Constraining gravity with a new precision E_G estimator using Planck + SDSS BOSS data

Lukas Wenzl,[1,*] Rachel Bean,[1] Shi-Fan Chen,[2] Gerrit S. Farren,[3,4] Mathew S. Madhavacheril,[5] Gabriela A. Marques,[6,7] Frank J. Qu,[3,4] Neelima Sehgal,[8] Blake D. Sherwin,[3,4] and Alexander van Engelen[9]

[1]*Department of Astronomy, Cornell University, Ithaca, New York 14853, USA*
[2]*School of Natural Sciences, Institute for Advanced Study, 1 Einstein Drive, Princeton New Jersey 08540, USA*
[3]*DAMTP, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 OWA, United Kingdom*
[4]*Kavli Institute for Cosmology Cambridge, Madingley Road, Cambridge CB3 0HA, United Kingdom*
[5]*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA*
[6]*Fermi National Accelerator Laboratory, P. O. Box 500, Batavia, Illinois 60510, USA*
[7]*Kavli Institute for Cosmological Physics, University of Chicago, Chicago, Illinois 60637, USA*
[8]*Physics and Astronomy Department, Stony Brook University, Stony Brook, New York 11794, USA*
[9]*School of Earth and Space Exploration, Arizona State University, 781 Terrace Mall, Tempe, Arizona 85287, USA*

(Received 23 January 2024; accepted 29 March 2024; published 30 April 2024)

The E_G statistic is a discriminating probe of gravity developed to test the prediction of general relativity (GR) for the relation between gravitational potential and clustering on the largest scales in the observable Universe. We present a novel high-precision estimator for the E_G statistic using CMB lensing and galaxy clustering correlations that carefully matches the effective redshifts across the different measurement components to minimize corrections. A suite of detailed tests is performed to characterize the estimator's accuracy, its sensitivity to assumptions and analysis choices, and the non-Gaussianity of the estimator's uncertainty is characterized. After finalization of the estimator, it is applied to *Planck* CMB lensing and SDSS CMASS and LOWZ galaxy data. We report the first harmonic space measurement of E_G using the LOWZ sample and CMB lensing and also updated constraints using the final CMASS sample and the latest *Planck* CMB lensing map. We find $\hat{E}_G^{\mathrm{Planck+CMASS}} = 0.36^{+0.06}_{-0.05}(68.27\%)$ and $\hat{E}_G^{\mathrm{Planck+LOWZ}} = 0.40^{+0.11}_{-0.09}(68.27\%)$, with additional subdominant systematic error budget estimates of 2% and 3%, respectively. Using $\Omega_{\mathrm{m},0}$ constraints from *Planck* and SDSS BAO observations, ΛCDM-GR predicts $E_G^{\mathrm{GR}}(z = 0.555) = 0.401 \pm 0.005$ and $E_G^{\mathrm{GR}}(z = 0.316) = 0.452 \pm 0.005$ at the effective redshifts of the CMASS and LOWZ based measurements. We report the measurement to be in good statistical agreement with the ΛCDM-GR prediction and report that the measurement is also consistent with the more general GR prediction of scale independence for E_G. This work provides a carefully constructed and calibrated statistic with which E_G measurements can be confidently and accurately obtained with upcoming survey data.

DOI: 10.1103/PhysRevD.109.083540

I. INTRODUCTION

Twenty-five years ago, the late-time accelerated expansion of the Universe was discovered [1,2], implying that the Universe is currently dominated by a component distinct from the clustering dark matter and Standard Model particles, coined "dark energy" that is driving this expansion. To date, the physical reason for dark energy remains unresolved [3]. Current observational constraints on the expansion history of the Universe can be well-explained by a cosmological constant Λ. Together with cold dark matter (CDM) and baryonic matter this ΛCDM model forms the current standard model of cosmology due to its simplicity and also its ability to fit a wide range of cosmological data (e.g. [4–7]). This however is not the only theoretical model of gravity that can fit the data. While simple in its inherently constant energy density, the fine-tuning and coincidence problems related to explaining the observed magnitude of Λ [8] have led to the consideration of other gravitational models. These include, for example, $f(R)$ gravity [9], Chameleon gravity [10], DGP [11], TeVeS [12] and others (see Clifton *et al.* [13] for an in-depth overview). These models generally contain screening mechanisms that hide their effects in high-density environments allowing them to pass gravity tests on solar system scales and they can be fine-tuned to match the expansion history of the Universe including the late-time accelerated expansion [14]. This degeneracy makes these modified gravity theories difficult to distinguish from distance measurements alone.

*ljw232@cornell.edu

 © 2024 American Physical Society

However, these alternative theories of gravity also have implications for the growth of large-scale structures (LSS) in the Universe differing from what is expected for general relativity. In particular, the Poisson equation which relates the gravitational potentials to overdensities is altered, and the potential experienced by relativistic particles like photons can be different from the potential experienced by nonrelativistic particles [15]. Similar to how comparing lensing around the sun with the gravitational dynamics of the planets can be used to test general relativity (GR) on solar system scales, one can devise tests on the scale of a significant fraction of the observable Universe to constrain gravity further [16]. A wide range of tests have been proposed that leverage the distinct growth of structure in modified gravity models to differentiate them from GR generally or ΛCDM specifically (see [14,17–19] for overviews).

One promising pathway to constrain gravity in this regime has been proposed in Zhang *et al.* [20]: the so-called E_G statistic, that combines a measurement of the divergence of the peculiar velocity field θ with a measurement of gravitational lensing $\nabla^2(\psi - \phi)$. A comparison between these two measurements is sensitive to changes in the Poisson equations and differences between the two gravitational potentials enabling it to differentiate between GR and modified gravity models with degenerate expansion histories [21,22]. The velocity field is measured by exploiting matter conservation so that on linear scales it can be found by measuring the linear growth rate f and matter overdensities δ. The underlying matter field dominated by dark matter cannot be observed directly so one generally instead leverages biased tracers like galaxies that are related to the underlying matter field by a bias factor b_g. The E_G statistic is specifically chosen as the ratio of two measurements that have the same dependence on the galaxy bias leading to a cancellation on linear scales. Spectroscopic redshift information is ideal to accurately determine the divergence of the peculiar velocity field, given it measures the 3D clustering of galaxies. An analogous quantity for photometric survey data is the D_G statistic, introduced by [23] and applied in several studies, e.g. [24–26]. D_G shares a conceptual resemblance to E_G, since it also combines lensing and clustering information to remove the dependence on galaxy bias. However, E_G has the advantage of being scale-independent in GR and is more directly related to deviations in the Poisson equation.

The initial proposal of the E_G statistic suggested the use of weak lensing measurements of background galaxies as the tracer of lensing [20]. This was used for the first measurement of E_G presented in Reyes *et al.* [27] and has been done in a range of further analyses [28–33]. The authors of Pullen *et al.* [21] proposed to use cosmic microwave background (CMB) lensing measurements as the lensing tracer allowing constraints on larger scales. Their proposed estimator (herein referred to as $\hat{E}_G^{\mathrm{Pullen}}$) has been applied to CMB lensing measurements from *Planck* and spectroscopic galaxy samples from the Sloan Digital Sky Survey (SDSS) [32,34,35].

Upcoming next-generation observatories offer the opportunity to significantly tighten constraints on the E_G statistic. These include for CMB lensing the Atacama Cosmology Telescope (ACT) [36], the South Pole Telescope (SPT) [37,38], Simons Observatory (SO) [39] and CMB-S4 [40] and upcoming spectroscopic galaxy samples from the Dark Energy Spectroscopic Instrument (DESI) [41] and the Spectro-Photometer for the history of the Universe, Epoch of Reionization, and Ices Explorer (SPHEREx) [42].

However, to fully realize the potential of larger statistical constraining power with upcoming datasets careful approaches are required. Precision estimators for which systematic errors are well within the smaller statistical uncertainties are needed. There are significant analysis challenges that need to be carefully modeled. In this work, we aim to prepare for upcoming data, creating a precision pipeline that allows accurate estimation of the E_G statistic using CMB lensing and spectroscopic galaxy data. We present a newly revised estimator for the E_G statistic building on previous work and making key advances.

We apply this new estimator for the E_G statistic to well-tested datasets, namely CMB lensing from *Planck* and galaxy clustering from SDSS BOSS, and report new constraints. The data used has been previously extensively tested for systematics (see e.g. [32,34,43,44]). We do not reproduce these tests on the data here but note that such tests will also be a key part of future work with new upcoming datasets, to carefully characterize the datasets prior to any E_G analysis to ensure constraints can be confidently attributed to cosmological, and not instrumental or astrophysical, effects.

The structure of this paper is as follows. In Sec. II we summarize the key results from the paper, the definition and motivation of the E_G statistic are discussed in Sec. III, and the datasets used for the analysis are described in Sec. IV. Then, in Sec. V, the new and updated E_G estimator applicable to precision CMB lensing and spectroscopic galaxy clustering data is motivated and presented. In Sec. VI the measurement of the observables from data is discussed. In Sec. VII the non-Gaussian uncertainty of the E_G estimator is investigated and statistical tests to compare with GR predictions are developed. The results of the measurements are presented in Sec. VIII and the results are discussed in Sec. IX.

II. SUMMARY OF KEY RESULTS

In this work, we introduce a novel precision estimator for the E_G statistic to test a key prediction of GR about the structure formation of the Universe that is distinct from predictions for alternative gravity models. We apply this new estimator to well-established data in the form of CMB

lensing measurements from *Planck* PR4 [45] and galaxy clustering measurements from SDSS BOSS DR12 [46], to present new constraints on the E_G statistic. GR predictions used in this work are based on the ΛCDM constraints given in the Planck 2018 results [5] using CMB anisotropy and baryon acoustic oscillations (BAO) from SDSS BOSS DR12 [4].

The presented estimator for the E_G statistics uses CMB lensing and spectroscopic galaxy clustering information. Key to the derivation is to establish the effective redshift of the observables used for the measurement and the introduction of a reweighting scheme that allows matching the effective redshift of the angular cross-power spectrum to the angular autopower spectrum and the redshift space distortions (RSD) analysis which are all combined for the estimator. Additionally, we minimize the approximations during the derivation, resulting in a precise estimator that is applicable to upcoming precision data and mainly limited by astrophysical systematics. The impact of astrophysical systematics can be minimized in upcoming precision data by calculating the E_G statistic in narrower bins in redshift.

The estimator, all applied corrections, and analysis choices were finalized and frozen before applying the estimator to the *Planck* and SDSS BOSS data. There were no major changes to the result after the initial calculation of the $\hat{E}_G$ and $\hat{E}_G^\ell$ estimates. We briefly summarize key insights from the measurements:

(i) GR predicts the E_G statistic to be scale-independent on linear scales while in many alternative theories of gravity, E_G is scale dependent. We find our observations to be statistically consistent with scale independence, in line with GR predictions. Our tests are based on constraints from two measurements with effective redshifts of $z_{\rm eff}^{\rm Planck+CMASS} = 0.555$ and $z_{\rm eff}^{\rm Planck+LOWZ} = 0.316$ sensitive to scales in the approximate range $k \in [0.02, 0.2]$ Mpc^{-1}. The binned measurements as a function of scale are shown in the top panels of Fig. 1.

(ii) Since our data is consistent with scale-independence we combine the constraints across the angular scales assuming E_G as a constant. We find overall constraints on E_G of $\hat{E}_G^{\rm Planck+CMASS} = 0.36^{+0.06}_{-0.05}({\rm stat})$ and $\hat{E}_G^{\rm Planck+LOWZ} = 0.40^{+0.11}_{-0.09}({\rm stat})$ which are visualized in the bottom panel of Fig. 1.

