individual sub-sample as well as between the two sub-samples. The importance of potential correlations between parts of a split was previously emphasised in Köhlinger et al. (2019), but here we provide a more direct confirmation. By incorporating the complete covariance matrix into our consistency

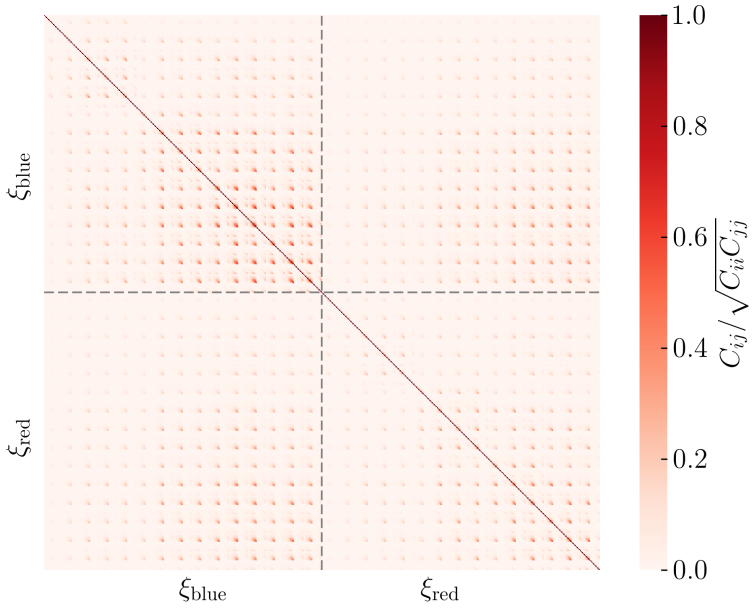

Figure 2.6: Analytical correlation matrix for the joint data vector. To highlight the correlations, the covariance C_{ij} is normalised using the diagonal $\sqrt{C_{ii}C_{jj}}$.

tests, we automatically account for all the data correlations.

We examined the relative contributions of the Gaussian and non-Gaussian terms to the complete covariance matrix. Generally, the Gaussian term dominates the diagonal parts of the covariance matrix, while the non-Gaussian term contributes more significantly to the off-diagonal regions. This pattern is also clearly outlined in Joachimi et al. (2021).

Given that our test model is especially sensitive to the difference $\Delta\xi$ between the two sub-samples, we built the covariance matrix of $\Delta\xi$ as $C_\Delta = C_{\rm blue} + C_{\rm red} - 2C_{\rm cross}$. We then compared this to the covariance matrices of the individual data vectors ($\xi_{\rm blue}$ or $\xi_{\rm red}$). Our findings indicate that the non-Gaussian contributions are significantly diminished in C_Δ, with an overall reduction of $\lesssim 75\%$ compared to $C_{\rm blue}$. The Gaussian contributions are also slightly reduced, primarily in the off-diagonal regions. These reductions in the covariance matrix C_Δ can be explained by the cancellation of sample variance. Hence, we confirmed that our test model is robust against uncertainties in the sample variance and alterations in the cosmological parameters.

2.6.2 Test setup

With the covariance matrix in place, we can explore the parameter space using a Bayesian analysis. Our main goal is to evaluate whether a unified set of nuisance parameters could adequately account for the residual systematics in the two sub-samples. Consequently, we opted to hold all cosmological parameters constant, under the assumption that the two sub-samples should intrinsically have identical values for these parameters.

To substantiate this assumption, we ran an additional test where we allowed cosmological parameters to vary. The results are in an agreement with our fixed-cosmology framework, revealing minimal degeneracy between cosmological parameters and test parameters. Therefore, we maintained our approach of fixing cosmological parameters, simplifying the likelihood function and preventing the unnecessary traversal of a high-dimensional parameter space.

To ensure robustness against any possible residual effects stemming from an inaccurate selection of cosmological parameters, we implemented two different settings. One setting used cosmological parameters derived from the KV450 cosmic shear analysis, while the other employed parameters from the *Planck* CMB results (see Table 2.2). This approach enabled us to factor in potential discrepancies in our choice of cosmological parameters.

Our test model, $\mathcal{H}_1$, contains six test parameters besides the nuisance parameters used in H20. These test parameters are a shift in IA amplitude, denoted as $A_{\rm IA,s}$, and shifts in redshift offsets, represented as $\delta_{z_i,\rm s}$. They are implemented in the two sub-samples as

$$X_{\rm blue/red} = X \pm X_{\rm s} \,, \tag{2.10}$$

where X stands for either the $A_{\rm IA}$ or δ_{z_i} parameters, while $X_{\rm s}$ designates the corresponding test parameters. The blue sub-sample is characterised by the positive sign, while the red sub-sample is associated with the negative sign.

While a discrepancy in the IA signal is anticipated, differences in the redshift offsets should be non-existent if the calibration pipeline is impervious to sample-related systematics. Any non-zero values for $\delta_{z_i,\rm s}$ would indicate the presence of residual systematics that are not fully encompassed by the common nuisance parameters. Therefore, our conclusions are primarily grounded on the posterior distributions of these test parameters.

For control purposes, we also established a base model, $\mathcal{H}_0$, with the same set of nuisance parameters as in H20, to model the joint data vector derived from our two sub-samples. It contains six free nuisance parameters: the amplitude of the IA signal $A_{\rm IA}$ (refer to Sect. 2.5.2) and the redshift offset δ_{z_i} for each tomographic bin i (refer to Sect. 2.3).

However, it is important to note that this assumption is stronger than what is mandated by data consistency. Given that the IA signal is dependent on the galaxy population, it is not expected to be identical across the two sub-samples.

The prior distributions for all free parameters are detailed in Table 2.3. The common nuisance parameters leverage the priors from H20. Specifically, $A_{\rm IA}$ employs a wide flat prior, while δ_{z_i} adopts Gaussian priors, the variance of which is determined using a spatial bootstrapping technique during the redshift calibration process (refer to Sect. 3.2 of H20). For the six new test parameters in the test model $\mathcal{H}_1$, we selected wide, uninformative priors. These choices, as demonstrated in Sect. 2.7, incorporate prior knowledge of redshift uncertainties into the common nuisance parameters, while simultaneously enabling a comprehensive exploration of the test parameters. We emphasise that the

main objective of our test is to ascertain whether the KV450 nuisance parameters are adequate for capturing any residual systematics.

Our testing method, which does not rely on Bayesian evidence to identify tensions, avoids the 'suspiciousness' issue associated with common model-selection methods (Lemos et al. 2019). In this regard, our approach is akin to the second tier of the Bayesian consistency tests proposed by Köhlinger et al. (2019). However, rather than duplicating the cosmological parameters and drawing conclusions from the posterior distributions of differences in these parameters, we concentrate on the nuisance parameters, particularly those tied to redshift calibration.

Another key difference lies in our approach to data splitting. We performed a colour-based split of the source galaxies, repeating measurements and calibrations for the sub-samples, whereas Köhlinger et al. (2019) based their analysis on a split of the measured correlation functions. Consequently, our method is more attuned to potential inconsistencies within the source samples, while their approach offers a more global test of residual systematics and their impact on the ultimate cosmological results. Thus, our test serves as a complementary examination of pipeline robustness alongside theirs.

2.7 Results

The primary outcomes of our consistency tests are depicted in Fig. 2.7, where we present the marginal posterior constraints of the five test parameters, $\delta_{z_i,\mathrm{s}}$, introduced in Sect. 2.6.2. Each of the five sections in the plot corresponds to one of the five tomographic bins. The two sets of results, drawn from the KV450 best-fit cosmology (represented by red lines) and the *Planck* best-fit cosmology (black lines), are in agreement, reinforcing the notion that our test model is not sensitive to the specific choice of cosmological parameters.

All of the values are consistent with zero within approximately 1.5σ, which suggests that the KV450 calibration pipelines are effectively correcting for these sample-related systematics. This finding further implies that there is no need to introduce additional nuisance parameters for the current analysis.