(iii) Within the framework of GR the well-determined expansion history of the Universe makes direct predictions for the growth history and thereby E_G. We compare our measurement to predictions for E_G based on a ΛCDM fit, specifically a measurement of $\Omega_{\rm m,0}$ from [5] using *Planck* CMB anisotropy and SDSS BOSS BAO measurements. We find that each of our measurements *Planck* + CMASS and *Planck* + LOWZ is statistically consistent with the respective ΛCDM-GR prediction. In addition to a consistency test of the data with GR the presented results can also be compared to predictions for alternative models of gravity, enabling constraints on their degrees of freedom.

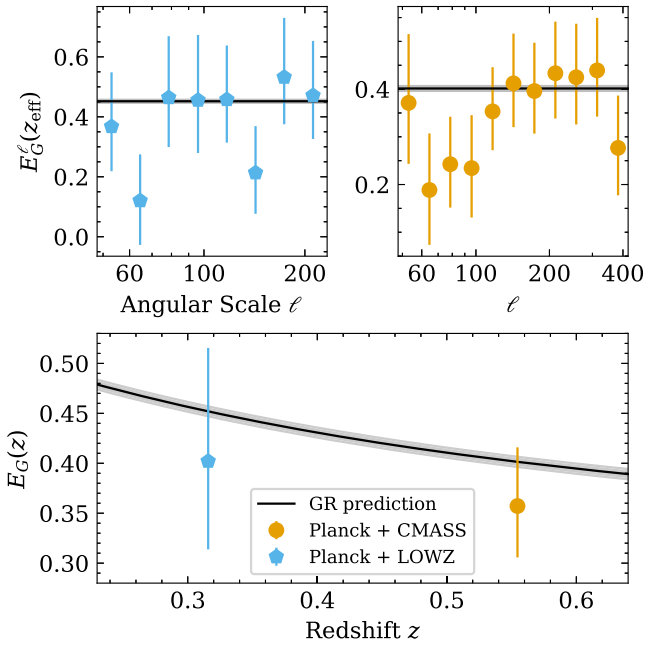

FIG. 1. Summary of the E_G statistic measurements presented in this work. Shown is E_G as a function of redshift (lower) and for each of the two measurements at their effective redshift as a function of angular scale ℓ (upper). We present a measurement using *Planck* CMB lensing combined with SDSS CMASS galaxies (orange circles) and combined with SDSS LOWZ galaxies (blue pentagons) with 68.27% confidence ranges for the error. The ΛCDM-GR expectation for the fiducial cosmology (assuming $\Omega_{\rm m,0} = 0.3111 \pm 0.0056$) is shown as a black line with a gray one sigma uncertainty band.

III. THE E_G STATISTIC

Modified gravity models, such as $f(R)$ gravity [9] and Chameleon gravity [10], offer an alternative framework to understand the observed accelerated expansion of the Universe. However, they typically have implications for the formation of cosmic structures that differ from GR.

For GR, the Einstein field equations are solved, under the assumption of isotropy and homogeneity on large scales, with the Friedmann-Lemaître-Robertson-Walker (FLRW) metric and the Friedmann equations. The dynamics of such a universe are described as perturbation fields in this background metric; ψ for the time component and ϕ for the spatial component. The perturbed FRW metric for a flat universe is given by $ds^2 = (1+2\psi)dt^2 - a^2(1+2\phi)d\mathbf{x}^2$ where $a(t)$ is the scale factor. With this metric, the dynamics of the universe are described by [47]

$$\nabla^2\psi = -4\pi G a^2 \rho(a)\delta, \tag{1}$$

$$\phi = -\psi, \tag{2}$$

where ρ is the background matter density and δ the matter density pertubations. In GR the perturbation fields of time and space are equal (up to a sign convention), giving the generalized Poisson equation relating the lensing of light $\nabla^2\phi$ and the fractional overdensity δ. This represents a prediction of GR about how gravitational lensing relates to the clustering in our Universe.

This relation changes for many alternative models of gravity in which the expansion history $a(t)$ is tuned to match GR well within observational uncertainties, however, the dynamics of the universe change. In general, for a large class of models, the modified Poisson equations can be described as [21]

$$\nabla^2\psi = -4\pi G a^2 \mu(k,a)\rho(a)\delta, \tag{3}$$

$$\phi = -\gamma(k,a)\psi. \tag{4}$$

Here, γ and μ are general functions that can be time and scale-dependent for a given modified gravity scenario. The μ function parametrizes the effective strength of gravity, and γ is the gravitational slip that quantifies the difference in the perturbation fields. We can also use the common reparametrization of γ to $\Sigma \equiv \frac{1}{2}\mu(1+\gamma)$, the lensing parameter, which is what the E_G statistic, defined below, more directly measures. In the GR limit μ, γ and Σ are 1. A large class of modified gravity scenarios can be mapped to these general functions [21,48–50].

The E_G statistic is constructed as a quantity to test the relation between lensing and matter clustering as predicted by the Poisson equations. For observations, the clustering is difficult to measure directly since the observable galaxies are a biased tracer of the underlying matter overdensities. The E_G statistic instead investigates the divergence of the peculiar velocity field θ which, on linear scales, is related to the underlying matter perturbations as $\theta = -f\delta$ where $f = d\ln\delta/d\ln a$ is the growth rate. The definition of the E_G statistic in harmonic space, as introduced in Zhang *et al.* [20], is given by

$$E_G(k,z) \equiv \left[\frac{\nabla^2(\psi-\phi)}{3H_0^2(1+z)\theta}\right]_k. \tag{5}$$

The definition is chosen so that one can build an estimator from power spectra involving galaxy tracers where the effect of galaxy bias cancels on linear scales and therefore the measured E_G statistic can conveniently be compared to the expectation value for GR. We discuss this in detail later in this section when discussing the 3D estimator.

By plugging in the Poisson equation and using $\theta = -f\delta$ we can find the expectation value for GR to be

$$E_G^{\mathrm{GR}}(z) = \frac{\Omega_{\mathrm{m},0}}{f(z)}, \tag{6}$$

where $\Omega_{\mathrm{m},0} = \frac{\rho}{\rho_{\mathrm{crit}}}$ is the energy density in the form of matter today, in units of the critical density $\rho_{\mathrm{crit}} = \frac{3H_0^2}{8\pi G}$. The value for E_G is predicted to be scale-independent for GR in the linear regime. More specifically we can quantify this for the current standard model of cosmology, ΛCDM, which represents a model of general relativity where the dominant components of the Universe today are a cosmological constant Λ, CDM, baryonic matter, and no spacial curvature. Within the ΛCDM model the growth rate can be well-approximated by $f(z) = \Omega_{\mathrm{m}}(z)^{0.55}$ [51] and therefore E_G is predicted from the background expansion using a constraint on $\Omega_{\mathrm{m},0}$ alone. Based on measurements of the expansion history and early Universe in the form of BAO and CMB anisotropy observations, the 6-parameter ΛCDM model is tightly constrained. In particular, for our fiducial cosmology based on *Planck* CMB anisotropies and SDSS BAO [5] we have $\Omega_{\mathrm{m},0} = 0.3111 \pm 0.0056$. This gives us a percent-level ΛCDM-GR prediction for the E_G statistic based on geometric measurements of the expansion history of the Universe against which we can compare the growth of large-scale structure-derived E_G estimates at the effective redshifts of the surveys used in this work.

Furthermore, for a given alternative gravity scenario with modifications γ and Σ, the expression for E_G is [21]

$$E_G^{\mathrm{ModGrav}}(k,z) = \frac{\Omega_{\mathrm{m},0}\Sigma(k,z)}{f(k,z)}, \tag{7}$$

where the growth rate $f(k,z)$ and $E_G(k,z)$ can now be scale dependent. The E_G statistic thus directly scales with modifications to the gravitational lensing potential quantified by Σ and is also sensitive to changes in the growth rate. Results on E_G can be compared to a given model of modified gravity by deriving the expected E_G value based on the modified Poisson equations of the model. We refer the interested reader to Pullen *et al.* [21] where E_G is derived for the example cases of $f(R)$ gravity and Chameleon gravity.

Based on the definition in Eq. (5), an effective 3D estimator with expectation value equal to E_G is given by [20,21]

$$\hat{E}_G(k,z) = \frac{c^2\hat{P}_{\nabla^2(\psi-\phi)g}(k,z)}{3H_0^2(1+z)\hat{P}_{\theta g}(k,z)}, \tag{8}$$

where H_0 is the Hubble constant and the hats indicate estimation from observables. The $\hat{P}_{\theta g}(k,z)$ is the cross-power spectrum between the divergence of peculiar velocities and galaxy clustering. On linear scales, the divergence of the peculiar velocities relates to overdensities, δ_g, of a galaxy tracer, as $\theta = -\beta\delta_g$ with $\beta = f/b_g$, where b_g is the bias of the tracer relative to the underlying matter field. Therefore we can approximate the correlation in the denominator as $\beta\hat{P}_{gg}(k,z)$ where we have the galaxy power

spectrum as well as β which can be constrained from an RSD measurement. In total the effect of galaxy bias cancels on linear scales for this estimator since $\beta \propto 1/b_g$, $\hat{P}_{gg} \propto b_g^2$ gives $\hat{P}_{\theta g} \propto b_g$ and we also have $\hat{P}_{\nabla^2(\psi-\phi)g} \propto b_g$. This allows a direct comparison of the measurement to the theory prediction.

The $\hat{P}_{\nabla^2(\psi-\phi)g}$ is the 3D power spectrum between lensing and galaxy clustering. While the E_G statistic is defined in 3D, the lensing information needed for the estimate is only available as a projected quantity on the sky. Therefore estimators of the E_G statistic are generally built for the projected statistics [20,21,27].

When combining projected quantities, as is explicit in the E_G statistic, one has to consider carefully the effective redshift of each observed angular power spectrum for the datasets employed to ensure a consistent apples-to-apples comparison.

IV. DATA

A. *Planck* CMB lensing data

We use CMB lensing measurements based on CMB observations with the *Planck* Satellite.[1] The official *Planck* 2018 public release 3 (PR3) was presented in [52] and the corresponding analysis of CMB lensing was presented in [53]. For this work, we use the latest reprocessing of the *Planck* data using NPIPE, referred to as PR4 [54]. PR4 contains approximately 8% more data and represents an improved processing pipeline compared to PR3, resulting in smaller errors overall. The *Planck* PR4 CMB power spectra give statistically consistent constraints for ΛCDM compared to *Planck* PR3 [55,56].

We use the CMB lensing map reconstruction based on the *Planck* PR4 data products presented in Carron *et al.* [45].[2] A quadratic estimator with additional optimal filtering compared to the PR3 lensing map which in combination with the additional CMB data, results in approximately 20% stronger constraints on the amplitude of the power spectrum. The amplitude of this *Planck* PR4 lensing map is statistically consistent with *Planck* PR3 lensing measurements, with the relative amplitude reported as 1.004 ± 0.024 (68% limits) [45]. The *Planck* CMB lensing maps have been extensively tested for systematic effects like foreground contamination in the context of cross-correlation analyses (e.g. [34,44,57]).

The *Planck* PR4 lensing map is estimated for the scales $8 \leq \ell \leq 2048$. We apply a low-pass filter $\exp(-(\ell/\ell_{\rm max})^{20})$ to the data with $l_{\rm max} = 1800$ to avoid small-scale noise bleeding into our measurement, for which we only consider scales $\ell < 420$, and we also apodize the mask to avoid sharp edges as done in White *et al.* [58]. We perform all calculations in equatorial coordinates and therefore rotate the *Planck* maps from the Galactic coordinates.

In addition to the *Planck* data, we also leverage in this analysis a set of 480 CMB simulated lensing maps used in Carron *et al.* [45]. They are CMB lensing reconstructions on the CMB map simulations originally presented in Planck Collaboration *et al.* [54]. These simulated CMB maps contain a fiducial cosmology signal, realistic noise, and systematics of the analysis pipeline.

B. SDSS BOSS data

For our analysis, we use the Baryon Oscillation Spectroscopic Survey (BOSS) catalogs of galaxies with spectroscopic redshift information [59]. BOSS was part of the Sloan Digital Sky Survey III [60], it covers 10,252 square degrees of sky and contains 1,198,006 galaxies. We use the final data release 12 (DR12) which was presented in Reid *et al.* [46][3] and whose cosmological analyses are summarized in Alam *et al.* [4]. The data consists of two large-scale structure galaxy catalogs for cosmological analysis: CMASS and LOWZ. For cosmological analysis, we use the CMASS galaxy catalog in the redshift range of $0.43 < z < 0.7$ and the LOWZ galaxy catalog in the range $0.15 < z < 0.43$.

In the following, we describe in detail how we construct the overdensity maps, used for the cosmological analysis, from the galaxy catalogs. Throughout the analysis HEALPix maps with nside $= 1024$ are used.

A mask is constructed containing the fractional coverage (f_i) of the survey for each pixel i of the overdensity map. SDSS provides completeness information per sector and veto areas as mangle polygon files. In Appendix A we describe in detail how we convert these to a pixelized map. We apply a minimum cutoff coverage fraction of 0.6 to the mask to avoid pixels with low coverage. This cutoff removes 5908 galaxies (0.76%) for CMASS and 3169 galaxies (0.88%) for LOWZ. Based on the mask the sky fraction of the galaxy sample is determined as

$$f_{\rm sky,gal} \equiv \sum_i f_i / N_{\rm pix}. \tag{9}$$

We also estimate $f_{\rm sky,gal,PlanckPR4overlap}$, the sky fraction for only the pixels that overlap with the Planck PR4 CMB lensing data, by only summing the pixels where the CMB lensing mask is nonzero. After accounting for the redshift cuts and cutoff on the mask the CMASS and LOWZ samples consist of 771,294 and 358,593 galaxies, respectively, and cover fractions $f_{\rm sky,gal}^{\rm CMASS} = 0.225$ and $f_{\rm sky,gal}^{\rm LOWZ} = 0.200$ of the sky. The Planck PR4 CMB lensing map covers 96% of both the CMASS map and LOWZ maps, resulting in

[1] https://www.cosmos.esa.int/web/planck.
[2] https://github.com/carronj/planck_PR4_lensing.
[3] https://data.sdss.org/sas/dr12/boss/lss/.

$f^{\mathrm{CMASS}}_{\mathrm{sky,gal,PlanckPR4overlap}} = 0.216$ and $f^{\mathrm{LOWZ}}_{\mathrm{sky,gal,PlanckPR4overlap}} = 0.192$.

To account for systematic effects in the data, weights are applied when constructing the overdensity maps. These weights account for trends in seeing, galactic latitude, redshift failures, and close pairs. The SDSS team provides these as inverse variance weights to correct for the effects. The weight for each CMASS or LOWZ galaxy g is given by [46]

$$w_g = (w_{\mathrm{NOZ}} + w_{\mathrm{CP}} - 1) \cdot w_{\mathrm{SEEING}} \cdot w_{\mathrm{STAR}} \tag{10}$$

accounting in order for redshift failures, close pairs with fiber collisions, effects of seeing, and bright star contamination. Additionally, the Feldman-Kaiser-Peacock (FKP) weights w_{FKP} given in Reid *et al.* [46] are applied. The FKP weights are redshift dependent and therefore need to be accounted for when calculating the galaxy redshift distribution $\mathrm{d}N/\mathrm{d}z$. Two separate overdensity maps are constructed; one with the weights $w_{\mathrm{tot}} = w_g w_{\mathrm{FKP}}$ for the autocorrelation and one with weights $w_{\mathrm{tot}} = w_g w_{\mathrm{FKP}} w_{\times}$ for the cross-correlation where additional weights [Eq. (24)] are applied to match the effective redshift of the autocorrelation. These weight choices are made to match the effective redshift of the angular power spectra with the effective redshift of the RSD analysis (see Sec. V A). Including the FKP weights is optimal for the RSD analysis, reducing the error on β significantly [61,62] and therefore to match the effective redshift of the angular power spectra we also need to include the FKP weights for the galaxy overdensity maps.

The galaxy overdensity maps δ_i are constructed as

$$\delta_i = \frac{n_i}{f_i \bar{n}} - 1, \tag{11}$$

where we used weighted galaxy counts per pixel $n_i = \sum_{g \in i} w_{\mathrm{tot}}$ and the average weighted galaxy count $\bar{n} = 1/N \sum_i n_i/f_i$. N is the number of pixels with coverage above the cutoff of 0.6, and f_i is the mask.

The impact of w_g weights on E_G analyses has been investigated in [34] where it was reported that only the seeing and star weights significantly affect the results compared to the size of the statistical uncertainty. Both of these weights are well-motivated to correct observational systematics and have been shown to be necessary to recover unbiased clustering measurements [63]. The other weights ($w_{\times}$ and w_{FKP}) are only applied based on redshift, not position so cannot introduce potential contamination to the correlation but they do affect the measured values as they change the effective redshift of the measurements.

For the autocorrelation of the galaxies, shot noise needs to be accounted for. The coupled shot noise can be estimated analytically as [64,65]

$$\tilde{N}^{\mathrm{shot}} = \frac{f_{\mathrm{sky,gal}}}{n_{\mathrm{eff}}}, \tag{12}$$

$$n_{\mathrm{eff}} = \frac{(\sum_g w_{\mathrm{tot}})^2}{4\pi f_{\mathrm{sky,gal}} \sum_g w_{\mathrm{tot}}^2}, \tag{13}$$

where n_{eff} is the effective number density of galaxies.

V. NEW PRECISION E_G ESTIMATOR FOR CMB LENSING AND GALAXY CLUSTERING

A. Angular power spectra and effective redshift

In harmonic space, we can measure the angular (cross) power spectra of the observed maps. These angular power spectra are sensitive to the underlying clustering quantified by a 3D matter power spectrum $P(k, z)$. This clustering is measured as projected onto the sky where the projection is described by window functions W that characterize the sensitivity of the observed tracers to the underlying clustering as a function of distance. The observed angular power spectra are therefore generally described under the Limber approximation as

$$C_\ell^{AB} \equiv \int \mathrm{d}\chi \frac{W_A(\chi) W_B(\chi)}{\chi^2} P\left(k = \frac{\ell + 1/2}{\chi}, z(\chi)\right), \tag{14}$$

where χ is the comoving distance to redshift z, and ℓ is the multipole moment.

The angular cross-power spectrum between CMB lensing and galaxy clustering as well as the angular autopower spectrum for galaxy clustering are given by [21]

$$C_\ell^{\kappa g} = \int \mathrm{d}z \frac{\hat{W}_\kappa(z) W_g(z)}{\chi^2(z)} P_{\nabla^2(\psi-\phi)g}\left(k = \frac{\ell + 1/2}{\chi(z)}, z\right) \tag{15}$$

$$\overset{(\mathrm{GR})}{=} \int \mathrm{d}z \frac{W_\kappa(z) W_g(z)}{\chi^2(z)} P_{\delta g}\left(k = \frac{\ell + 1/2}{\chi(z)}, z\right), \tag{16}$$

$$C_\ell^{gg} = \int \mathrm{d}z \frac{H(z)}{c} \frac{W_g^2(z)}{\chi^2(z)} P_{gg}\left(k = \frac{\ell + 1/2}{\chi(z)}, z\right), \tag{17}$$

where g refers to galaxy clustering and κ refers to CMB lensing, and $P_{\delta\delta}$ to the matter power spectrum. For the cross-correlation in the case of GR, the Poisson equation can be used to express the measurement as a function of the matter power spectrum. Galaxy clustering is a biased tracer of the underlying matter power spectrum which on linear scales can approximately be described as a linear bias factor b_g: $P_{\delta g} = b_g P_{\delta\delta}$, $P_{gg} = b_g^2 P_{\delta\delta}$ and $P_{\nabla^2(\psi-\phi)g} = b_g P_{\nabla^2(\psi-\phi)\delta}$.

Each tracer considered has a specific kernel function $W_A(z)$ characterizing it. The lensing kernel for a source at redshift z_{S} is given by

$$W_\kappa(z, z_S) = \frac{3H_0^2\Omega_{m,0}}{2c^2}\hat{W}_\kappa(z, z_S), \tag{18}$$

$$\hat{W}_\kappa(z, z_S) \equiv (1+z)\chi(z)\left(1 - \frac{\chi(z)}{\chi(z_S)}\right). \tag{19}$$

For CMB lensing the source is the surface of the last scattering z_*, therefore the CMB lensing kernel is given by $W_\kappa(z) = W_\kappa(z, z_*)$, $\hat{W}_\kappa(z) = \hat{W}_\kappa(z, z_*)$.

For galaxy clustering, the kernel is given by the weighted and normalized galaxy redshift distribution

$$W_g(z) = \frac{dN}{dz}. \tag{20}$$

Note that beyond this description observed angular power spectra have additional subdominant contributions like contamination of the galaxy clustering from foreground gravitational lensing, referred to as magnification bias (discussed in Sec. V C 3). Throughout the analysis, angular power spectra for the fiducial cosmology are calculated using CCL[4] [66] with the underlying 3D power spectra calculated using CAMB [67,68].

Angular power spectra between two tracers sample the underlying 3D power spectra at a range of redshift as defined by the combination of the two window functions $W_A(\chi), W_B(\chi)$. The effective redshifts of the angular power spectra are given by [44]

$$z_{\rm eff}^{\rm AB} = \frac{\int d\chi \chi^{-2} W_A(\chi) W_B(\chi) z(\chi)}{\int d\chi \chi^{-2} W_A(\chi) W_B(\chi)}, \tag{21}$$

and more specifically for the angular cross-power spectrum $C_\ell^{\kappa g}$ and angular autopower spectrum C_ℓ^{gg} as

$$z_{\rm eff}^{\rm cross} = \frac{\int dz \chi^{-2} \hat{W}_\kappa(z) W_g(z) z}{\int dz \chi^{-2} \hat{W}_\kappa(z) W_g(z)}, \tag{22}$$

$$z_{\rm eff}^{\rm auto} = \frac{\int dz \chi^{-2}(z) H(z) c^{-1} W_g^2(z) z}{\int dz \chi^{-2}(z) H(z) c^{-1} W_g^2(z)}. \tag{23}$$

It is important to note, therefore, that the effective redshift of the auto and cross-correlation for the same galaxy sample do not match in general. Comparing the angular power spectrum measured at different effective redshifts would bias the E_G estimation. This can be avoided by reweighting each galaxy for the cross-correlation by an additional weight,

$$w_\times(z) = \frac{dN}{dz}\frac{1}{\hat{W}_\kappa(z) I}, \tag{24}$$

$$I = \int dz \frac{W_g^2(z)}{\hat{W}_\kappa(z)}, \tag{25}$$

evaluated at our fiducial cosmology where I is the normalization so that the kernel for the reweighted sample,

$$W_g^* \equiv \frac{dN^*}{dz} = \frac{dN}{dz} w_\times(z), \tag{26}$$

is correctly normalized ($\int dz W_g^* = 1$). The effective redshift of the angular cross-power spectrum with the reweighting, $C_\ell^{\kappa g *}$, then matches the effective redshift of the angular autopower spectrum. Additionally, the effective redshift for an RSD analysis based on the 3D clustering of the same galaxy data matches the effective redshift of the 2D clustering analysis [44] which in total gives

$$z_{\rm eff} \equiv z_{\rm eff}^{\rm cross*} = z_{\rm eff}^{\rm auto} = z_{\rm eff}^{\beta}. \tag{27}$$

These three measurements at the same effective redshift can be combined to build an estimator for the E_G statistic. Note for this to hold the galaxy sample needs to be treated the same for all three measurements, including the same weights except for the additional reweighting for the cross-correlation.