The two tomographic bins that exhibit slightly non-zero differences are the second bin (approximately 1.2σ) and the third bin (approximately 1.3σ). The interpretation of such a level of difference is nuanced, given the statistical power of current data. We emphasise that the $\delta_{z_i,\mathrm{s}}$ parameters we constrained here represent the shifts in the redshift offsets within the two sub-samples. These shifts are expected to be larger than the mean redshift offsets (δ_{z_i}), considering the notable redshift differences between the two sub-samples and the breadth of the DIR redshift distributions (refer to Fig. 2.2).

As demonstrated in Table 2.3, all $\delta_{z_i,\mathrm{s}}$ values are less than the width of the underlying redshift distributions and are approximately zero within the uncertainties. This observation attests to the overall precision of the DIR redshift distributions.

Table 2.3 presents the posterior results for all free parameters, as well as the best-fit χ^2 values for all models. While we do not base our conclusions strictly on the χ^2 test, due to the complex nature of Bayesian models where dimensionality is not directly

defined by the number of free parameters (see, for example, Handley & Lemos 2019a), it is still informative to compare these values. Upon considering the number of free parameters, it appears that the test model $\mathcal{H}_1$ is statistically on par with the control model $\mathcal{H}_0$. This further reinforces our initial conclusion regarding the suitability of the current nuisance parameters in accounting for any remaining systematic errors.

Figure 2.8 shows the contour plot for the test model, with a notable feature being the pronounced degeneracy between $A_{\rm IA,s}$ and $\delta_{z_i,{\rm s}}$ in the lower redshift bins (refer to Fig. 2.8). This correlation primarily contributes to the ambiguity in the test parameters. The complex relationship between the IA signal and redshift uncertainties was also observed in Wright et al. (2020b), where an updated redshift calibration of the KV450 data resulted in a negligible IA amplitude. Our observation underlines the complexity inherent in deciphering the apparent IA signal.

We performed an additional test in which we constrained $\delta_{z_i,{\rm s}} = 0$ in the test model $\mathcal{H}_1$. The result was a significantly positive $A_{\rm IA,s}$ value, suggesting $A_{\rm IA,blue} > A_{\rm IA,red}$. This contradicts established IA studies (see Joachimi et al. 2015 for a review), implying that IA parameters could potentially mask issues with redshift estimates. Therefore, we must exercise caution when interpreting the IA parameters.

To further investigate the influence of the IA parameters in our test model, we executed an additional test, $\mathcal{T}_1$, where $A_{\rm IA,s}$ was held constant at zero. This approach maximises the shifts of the redshift offsets by disregarding the IA difference in the two sub-samples. Even with this conservative estimate, the shifts are $\lesssim 2.1\sigma$ for all redshift bins, with the most significant values still appearing in the third bin (refer to Table 2.3).

2.8 Summary and discussion

We conducted an internal consistency test on the KV450 cosmic shear analysis by dividing source galaxies based on colour, yielding two statistically equivalent sub-samples comprising distinct galaxy populations (refer to Figs. 2.1, 2.2, and 2.3). Uniform measurements and calibrations were applied to these sub-samples, and we evaluated changes in the two-point correlation functions due to known variations in redshift distributions and multiplicative biases (see Fig. 2.5). With cosmological parameters fixed, we used a Bayesian analysis and specific test parameters to scrutinise the internal consistency of observational nuisance parameters, focusing on those related to redshift distributions. We noticed a degeneracy between redshift uncertainties and the inferred IA amplitude for lower redshift bins. However, we found no signs of internal inconsistency in the KV450 data, affirming the adequacy of the current approach—utilising a common set of nuisance parameters to linearly shift redshift distributions—for addressing residual systematics in redshift calibration.

The internal consistency test we introduced is resilient to uncertainties in background cosmology and cosmic variance. It can be employed in upcoming cosmic shear surveys prior to making any cosmological inference. This weak sensitivity to cosmology aligns with the 'shear-ratio' test (Jain & Taylor 2003; Schneider 2016; Unruh et al. 2019), which has been used to verify the precision of redshift distributions in current cosmic

Table 2.3: Priors and posterior results for all models.