In the original SDSS BOSS analyses, which were referenced for the redshift of the measurement in previous E_G analyses [32,34], the weighted mean redshift of the galaxy sample was used to estimate the redshift of the measurement. This is less accurate than the above prescription, especially for angular power spectra [44].

The difference in effective redshift between auto and cross-correlation has been noted in earlier E_G estimations using CMB lensing and galaxy clustering. It was addressed in these works by a correction that relies on accurate simulations and a halo occupation distribution (HOD) [32,34,35]. In this work, by reweighting, we remove the need for such a correction.

B. Derivation of the estimator

In this section, we present a novel estimator for E_G using CMB lensing and galaxy clustering. We build on previous analyses [21,32,34,35] but make key improvements by carefully matching the effective redshifts of the observables, reducing the corrections needed, and minimizing the error introduced from approximations.

To define a consistent projected E_G estimator the 3D power spectra in Eq. (8) need to be projected with the same set of window functions. Based on the discussion of effective redshift a convenient choice is using the galaxy kernel resulting in an estimator as a function of scale ℓ at the redshift $z_{\rm eff}$,

[4]https://github.com/LSSTDESC/CCL.

$$\hat{E}_G^\ell(z_{\rm eff}) = \frac{2c^2\hat{C}_\ell^{\nabla^2(\psi-\phi)g}}{3H_0^2\beta\hat{C}_\ell^{gg}}, \tag{28}$$

where $\hat{C}_\ell^{gg}$ is the angular autopower spectrum of the galaxy sample, β is measured from an RSD analysis at the same effective redshift as the autocorrelation and

$$\hat{C}_\ell^{\nabla^2(\psi-\phi)g} = \frac{1}{2}\int \mathrm{d}z \frac{H(z)}{c(1+z)}\frac{W_g^2(z)}{\chi^2(z)}\hat{P}_{\nabla^2(\psi-\phi)g}(k,z). \tag{29}$$

This helper function is closely related to the observable cross-correlation $C_\ell^{\kappa g*}$. We, under careful consideration of the approximations needed, equate this expression with $C_\ell^{\kappa g*}$, minimizing the overall error of the estimator.

By applying the reweighting scheme [Eq. (24)] to the cross-correlation analyses in order to match the effective redshift of the cross-correlation to the autocorrelation, the galaxy kernel for the cross-correlation is

$$W_g^*(z) = W_g(z)w_\times(z). \tag{30}$$

Plugging in $w_\times(z)$ we have that the projection kernel $W_g^2(z)$ can be expressed as

$$W_g^2(z) = \hat{W}_\kappa(z)W_g^*(z)I. \tag{31}$$

This reweighting of the galaxy sample removes the mismatch in effective redshift that would otherwise be present between the auto and cross-correlations. Without this reweighting a scale-dependent bias of multiple percent, depending on $\mathrm{d}N/\mathrm{d}z$, would be present in the estimator (see Appendix B). We note that the reweighting also affects the variance of the weights which affects the shot noise in the galaxy data. In practice for our particular case, where we also apply the redshift-dependent FKP weights the redshift-dependent weights partially cancel so that the overall effect of the reweighting is a decrease in variance compared to the sample for the autocorrelation increasing the constraining power of the cross-correlation.

Using Eq. (31) the helper function can be well-approximated with a cross-correlation measurement as

$$\hat{C}_\ell^{\nabla^2(\psi-\phi)g} \approx \frac{H(z_{\rm eff})I}{c}\int \mathrm{d}z \frac{\hat{W}_\kappa(z)W_g^*(z)}{\chi^2(z)}\hat{P}_{\nabla^2(\psi-\phi)g}(k,z) \tag{32}$$

$$= \frac{H(z_{\rm eff})I}{c}C_\ell^{\kappa g*}, \tag{33}$$

where $C_\ell^{\kappa g*}$ is our cross-correlation measurement with the reweighted galaxy sample and we approximate $H(z)$ as slowly varying within the redshift sample.

In summary, the new E_G estimator based on CMB lensing and a galaxy sample is given by

$$\hat{E}_G^\ell(z_{\rm eff}) \approx \Gamma(z_{\rm eff})\frac{C_\ell^{\kappa g*}}{\beta C_\ell^{gg}}, \tag{34}$$

where

$$\Gamma(z_{\rm eff}) \equiv \frac{2cH(z_{\rm eff})}{3H_0^2}\int \mathrm{d}z \frac{W_g^2(z)}{\hat{W}_\kappa(z)} \tag{35}$$

and $z_{\rm eff}$ is the effective redshift of the observables given by Eq. (27).

C. Accuracy of the E_G estimator

Throughout the derivation of the E_G estimator presented in Sec. V B, some approximations were needed. While the overall bias is minimized by carefully matching the effective redshift of each measurement through the use of a reweighting scheme, β, and H are assumed as slowly varying within the redshift range of the galaxy sample and evaluated at the effective redshift of the measurement. Astrophysical effects like an evolution of the galaxy bias as a function of redshift and the effect of magnification bias can affect the result. We now characterize the accuracy of the estimator as defined in Eq. (34) and its sensitivity to astrophysical effects for the *Planck* CMB lensing maps and SDSS BOSS galaxy samples considered in this work. In this work, we neglect errors in the redshift estimation of the galaxies since they are based on spectroscopic observations and are expected to be small. For the reweighting, the comoving distance across the redshifts covered by the galaxy sample at the fiducial cosmology is used [see Eqs. (19) and (24)]. This is not a concern since the E_G statistic tests models that reproduce the expansion history as measured for our Universe but predict deviations in the growth of structure. We neglect the small uncertainty in the fiducial cosmology for the reweighting in this analysis.

First, in Sec. V C 1, we consider the numerical accuracy of the estimator when ignoring the astrophysical complications of dependencies in the galaxy bias and the effect of magnification bias. Then, in Sec. V C 2 we quantify the systematic bias from the redshift evolution of the galaxy bias. Finally, in Sec. V C 3, we consider that the observables are affected by magnification bias for which we quantify the effect on our estimator and find a correction in the case of CMASS.

1. Systematic without astrophysical complications

The overall systematic bias of the estimator as a result of the approximations made during the derivation can be quantified by comparing the estimator calculated from the analytic observables as defined in Eq. (34) for the fiducial cosmology with the $E_G^{\rm GR}$ value at the fiducial cosmology.

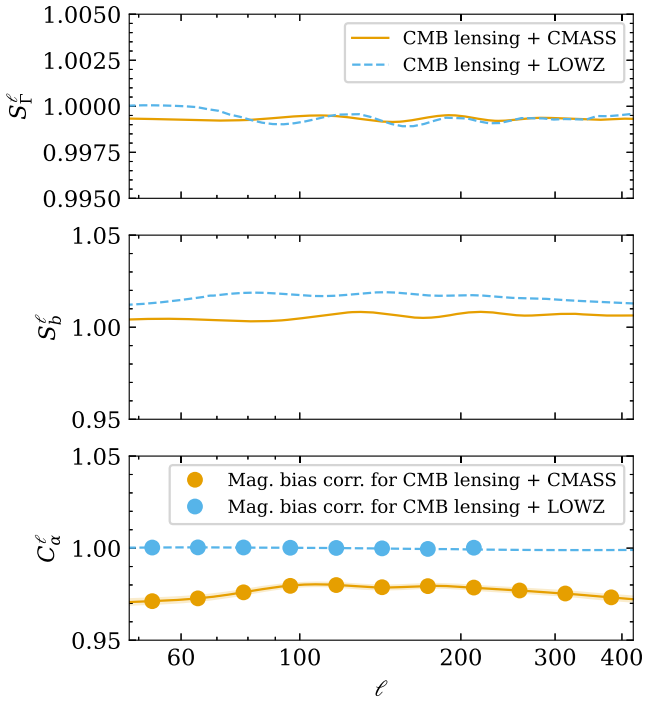

FIG. 2. Accuracy of the E_G estimator presented in this work for CMB lensing combined with the CMASS (orange, solid line) and LOWZ (blue, dashed line) galaxy samples with one indicating consistency between estimator and input. (Upper panel) The overall systematic error of the estimator, S_Γ^ℓ, before considering astrophysical systematics. (Middle) Our estimate of the systematic error from bias redshift evolution within the galaxy samples of CMASS and LOWZ, S_b^ℓ. (Lower) The overall impact of magnification bias, C_α^ℓ, which we correct for with the binned correction shown as dots.

The overall systematic bias is given by

$$S_\Gamma^\ell \equiv \frac{E_G^{\rm GR}(z_{\rm eff})}{\hat{E}_G^\ell(z_{\rm eff})}. \tag{36}$$

The upper panel of Fig. 2 shows the result for S_Γ^ℓ at our fiducial cosmology. The $E_G^{\rm GR}$ value is calculated using Eq. (6) for our fiducial cosmology. The $\hat{E}_G^\ell(z_{\rm eff})$ value is calculated by plugging into Eq. (34) the angular power spectra and β calculated for our fiducial cosmology with a constant galaxy bias of $b_g = 2.0$ and no magnification bias using CCL (see Sec. V A). For the estimator presented in this work, we find that the S_Γ^ℓ systematic is negligible compared to the observational uncertainty, with $|S_\Gamma^\ell - 1| < 0.3\%$ on all scales in our cosmological ranges and for both CMB lensing + CMASS and CMB lensing + LOWZ. We conservatively include the upper limit of 0.3% in our systematic error budget.

For a previous harmonic space estimator for E_G presented in Pullen *et al.* [34] a similar quantity referred to as C_Γ has been investigated; the quantity S_Γ^ℓ also considers the redshift evolution of the growth f but is otherwise analogous (not considering the redshift evolution of f leads to an amplitude difference of around 0.1% so they are equivalent within the accuracy relevant to this discussion). Therefore, we can compare the overall systematic of the two estimators. For the previous estimator, the systematic was at the level of a few percent, depending on $\mathrm{d}N(z)/\mathrm{d}z$, and required a correction based on N-body simulations and assuming a HOD for the galaxy samples [21,34,35]. In Appendix B we compare the estimator presented in this work with the one presented in Pullen *et al.* [21] in detail and we discuss the impact of each difference. For the estimator presented in this work, no correction based on simulations needs to be applied before considering astrophysical effects since the bias S_Γ is negligible, and therefore the results do not rely on the accuracy of simulations or choice of HOD.

2. Systematic from bias evolution with redshift

Next, one can consider the effect of redshift evolution of the galaxy bias. So far a constant galaxy bias has been assumed on linear scales. To separate β and C_ℓ^{gg} there is an implicit assumption that the galaxy bias b_g is not only scale independent but also constant in redshift. If there is a significant evolution of the galaxy bias with redshift within a given galaxy sample this can introduce a systematic bias. Given a bias evolution $b_g(z)$ we can analytically estimate the systematic bias to the estimator from this redshift evolution as

$$S_b^\ell = \frac{C_\ell^{gg}}{b_g C_\ell^{\delta g}}, \tag{37}$$

$$C_\ell^{\delta g} \equiv \int \mathrm{d}z \frac{H(z)}{c} \frac{W_g^2(z)}{\chi^2(z)} P_{\delta g}(k, z), \tag{38}$$

where b_g is the constant galaxy bias at the effective redshift as before but the power spectra projected in the angular power spectra are given by $P_{gg} = b_g^2(z) P_{\delta\delta}$ and $P_{\delta g} = b_g(z) P_{\delta\delta}$ with $b_g(z_{\rm eff}) = b_g$. We implement this calculation using CCL. The implementation of `ccl.NumberCountsTracer` is already set up to handle a redshift-dependent bias as an input which we use to calculate C_ℓ^{gg} here. For $C_\ell^{\delta g}$ we correlate one `ccl.NumberCountsTracer` with the redshift-dependent bias and one with a bias equal to one resulting in the quantity of Eq. (38).

This characterization of systematic error due to the deviation of galaxy bias from a constant is identical to "$C_b(\ell)$" in previous analyses [34,35]. In Pullen *et al.* [34] this was studied for CMASS based on power spectra from N-body simulations and assuming an HOD testing for the impact of bias evolution and also scale dependence of the

bias. They report systematics of between 0% and 1% over their cosmological range.

We can also study this systematic from measurements of $b_g(z)$ based on the data. To do this measurement one has to assume GR, so we can not directly differentiate an astrophysical bias evolution within the sample from a deviation from GR that affects the clustering amplitude. Therefore, we can not easily correct this systematic by constraining the biased evolution of the galaxy samples. However, we can study how large the effect on our estimator is to inform if a measured disagreement with GR could be caused by the bias evolution informing our systematic bias estimate. If it is found that the systematic bias from bias evolution is relevant compared to the observational uncertainty, narrower bins in redshift would need to be used to reduce the systematic bias. We note that in principle there could be a scenario where there is a strong bias evolution with a redshift that is largely canceled by an opposite deviation from GR in the growth history. In such a case, the systematic bias would not be fully captured by this estimate. We neglect this in this work but note that this could be investigated in future work with higher constraining power by comparing E_G results for narrower bins in redshift with results for the full samples.

In Farren *et al.* [69] the galaxy bias was constrained for CMASS and LOWZ in redshift increments of $\Delta z = 0.05$ assuming a ΛCDM growth history. Using the values reported in their Table 13 and performing a linear least squared fit results in slopes s of approximately $s = [1.2, 1.6, 1.8, 1.8]$ for CMASS North, CMASS South, LOWZ North, and LOWZ South. We use $b_{g,\mathrm{LOWZ}}(z) = 2.0 + 1.8(z - 0.316)$ and conservatively assume the steeper evolution for CMASS $b_{g,\mathrm{CMASS}}(z) = 2.0 + 1.6(z - 0.555)$. Note that for this test the focus is the estimate of the evolution with redshift, the test is only weakly sensitive to small differences in the overall amplitude from the fiducial value of 2.0. We note that the best-fit for the values from Farren *et al.* [69] for the LOWZ amplitude is approximately 2.0 for both North and South and for the CMASS amplitude is 2.2 and 2.1 for the North and South samples respectively, all consistent with the overall b_g fitted to data in this work (see Sec. VI B 2).

The middle panel in Fig. 2 shows the constraints on S_b^ℓ. Given the estimated bias redshift evolution, the systematic bias in E_G when using CMASS is at most 1% and when using LOWZ at most 2%. These systematic biases are much smaller than the statistical uncertainty of the measurement and therefore bias evolution with redshift is not of significant concern for the analysis. We account for the effect in our systematic error budget with 1% uncertainty on E_G for CMASS and 2% for LOWZ.

3. Correcting for magnification bias

So far in the discussion, we have neglected the effect of magnification bias on the observable cross- and autocorrelation. The brightness of observed galaxies, similarly to their shape, gets modulated by lensing along the line of sight from intervening gravitational potential [70–75]. For surveys with a detection threshold, this magnification affects the local observed number density of galaxies and thereby introduces a systematic effect known as the magnification bias.

In the literature, this is known to be a significant systematic for current-generation surveys, especially on the cross-correlation signal, and therefore needs to be accounted for in cosmological analyses [57,76–81]. Specifically for the E_G statistic it has been highlighted as an important systematic that should be accounted for [82–84] and has been included in a recent E_G constraint using SDSS eBOSS [35]. However, the magnification bias effect has been neglected by previous E_G estimates using SDSS BOSS and Planck CMB lensing [32,34].

The magnification bias effect can be quantified via the function $\alpha(z)$ that describes the sensitivity of the sample selection to lensing. For a review of the functional form of the impact of the magnification bias on the observable angular auto and cross-power spectra, we refer the interested reader to Wenzl *et al.* [81]. We leverage the recent measurements of α for the LOWZ and CMASS samples from [81] for our specific set of analysis choices and accounting for the full photometric selections of the samples.

For LOWZ we use $\alpha_{\mathrm{LOWZ}} = 2.47 \pm 0.11$ where we use the mean value between the best-fit value for the galaxy sample with and without reweighting for the cross-correlation. The uncertainty conservatively combines statistical uncertainty (0.02), systematic uncertainty from redshift evolution (0.07) as well as half the difference between with and without reweighting for the cross-correlation (0.02).

For CMASS we account for the observed redshift evolution by using the best-fit function,

$$\alpha_{\mathrm{CMASS}}(z) = 2.71 + 8.78(z - 0.55), \tag{39}$$

with statistical uncertainties for the constant and slope of 0.08 and 1.26. We conservatively widen the uncertainty of the constant to 0.10 to account for the reported systematic uncertainty from the light profile choice of 0.02.

In the estimate of $\hat{E}_G$, we account for the effect of magnification bias by including a correction term C_α^ℓ throughout our analysis by replacing the Γ factor in Eq. (34) as

$$C_\alpha^\ell \equiv \frac{C_\ell^{\kappa g*}}{C_\ell^{\kappa g*,\mathrm{mag}}} \frac{C_\ell^{\mathrm{gg,mag}}}{C_\ell^{gg}}, \tag{40}$$

$$\Gamma \to \Gamma_\ell \equiv \Gamma C_\alpha^\ell, \tag{41}$$

where the $C_\ell^{\kappa g*,\mathrm{mag}}$ and $C_\ell^{\mathrm{gg,mag}}$ are the angular cross- and autopower spectra additionally including the effect of

magnification bias calculated as described in [81]. All angular power spectra are calculated at the fiducial cosmology and binned using the mode coupling matrix as described in Sec. VI A. The correction factors for LOWZ and CMASS are shown in Fig. 2. For LOWZ the correction is negligible, with corrections of $< 0.1\%$ in the cosmological range of this analysis. For the case of *Planck* + CMASS, the magnification bias impact is 2–3% and we correct this effect in our analysis with the binned corrections shown as dots in the figure. Note that this correction is such that it reduces the amplitude of the E_G measurement for both SDSS BOSS samples and any other galaxy sample with $\alpha > 1$.

When plotting the measurement of angular power spectra compared to theory spectra we also include the magnification bias contribution in the comparison theory curves. The magnification bias contribution is also consistently included in the simulated galaxy maps and covariance estimation. From the uncertainty of the magnification bias, we estimate the uncertainty of the correction applied. Like the correction itself, this is negligible for LOWZ. For CMASS the uncertainty of $\alpha(z)$ translates to at most 0.2% uncertainty in C_α^ℓ, shown in Fig. 2 as a narrow band around the estimate. Based on this we account for a 0.2% error for our systematic budget of the estimate involving CMASS.

VI. MEASURING THE OBSERVABLES FROM DATA

In this section, we describe the signal extraction approach and discuss how we characterize the uncertainty on our observables and validate our pipeline. The angular power-spectra estimation approach is outlined in Sec. VI A. Sections VI B and VI C describe the use of simulations for covariance estimation and to perform pipeline validation and Sec. VI E describes a test of the shot noise estimate. The power spectrum measurements from the *Planck* PR4 CMB lensing map and the SDSS BOSS galaxy samples are summarized in Sec. VI F and the RSD parameter β in Sec. VI G.

A. Angular power spectrum estimation

To measure the angular auto and cross-power spectra from our observed maps we use a pseudo-Cl estimator. Given two maps (A, B) expanded in spherical harmonics ($a_{\ell m}$, $b_{\ell m}$) one can estimate the auto or cross-power spectrum at each multipole ℓ as

$$\hat{D}_\ell^{AB} = \frac{1}{2\ell+1}\sum_{m=-\ell}^{\ell} \mathrm{Re}(a_{\ell m} b_{\ell m}^*). \tag{42}$$

In general, our observations only cover a fraction of the sky so that the overall amplitude of this simple estimate $\hat{D}_\ell$ is approximately a factor $f_{\rm sky}$, the sky fraction of the overlap between the maps, lower than the expectation for the full sky. Additionally, complex mask geometries couple originally independent modes at neighboring scales. To reduce the correlation between the measurements, the large noise of the individual multiples, and for computational efficiency, we measure the angular power spectra in band powers i.[5] We use the unbiased pseudo-C_ℓ estimator NaMaster[6] presented in Alonso *et al.* [85]. The estimates for the angular power spectra are given by

$$\hat{C}_i^{AB} = \sum_{i'} [M^{-1}]_{ii'} \frac{1}{N_{i'}} \sum_{\ell \in i'} \hat{D}_\ell^{AB}, \tag{43}$$

where N_i is the number of multipoles ℓ in band power i and M is the so-called binned coupling matrix. This binned coupling matrix is defined by the geometry of the observational masks of the two maps being correlated as well as the band powers and it accounts for both the loss of overall amplitude from partial sky coverage and the coupling between modes (see [85] for further details).

Additionally, for the galaxy autocorrelation C_ℓ^{gg} the shot noise in the map correlates with itself. This contribution needs to be subtracted to estimate the underlying cosmological correlation. We do this before binning as

$$\hat{D}_\ell \to \hat{D}_\ell - \tilde{N}^{\rm shot}, \tag{44}$$

where the coupled shot noise estimate $N^{\rm shot}$ is from Eq. (12).

We use a map resolution parametrized by nside = 1024 throughout our analysis.[7] Since the galaxy map is constructed in real space we need to account for the loss of power due to the finite pixelization used [65]. To do this we apply the pixel window function specific to our map resolution as an effective beam function for the galaxy field objects in NaMaster. For the autocorrelation, this is done after shot-noise subtraction.

We also use the mode coupling matrix to find the expectation value of the angular power spectrum for each band power given by

$$\langle \hat{C}_\ell \rangle = \sum_{\ell'} M_{\ell\ell'} C_{\ell'}, \tag{45}$$

where C_ℓ is the unbinned theory spectrum. We apply this throughout our analysis when comparing theory curves to the measured $\hat{C}_i^{AB}$ and to bin scale-dependent correction factors into the band powers.

[5] In this section we use the subscript i for clarity but then simplify to using ℓ for the rest of the work referring to the effective ℓ of each band power which we also use in plots.

[6] https://github.com/LSSTDESC/NaMaster.

[7] This represents a map with $N_{\rm pix} = 12 \cdot \mathrm{nside}^2 = 12,582,912$ pixel.

We use a log-spaced binning scheme for our analysis. For computational accuracy and to avoid leakage at the edges we perform all computations in the full range given by our map resolution ($\ell \in [2, 3071]$).

For our cosmological analysis, however, we only use linearly dominated scales up to $k = 0.2$ Mpc^{-1}, a real space cutoff of around 30 Mpc (comoving). For our logarithmic binning, these scale cuts limit us to $\ell_{\rm max} = 420$ and $\ell_{\rm max} = 233$ for the effective redshifts of CMASS and LOWZ, respectively. We also limit the cosmological analysis on large scales to $\ell_{\rm min} = 48$ following the recommendation in Chen *et al.* [44] to conservatively avoid large scales where foregrounds in the galaxy maps could introduce spurious correlations with CMB lensing.