Parameter	Prior	KV450			Planck		
		$\mathcal{H}_0$	$\mathcal{H}_1$	$\mathcal{T}_1$	$\mathcal{H}_0$	$\mathcal{H}_1$	$\mathcal{T}_1$
$A_{\rm IA}$	$[-6, 6]$	$1.442^{+0.826}_{-0.898}$	$1.049^{+0.818}_{-0.871}$	$0.976^{+0.776}_{-0.804}$	$1.741^{+0.507}_{-0.533}$	$1.358^{+0.463}_{-0.495}$	$1.340^{+0.466}_{-0.476}$
δ_{z_1}	0.000 ± 0.039	$-0.012^{+0.037}_{-0.037}$	$-0.000^{+0.035}_{-0.038}$	$0.001^{+0.035}_{-0.037}$	$-0.037^{+0.028}_{-0.036}$	$-0.008^{+0.036}_{-0.040}$	$-0.005^{+0.038}_{-0.039}$
δ_{z_2}	0.000 ± 0.023	$-0.006^{+0.019}_{-0.023}$	$-0.001^{+0.022}_{-0.021}$	$-0.000^{+0.021}_{-0.022}$	$-0.011^{+0.019}_{-0.019}$	$-0.003^{+0.020}_{-0.022}$	$-0.002^{+0.021}_{-0.019}$
δ_{z_3}	0.000 ± 0.026	$0.009^{+0.023}_{-0.022}$	$0.006^{+0.022}_{-0.023}$	$0.006^{+0.022}_{-0.026}$	$0.020^{+0.020}_{-0.018}$	$0.019^{+0.020}_{-0.021}$	$0.021^{+0.020}_{-0.020}$
δ_{z_4}	0.000 ± 0.012	$-0.002^{+0.012}_{-0.011}$	$-0.001^{+0.012}_{-0.011}$	$-0.002^{+0.012}_{-0.012}$	$0.003^{+0.011}_{-0.012}$	$0.003^{+0.012}_{-0.012}$	$0.003^{+0.012}_{-0.013}$
δ_{z_5}	0.000 ± 0.011	$0.002^{+0.011}_{-0.011}$	$0.003^{+0.012}_{-0.010}$	$0.002^{+0.011}_{-0.011}$	$0.006^{+0.012}_{-0.011}$	$0.005^{+0.011}_{-0.010}$	$0.006^{+0.011}_{-0.011}$
$A_{\rm IA,s}$	$[-6, 6]$	-	$0.571^{+1.178}_{-1.337}$	-	-	$0.536^{+0.793}_{-0.967}$	-
$\delta_{z_1,{\rm s}}$	$[-0.3, 0.3]$	-	$0.032^{+0.142}_{-0.085}$	$0.079^{+0.076}_{-0.069}$	-	$0.004^{+0.122}_{-0.098}$	$0.072^{+0.057}_{-0.066}$
$\delta_{z_2,{\rm s}}$	$[-0.3, 0.3]$	-	$0.080^{+0.087}_{-0.068}$	$0.116^{+0.048}_{-0.055}$	-	$0.039^{+0.059}_{-0.053}$	$0.069^{+0.032}_{-0.033}$
$\delta_{z_3,{\rm s}}$	$[-0.3, 0.3]$	-	$0.066^{+0.057}_{-0.051}$	$0.087^{+0.037}_{-0.041}$	-	$0.040^{+0.042}_{-0.039}$	$0.060^{+0.027}_{-0.030}$
$\delta_{z_4,{\rm s}}$	$[-0.3, 0.3]$	-	$0.002^{+0.048}_{-0.050}$	$0.014^{+0.044}_{-0.045}$	-	$0.009^{+0.039}_{-0.039}$	$0.019^{+0.037}_{-0.037}$
$\delta_{z_5,{\rm s}}$	$[-0.3, 0.3]$	-	$-0.002^{+0.053}_{-0.051}$	$0.005^{+0.051}_{-0.050}$	-	$0.008^{+0.048}_{-0.046}$	$0.015^{+0.046}_{-0.046}$
$N_{\rm data}$	-	390	390	390	390	390	390
$N_{\rm para}$	-	6	12	11	6	12	11
χ^2	-	366.8	356.4	356.1	357.5	364.5	364.4