The 11 log-spaced bins in the cosmological range used throughout the analysis have edges [48, 59, 71, 87, 106, 129, 157, 191, 233, 283, 345, 420] where each bin is inclusive of the lower edge and exclusive of the upper edge as per NaMaster convention. In the following, we will label each bin by their effective ℓ which are given by [53, 64.5, 78.5, 96, 117, 142.5, 173.5, 211.5, 257.5, 313.5, 382] with the analysis for LOWZ only using the first eight bins.

B. Simulated data

To validate our estimates of the angular correlation functions, and to estimate the covariance of the measurements, we consider simulated realizations of the observables. The CMB lensing reconstruction has a known transfer function, in Sec. VI B 1 we use the simulated realizations of the CMB lensing maps, described in Sec. IV A, to estimate a norm correction to account for this. In Sec. VI B 2, we create realizations of the galaxy overdensity maps with and without reweighting for the cross-correlation. We create them with the same level of shot noise as our galaxy samples and give them the expected correlation with the CMB lensing realizations and to each other. Finally, in Sec. VI D we use our simulated observables to test our measurement code against the input.

For the theory curves to draw our simulated realizations from, we use our fiducial cosmology discussed in Sec. II.

1. Simulated CMB lensing maps and norm correction

We use the set of 480 simulations used in Carron *et al.* [45] to estimate the norm correction of the *Planck* PR4 CMB lensing maps and to estimate the covariances of our cross-correlations with the CMASS and LOWZ galaxy samples. For all of these, we know the Gaussian realization of the CMB lensing angular power spectrum used as an input for the simulation ($a^{\kappa}_{\ell m}$).

It is well known that lensing reconstruction with quadratic estimators in the presence of a mask can lead to offsets of the overall normalization [69,86,87]. This normalization offset was not accounted for in previous E_G measurements with *Planck* and SDSS [32,34,35].

In the context of cross-correlation with galaxy samples, Farren *et al.* [69] showed that this overall normalization can be corrected based on end-to-end simulations of the pipeline. Using the approach in [69] and Krolewski *et al.* [88], we calculate the Monte Carlo correction based on the set of 480 simulations as the ratio between the cross-correlation of appropriately masked CMB lensing input realizations with the cross-correlation of the CMB lensing input realization to the map reconstruction. The functional form is given by

$$\text{mc-corr}(\ell) \equiv \frac{\bar{C}_\ell^{\kappa_{g-\text{mask}},\kappa_{\text{CMB-mask,apodized}}}}{\bar{C}_\ell^{\kappa_{g-\text{mask}},\hat{\kappa}_{\text{apodized}}}}, \tag{46}$$

$$\hat{a}^{\kappa}_{\ell m} \to \text{mc-corr}(\ell) \cdot \hat{a}^{\kappa}_{\ell m}, \tag{47}$$

where $\kappa_{g-\text{mask}}$ are the input κ maps $a_{\ell m}$ with the mask of the galaxy sample to be correlated with applied, κ_{apodized} are the reconstructed maps $\hat{a}^{\kappa}_{\ell m}$ with the same apodization and low-pass filter we apply to the data and $\kappa_{\text{CMB-mask,apodized}}$ are the input κ maps $a_{\ell m}$ with the CMB lensing mask applied including the same apodization that is applied to the data. We here use a hat to indicate maps estimated with our pipeline and no hat for the input maps. Note that the *Planck* PR4 CMB lensing mask is binary. We take the average angular power spectrum $\bar{C}_\ell$ over all 480 simulations. The correlation is calculated without accounting for the mask coupling in [see Eq. (42)] since we do not want to account for the coupling matrix in our norm correction to avoid correcting it twice. We use the implementation `healpy.sphtfunc.anafast` [89,90].

For the *Planck* PR4 κ map, anisotropic filtering was used which can make the norm correction dependent on the specific area of overlap of the galaxy sample that we correlate with. While for the full map, the norm correction is approximately 2%, we find that for the overlap with CMASS and LOWZ, the Monte Carlo correction (mc-correction) for each is approximately 4% up in the cosmological range. We apply the mc-correction throughout our analysis. We recalculate the mc-correction for each galaxy sample and the North and South patches separately when considering them separately in our analysis. The norm correction affects the amplitude of the angular cross-correlation measurements and thereby directly shifts our E_G measurement. We will discuss the level of impact the norm correction has on our measured result in Sec. VIII B.

2. Simulated correlated galaxy maps

While the E_G statistic is designed to optimally cancel the galaxy bias to create simulated realizations of the observation we need to estimate the galaxy bias for each galaxy sample. We iteratively update the galaxy bias to ensure the

amplitudes of the theory curves for the galaxy autocorrelation used to make our simulations match well with the clustering amplitude in our galaxy samples. For this purpose, we measured the C_ℓ^{gg} and then fitted our theory curve to the data with only the galaxy bias as a free parameter. We performed the fit in the cosmological ranges for CMASS and LOWZ and used relative weights for each bin informed by the analytic error (see Sec. VI C for the description of our analytic error calculation).

For the initial analytic error, we assumed our fiducial cosmology and an initial guess of 2.0 for the galaxy bias for both the CMASS and LOWZ samples, consistent with the literature [91,92]. We also include magnification bias in the theory curves since the effect is in the measurement and to ensure that our end-to-end pipeline test covers the magnification bias correction. For our full CMASS (LOWZ) galaxy map and fiducial cosmology, we find a best-fit value of $b_g = 2.14$ ($b_g = 2.02$). We use these best-fit galaxy bias values, the magnification bias measurement, and our fiducial cosmology to define the theory curves used to create the simulated galaxy maps.

With the theory curves established, we now leverage them to create the galaxy maps with the expected cross-correlation with our set of CMB lensing simulations and also the expected autocorrelation including the shot noise contribution. Given the need to consider two different weighting schemes for the auto and cross-correlation analysis [see Eq. (34)], two versions of each galaxy map simulations are created to effectively match the respective redshift distributions of the two weighting schemes while also being correctly correlated to each other. Given our theory spectra, we build the $a_{\ell m}$ of the simulated maps by combining (building on the approaches in [69,93]),

$$a_{\ell m}^{g^X} = \frac{C_\ell^{\kappa g^X}}{C_\ell^{\kappa\kappa}} a_{\ell m}^{\kappa} + a_{\ell m}^{g^X,\mathrm{uncorr}} + a_{\ell m}^{g,\mathrm{noise}}, \tag{48}$$

where $g^X \in [g, g^*]$ and the contributions are drawn so that

$$\left\langle a_{\ell m}^{g^X,\mathrm{uncorr}}, \left(a_{\ell m}^{g^Y,\mathrm{uncorr}}\right)^* \right\rangle = \delta_{\ell\ell'}\delta_{mm'}\left(C_\ell^{g^X g^Y} - \frac{C_\ell^{\kappa g^X} C_\ell^{\kappa g^Y}}{C_\ell^{\kappa\kappa}}\right) \tag{49}$$

$$\left\langle a_{\ell m}^{g^X,\mathrm{noise}}, \left(a_{\ell m}^{g^X,\mathrm{noise}}\right)^* \right\rangle = \delta_{\ell\ell'}\delta_{mm'} n_{\mathrm{eff}}^{g^X}. \tag{50}$$

Here $n_{\mathrm{eff}}^{g^X}$ is the effective noise for each galaxy weighting scheme as defined in Eq. (13).

For the implementation, we use the `synalm` function from HEALPY and then we apply the respective galaxy mask to the sims. The $a_{\ell m}^{\kappa}$ are the set of Gaussian realizations of the CMB lensing field used for the generation of the simulated CMB lensing maps. We note that the $a_{\ell m}^{\kappa}$ were generated for a slightly different input $C_\ell^{\kappa\kappa}$ than our fiducial cosmology. To account for this and avoid a slight bias in the simulated maps, we use the exact $C_\ell^{\kappa\kappa}$ used to generate the $a_{\ell m}^{\kappa}$.

For the shot noise in the galaxy maps, we assume Gaussianity and sample the noise for each pixel according to our effective noise as well as the coverage fraction for each pixel. The shot noise between the galaxy maps with and without reweighting is correlated, to account for this we use the same realization of the standard normal for both and scale them according to the shot noise in each. This approach ensures that the simulated maps are correctly correlated between them so we can accurately estimate the full covariance including cross-covariances.

C. Covariance estimation

We need to carefully characterize the covariance matrix of our measurement. Due to the complicated mask geometry of the galaxy samples, measurements at different scales become correlated. As discussed in Sec. VI A we are using wide bins which reduce the correlation between bins, but neighboring bins will still have some amount of correlation. Additionally, the auto and cross-correlation at similar scales are also expected to be correlated with each other as both scale with the amount of clustering in the galaxy data.

We use multiple approaches to estimate the full covariance to check for consistency and validate our approach. We use our set of simulated maps to estimate the covariance for our baseline. We then compare the simulated covariance to an analytic estimate where we compare both the marginalized errors and the off-diagonals. Finally, we perform an internal error estimate on the data using a jackknifing technique to validate the marginalized errors.

As our baseline, we estimate the sample covariance from our set of $N_{\mathrm{sims}} = 480$ measurements of the angular power spectra on our simulated maps as

$$\widehat{\mathrm{Cov}}(X_\ell, Y_{\ell'}) = \frac{1}{N_{\mathrm{sims}} - 1}\sum_{k=1}^{N_{\mathrm{sims}}} (X_\ell^{(k)} - \bar{X}_\ell)(Y_{\ell'}^{(k)} - \bar{Y}_{\ell'}), \tag{51}$$

where $X_\ell, Y_\ell \in [\hat{C}_\ell^{\kappa g^*}, \hat{C}_\ell^{gg}]$ are the angular power spectra for our simulated maps discussed in Sec. VI B and $\bar{X}_\ell, \bar{Y}_\ell$ are the means over all simulations. With this, we can estimate the covariance of each observable as well as the cross-covariance between the auto and cross-correlations. We note that here it is crucial that the galaxy sample with and without reweighting are correctly correlated in our simulated maps so that the cross-covariance is captured in the estimate. While the estimate of the covariance from simulations is unbiased, the inverse is a biased estimate. We correct for this by applying a Hartlap correction when inverting the covariance matrix given by [94]

$$\widehat{\mathrm{Cov}}^{-1} \to \widehat{\mathrm{Cov}}^{-1}\left(1 - \frac{N_d + 1}{N_{\mathrm{sims}}}\right), \tag{52}$$

where N_d is the length of the data vector considered.[8]

We compare our estimate from simulations with an analytic estimate as well as jackknife errors to test for consistency. We analytically estimate the full covariance including the cross-covariances assuming Gaussianity and using our theory power spectra based on our fiducial cosmology and measured galaxy bias (see Sec. VI B). In the absence of a mask, the off diagonals would be zero and the error for each multipole would be given by [65]

$$(\sigma_\ell^{XY})^2 = \frac{1}{2\ell + 1}[C_\ell^{XX} C_\ell^{YY} + (C_\ell^{XY})^2], \tag{53}$$

where $X, Y \in [\kappa, g, g*]$. Here C_ℓ^{XX} and C_ℓ^{YY} refer to the autocorrelation inclusive of the noise in the map. For the galaxy autocorrelation, we use the theory autocorrelation plus our shot noise estimate N_ℓ. For the CMB lensing autocorrelation we use the theory autocorrelation plus the noise curve of the *Planck* PR4 CMB lensing map. The single noise curve does not account for the anisotropy of the noise in the CMB lensing map which can make the analytic covariance less accurate overall [45]. The complex mask geometry leads to coupling between different modes which can approximately be analytically modeled [95,96]. We use the implementation presented in García-García *et al.* [97] which introduced the `NmtCovarianceWorkspace` class within NaMaster. The implementation takes as inputs the angular power spectra inclusive of noise as well as the mask geometry in the form of the coupling matrices introduced in Eq. (43). We note that we include magnification bias for the theory spectra here but it has very little impact. For the cross-correlation of *Planck* CMB lensing and CMASS, the covariance increases by about half a percent compared to not including magnification bias.

The off diagonals of the simulated and analytic covariance for *Planck* + CMASS are visualized as a correlation matrix in Fig. 3. The correlation matrix is calculated as $\mathrm{Cov}(C_\ell^{XY}, C_\ell^{AB})/(\sigma_\ell^{XY}\sigma_\ell^{AB})$. We find good agreement between the two techniques up to the statistical uncertainty of the simulated covariance. Additionally, we find, as expected, a significant correlation between the auto and cross-angular power spectra with correlation factors up to 0.35 on our largest scales in the cosmological analysis range. We also compare the marginalized errors of the covariances in the middle panels of Fig. 6 where we show that in the cosmological analysis range the errors using our simulated covariance and analytic covariance agree at the percent level. We also show the same consistency of the covariance estimates for *Planck* + LOWZ in Appendix C.

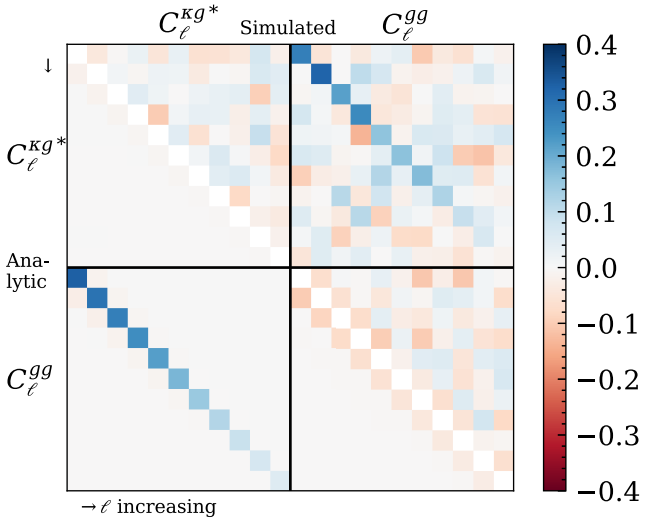

FIG. 3. Visualizing the off-diagonal correlation matrix between $C_\ell^{\kappa g*}$ and C_ℓ^{gg} for *Planck* and CMASS. Shown is the correlation matrix defined as $\mathrm{Cov}(C_\ell^{XY}, C_\ell^{AB})/(\sigma_\ell^{XY}\sigma_\ell^{AB})$ for the bins in the cosmological range. The upper triangle shows the simulated covariance which is the baseline for the results presented in this work and the lower triangle shows the analytic covariance. An arrow indicates the direction of increasing ℓ for each axis. Correlations between C_ℓ^{gg} and $C_\ell^{\kappa g*}$ are up to around 0.35 for our largest-scale bin. The off-diagonals of the simulated covariance match the analytic covariance well within statistical uncertainty. For a comparison of the diagonals of the covariances see Fig. 6. For *Planck* + LOWZ see Appendix C.

Estimating the covariance from simulations or analytic modeling assumes that we include all sources of noise present in the data. We can validate these error estimates by comparing them to an internal estimate from the data using a jackknife approach. For jackknife resampling, we split the overlapping area into N_{jack} patches.[9] From these patches, we create N_{jack} maps with one patch left out in each. For each of these, we calculate the auto and cross-correlations. We can then estimate the covariance as [98]

$$\widehat{\mathrm{Cov}}(X_\ell, Y_{\ell'}) = \frac{N_{\mathrm{jack}} - 1}{N_{\mathrm{jack}}} \sum_{k=1}^{N_{\mathrm{jack}}} (X_\ell^{(k)} - \bar{X}_\ell)(Y_{\ell'}^{(k)} - \bar{Y}_{\ell'}), \tag{54}$$

where $X_\ell, Y_\ell \in [\hat{C}_\ell^{\kappa g*}, \hat{C}_\ell^{gg}]$. For the autocorrelations, we need to account for the change in shot noise in the galaxy map when removing a patch. We adapt the analytic shot noise estimate [Eq. (12)] for the subsample as

[8]For considerations of each 2 pt correlation function this is the number of bins in the cosmological range. For the $\hat{E}_G$ estimate we invert the full covariance of both 2 pt correlation functions doubling N_d and for the bin-wise estimate $\hat{E}_G^\ell$ we have $N_d = 2$.

[9]We recursively bisect the overlapping area along lines of RA and DEC. We split so that each patch contains equal $\sum f_i$ based on our galaxy mask and split along the more spread out direction.

$$\tilde{N}^{\mathrm{shot},(\mathrm{k})} = N^{\mathrm{shot}} \frac{N_{\mathrm{jack}-1}}{N_{\mathrm{jack}}}. \tag{55}$$

For the cross-covariance between $\hat{C}_\ell^{\kappa g*}, \hat{C}_\ell^{gg}$ we restrict the autocorrelation to the same footprint as the cross-correlation and use the same splits to capture the correlation correctly.

The smallest patch is chosen to still contain the largest scales of the cosmological range following Pullen *et al.* [34]. Based on this we decided to use 32 patches. Because of this small number, the statistical uncertainty of the marginalized jackknife error for each X_ℓ is large compared to the other methods above. Additionally, jackknife resampling overestimates the covariance of the measurement systematically [99,100]. This is easiest to see in real space where 2 pt functions are based on pairs of objects. Pairs across patches are doubly removed from the sampling and therefore the variance is overcounted [100].

We use the jackknife errors to validate the marginalized bin-wise errors of our angular power spectrum measurements. They are a conservative overestimate of 10–20% with large statistical uncertainty [99] and therefore we interpret them as validating our error estimation if they are on average 10–20% larger than our baseline error estimate and there are no outliers more than 50% larger. In the middle panels of Fig. 6 we compare the marginalized jackknife errors to the other techniques finding that they are on average around 20% higher with statistical variation between 0% and 50% higher, consistent with our expectations. From this reasonable agreement, we conclude that in the cosmological range, the noise in the data is well-represented in the simulations we use to estimate our baseline covariance.

We do not have sufficient samples to accurately estimate the full covariance including the off-diagonal elements with the jackknifing technique and therefore do not give a constraint on $\hat{E}_G$ marginalized across scales for the jackknife errors. We, however, do use them to estimate the bin-wise $\hat{E}_G^\ell$ where we only need the marginalized error per bin and the cross-covariance between the two. We again apply Hartlap corrections when inverting the covariance where in Eq. (52) we replace N_{sims} with N_{jack}.

D. Input recovery on simulations

Based on our set of 480 simulated map realizations we test if our analysis pipeline accurately recovers the input angular power spectra for the simulations. In Fig. 4 we show the ratio of the mean recovered cross and autocorrelation signals, $\bar{\hat{C}}_\ell^{\kappa g*}$ and $\bar{\hat{C}}_\ell^{gg}$, for our set of 480 simulations to the input angular power spectra. Shown is the uncertainty of a single measurement σ (orange/blue) and the uncertainty on the mean $\sigma/\sqrt{N_{\mathrm{sims}}}$ (red).

The level of agreement with the input is quantified by calculating the probability to exceed the given χ^2 value

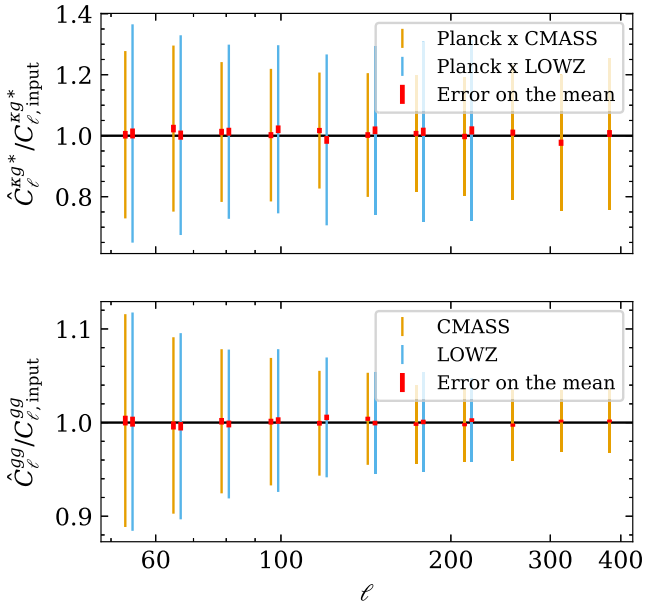

FIG. 4. Demonstration of pipeline accuracy on simulated data. The ratio of the mean estimated cross-correlation, $\hat{C}_\ell^{\kappa g}$, (upper) and autocorrelation, $\hat{C}_\ell^{gg}$, (lower) with the input theory spectrum is shown for the cases involving CMASS (orange) and LOWZ (blue, slightly shifted to the right for readability) with the statistical uncertainty of a single measurement, σ, and of the mean for $N_{\mathrm{sims}} = 480$ simulations, $\sigma/\sqrt{N_{\mathrm{sims}}}$ (red).

(PTE) over the set of scales in the range of cosmological scales to be used to analyze the experimental data. For this, the covariance on the mean is the full simulation-based covariance in the cosmological range divided by N_{sims}. The χ^2 is for the difference between the mean measurement $\bar{\hat{C}}_\ell^{XY}$ and the input C_ℓ^{XY} given by

$$\chi^2_{\mathrm{recovery}} = \sum_{\ell\ell'} d_\ell^{\mathrm{recovery}} \left(\frac{1}{N_{\mathrm{sims}}}\mathrm{Cov}\right)^{-1}_{\ell\ell'} d_{\ell'}^{\mathrm{recovery}}, \tag{56}$$

where $d_\ell^{\mathrm{recovery}} = \bar{\hat{C}}_\ell^{XY} - C_\ell^{XY}$ and XY is either $\kappa g*$ or gg. The PTE value for this $\chi^2_{\mathrm{recovery}}$ is used to indicate if our measurements on the simulations are statistically consistent with the input. Since we already confirmed the accuracy of the covariance estimation a low PTE value here would indicate a bias in the analysis pipeline. Given the number of simulations N_{sims} this test is sensitive to systematic biases of $\sigma/\sqrt{N_{\mathrm{sims}}}$ where σ is the measurement uncertainty for an individual bin. To complement the PTE test, we also fit the input theory curve to the mean measurement on the sims with a free amplitude A and calculate the difference in amplitude $\Delta A = A - 1$ to report the accuracy of the test in absolute terms.

The resulting PTE values and the difference in fitted amplitude to the input are summarized in Table I. For each data set and statistic, the mean measurement is statistically consistent with the input, showing no significant

TABLE I. Comparison of the mean signal recovered from 480 simulated CMB lensing and galaxy clustering maps using the analysis pipeline compared to the input statistic. The PTE values test if the mean signal is statistically consistent with the input. In addition, the theory curve is fitted to the mean estimated signal with a free amplitude A. Values for $\Delta A = A - 1$ are listed.

Input recovery	*Planck* + CMASS	*Planck* + LOWZ
$C_\ell^{\kappa g*}$	PTE = 0.10 $\Delta A = (0.5 \pm 0.3)\%$	PTE = 0.38 $\Delta A = (0.9 \pm 0.5)\%$
C_ℓ^{gg}	PTE = 0.49 $\Delta A = (0.0 \pm 0.1)\%$	PTE = 0.59 $\Delta A = (0.1 \pm 0.1)\%$

disagreement. We note that for the galaxy autocorrelation simulation recovery test for the LOWZ sample, we originally observed a marginal failure with a PTE value of 0.03. The amplitude of this mean recovered autocorrelation was still only around $(-0.2 \pm 0.1)\%$ low. We do enough PTE tests that a marginal failure can be a statistical outlier. Since this is a simulation-based test we reran the simulation generation to do another independent test. The second run passed without issue (PTE 0.59) therefore we concluded the failure was just statistical and we proceeded with the second run. We discuss the summary of our PTE statistics in Sec. IX.