The table shows parameters for the test model $\mathcal{H}_1$ (12 free parameters), control model $\mathcal{H}_0$ (6 common parameters), and test setting $\mathcal{T}_1$ (ignoring IA signal differences). The initial six are common nuisance parameters; the subsequent six address potential discrepancies between sub-samples (see Eq. 2.10). Bracketed priors denote top-hat ranges while values with errors indicate Gaussian distributions. Displayed results are mean posterior values with χ^2 values for maximum likelihood.

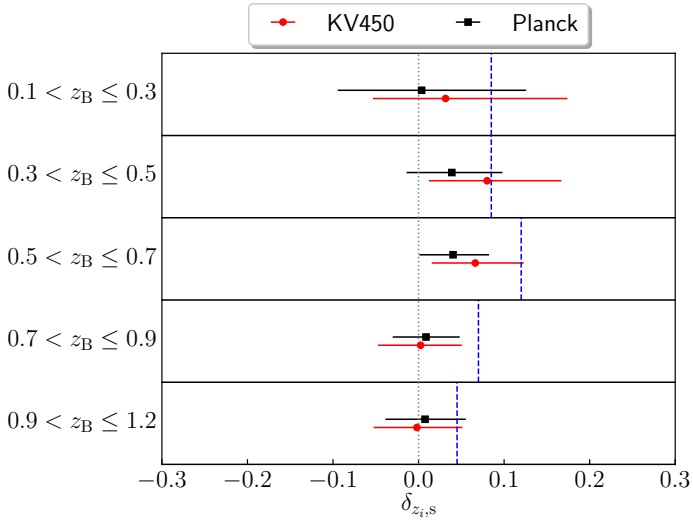

Figure 2.7: Constraints on $\delta_{z_i,\rm s}$ for each tomographic bin under the $\mathcal{H}_1$ model. The errors depicted represent the 68% credible intervals derived from the MCMC analysis. For contextual comparison, the vertical blue lines indicate half of the mean differences between the reconstructed DIR redshift distributions of the two sub-samples (as seen in Fig. 2.2).

shear surveys (Heymans et al. 2012; H20; Giblin et al. 2021). The 'shear-ratio' test, a cross-correlation method, relies on the galaxy-galaxy lensing signals from two or more source samples at varying redshift bins. Consequently, these two tests are sensitive to different systematics, making them complementary tools.

While our discussion primarily focused on redshift calibration, we discovered that the test is also contingent on our assumptions about IA signals (refer to Fig. 2.8). Without an extensive exploration of IA models, our test is already capable of identifying the degeneracy between the IA signals and redshift uncertainties, a concept previously hinted at in other studies(see Sect. 6.6 of Hildebrandt et al. 2017). Recently, Samuroff et al. (2019) conducted an analysis similar to ours, but based on DES data. Their focus was the IA signal and cosmological parameters, while they marginalised over observational nuisance parameters. While this differs from our approach, it links to our test via the IA signals, which both tests examined. They achieved more precise constraints on the IA signals in sub-samples by employing a range of IA models. We can similarly enhance our test model to glean deeper insights into IA signals and their correlation with other nuisance parameters in future cosmic shear data.

Acknowledgements

We thank Shahab Joudaki for carefully reading the manuscript and providing useful comments. SSL is supported by NOVA, the Netherlands Research School for Astronomy.

KK acknowledges support by the Alexander von Humboldt Foundation. HHo acknowledges support from Vici grant 639.043.512, financed by the Netherlands Organisation for Scientific Research (NWO). HHi is supported by a Heisenberg grant of the Deutsche Forschungsgemeinschaft (Hi 1495/5-1) as well as an ERC Consolidator Grant (No. 770935). This work is based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs 100.A-0613, 102.A-0047, 179.A-2004, 177.A-3016, 177.A-3017, 177.A-3018, 298.A-5015, and on data products produced by the KiDS consortium.

Author Contributions: All authors contributed to the development and writing of this paper. The authorship list is given in two groups: the lead authors (SSL, KK, HHo), followed by an alphabetical group of key contributors to both the scientific analysis and the data products.

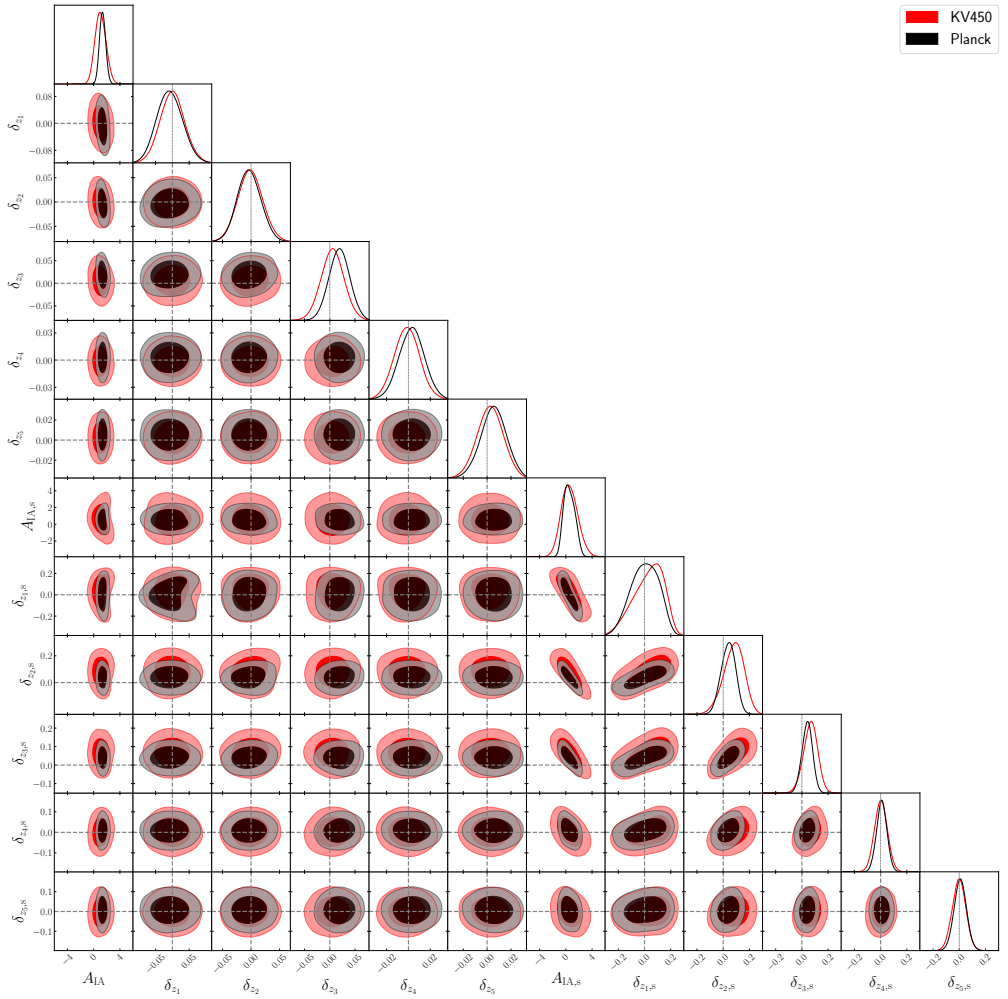

Figure 2.8: Contour plots outlining the 68% and 95% credible regions for all free parameters within the $\mathcal{H}_1$ model. The plotting ranges align with the prescribed prior ranges. Dashed lines demarcate zero values, which would represent the ideal case. The two different colours correlate with the two sets of results derived from the KV450 and *Planck* cosmological values, respectively. The faint degeneracy observed between δz_i, s in the lower redshift bins is an artefact of the substantial degeneracy between $A_{\rm IA,s}$ and δ_{z_i},s. This degeneracy dissipates in the $\mathcal{T}_1$ test setting, wherein $A_{\rm IA,s}$ is held constant at zero.