We conclude that our analysis pipeline accurately estimates the angular power spectra to at least within $\sigma/\sqrt{N_{\rm sims}} \approx 0.05\sigma$ where σ is the measurement uncertainty for individual bins. In absolute terms, the cross-correlation measurement amplitudes are shown to be accurate to at least the percent level and the galaxy autocorrelation measurement amplitudes are accurate to at least the 0.1 percent level. We conservatively account for the level at which we validated the pipeline in the systematic error budget for the E_G estimate with 1% for *Planck* + CMASS and 2% for *Planck* + LOWZ.

Our test covers the entire angular power-spectrum estimation pipeline as described in Sec. VI A and the norm correction of the CMB lensing map as described in Sec. VI B 1.

E. Shot noise test

Measuring the angular autopower spectrum of the galaxy samples with a pseudo-Cl estimator relies on accurate subtraction of the shot noise as described in Eq. (44). We use an analytic estimate of the shot noise that accounts for the variance of the weights applied to the galaxy samples [see Eq. (12)]. We can check our analytic estimate of the shot noise by comparing it to a measurement from data. To estimate the shot noise from data we split the galaxy sample into two equal parts and construct the galaxy overdensity for both. Each overdensity map δ_i contains the cosmological signal δ as well as shot noise n_i. The shot noise in each map is independent since the two subsamples of the data are distinct. By taking the difference between the two maps, the cosmological contribution cancels and we are left with only shot noise in the map when taking the autocorrelation [101],

$$\langle|\delta_1 - \delta_2|^2\rangle = \langle|n_1|^2\rangle + \langle|n_2|^2\rangle. \tag{57}$$

Each sample has half the number density of the full galaxy sample, so the total shot noise is four times higher than our full galaxy sample. We can therefore get an observational estimate of the shot noise contained in the sample by calculating the autocorrelation of the difference map (without subtracting any shot noise) and dividing it by 4. Since the map contains shot noise and no signal this measurement is not affected by a pixel window function and we therefore do not apply a correction for it. To compare our analytic estimate of the coupled shot noise with the measurement from data we need to apply the inverse of the binned coupling matrix as done for the theory spectra [see Eq. (45)]

$$N_\ell^{\rm shot} = \sum_{\ell'} [\mathcal{M}^{-1}]_{\ell\ell'} \tilde{N}^{\rm shot}. \tag{58}$$

In Fig. 5 we compare the analytic estimates to the estimates from data for both the CMASS and LOWZ samples. On large scales the statistical uncertainty from the observational estimate is large but especially on small scales we see that it matches our analytic estimate very well visually. To quantify this we divide the difference map estimate by the analytic estimate. In the range $757 \leq \ell < 2024$ the mean and standard deviation of the ratios for CMASS are

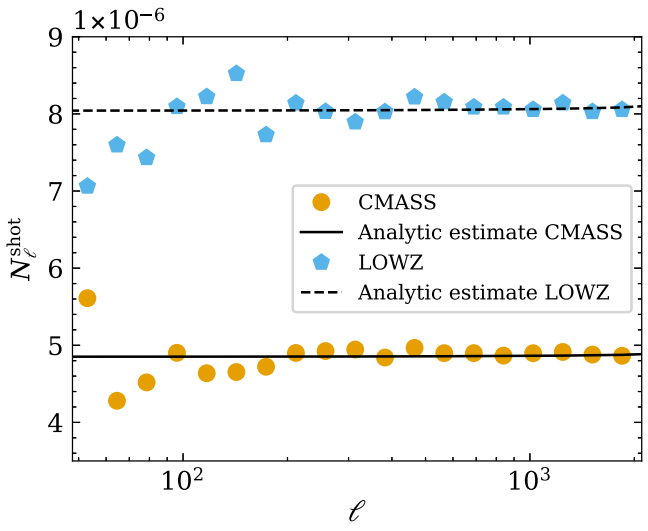

FIG. 5. Comparing our analytic estimate of the shot noise in the galaxy sample [solid black line for CMASS, dashed black line for LOWZ] with an estimate from data for both CMASS [orange circles] and LOWZ (blue pentagons). The data estimate has large statistical uncertainty on large scales and gets more accurate for small scales (high ℓ). The analytic estimate matches the constraint from the data well.

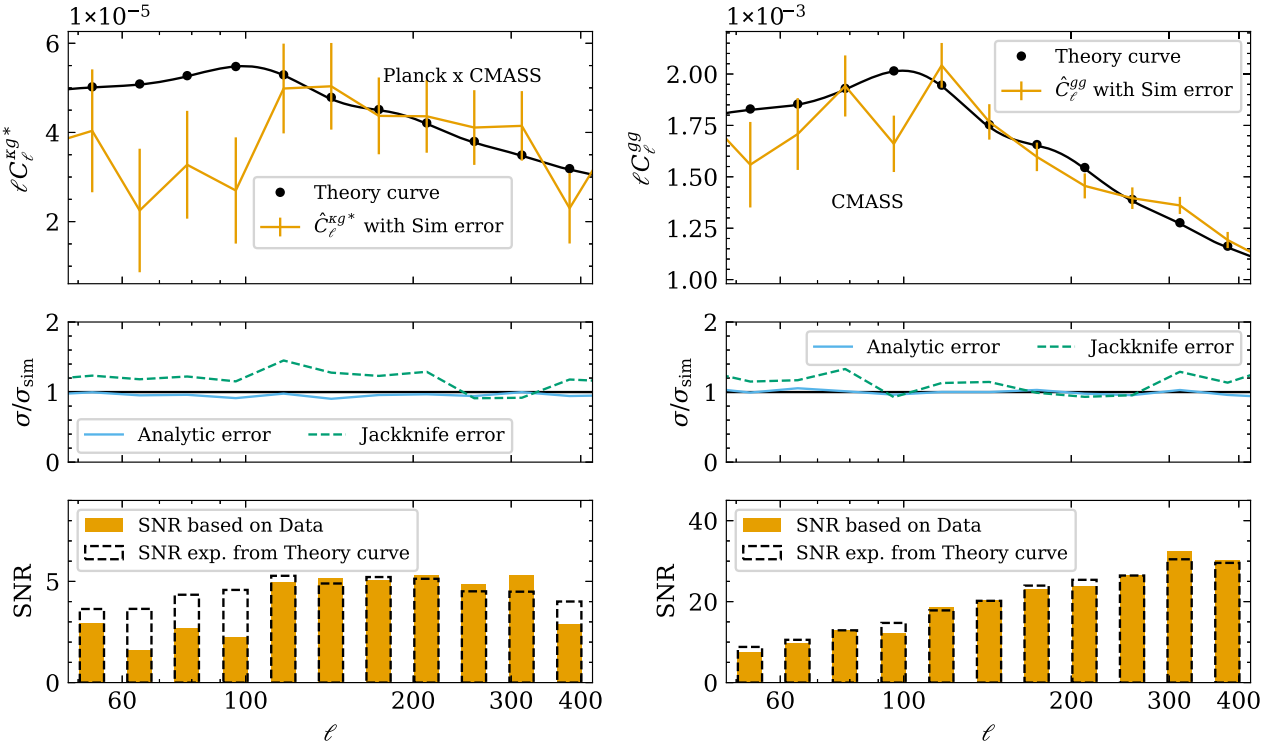

FIG. 6. Measurement of (Left) the angular cross-power spectrum, $C_\ell^{\kappa g*}$, for the CMB lensing map based on *Planck* PR4 and the reweighted CMASS galaxy sample and (Right) the angular auto-CMASS galaxy power spectrum. (Top) The angular power spectrum with simulation-based errors as well as our theory curve. (Middle) Comparison of the analytic (solid blue line) and jackknife error estimates (dashed green) to the baseline marginalized errors from the simulations. (Lower) The signal-to-noise ratio, calculated as the ratio of measurement and measurement uncertainty is shown (filled orange). For reference, the ratio of the theory expectation and the measurement uncertainty is also shown [dashed black]. For *Planck* + LOWZ see Appendix C.

1.003 ± 0.004 and for LOWZ are 1.001 ± 0.005. Therefore the analytic estimate agrees with the difference map estimate well within a percent on small scales and the two are consistent with each other within the statistical scatter of the difference map estimate.

We note that this shows that Eq. (12) correctly accounts for the variance in weights across the sky. A naive approach of assuming equal weights would simplify the shot noise estimate to $n_{\rm eff,naive} = \frac{N_{\rm gal}}{4\pi f_{\rm sky,gal}}$ with $N_{\rm gal}$ the number count of galaxies in the sample. However, SDSS has significant variance in the weights. Using the naive formula would underestimate the shot noise for the CMASS sample by around 21% and for the LOWZ sample by around 7% (both averaged over $757 \leq \ell < 2024$).

Overall this test gives credibility to our autocorrelation result as the subtraction of the shot noise is a crucial step and we have shown that the analytic estimate fits the data accurately. The shot noise estimate is also important for getting an accurate estimate of the analytic covariance for both cross- and autocorrelation.

F. Angular power spectrum measurements

In Fig. 6 we present the angular cross-correlation measurement $\hat{C}_\ell^{\kappa g*}$ for the *Planck* CMB lensing map and the CMASS galaxy sample, based on our pipeline and covariance estimate and, for reference only, the theory curve based on the fiducial cosmology and a fit of the galaxy bias to the amplitude of the galaxy autocorrelation measurement. We also present the galaxy autocorrelation of the CMASS galaxy sample. The same plot for *Planck* + LOWZ can be found in Appendix C.

Visually the measurements are consistent with the shape of the theory curves. For the cross-correlation between *Planck* and CMASS, we recover the known result in the literature that the measurement tends to be low compared to theory expectation on large scales ($\ell < 100$) although to a less significant degree than reported in e.g. [34].

In the bottom panels of the angular power spectrum measurement plots (Fig. 6 and Appendix C) the signal-to-noise ratios (SNR) are shown for each case. We also show the ratio of the theory curves with the measurement which can be more directly compared with the constraints of other surveys. For the log-spaced binning used in this work the expected SNR of the cross-correlation, which carries much larger uncertainty than the autocorrelation, is roughly constant between bins. Therefore, the uncertainties on the E_G estimates between bins are comparable, making them simpler to compare and combine.

G. β measurement from RSD analysis

We redo the RSD analysis on the original BOSS DR12 data vectors from Gil-Marín *et al.* [61] to measure β for CMASS and LOWZ. We do this to compare the constraints with updated theoretical modeling to the original results and also to be able to account for the cross-covariance between $f\sigma_8$ and $b\sigma_8$ when calculating the derived parameter β which is not included in the published results. The data vectors from Gil-Marín *et al.* [61] are combined measurements of the power spectra for the full samples obtained by sky-averaging the NGC and SGC measurements, along with window matrices connecting the translation-invariant theoretical three-dimensional power spectrum P_ℓ of the galaxies to that of the windowed observed field,

$$P_\ell^{\rm obs}(k_i) = \sum_{\ell'=0,2} W_{ij}^{\ell\ell'} P_{\ell'}(k_j). \tag{59}$$

These three-dimensional measurements were computed in Gil-Marín *et al.* [61] assuming a fiducial cosmology with $\Omega_m^{\rm fid} = 0.31$.[10]

The ratio β compares the amplitude of the real and redshift-space clustering signal in galaxies, which become contaminated by the nonlinearities of structure formation towards smaller scales (higher k). As in the official BOSS results [61] we fit the linear theory clustering-weighted bias and growth rate, $b\sigma_8(z)$ and $f(z)\sigma_8(z)$ ($\beta = f/b$), assuming a fiducial linear power spectrum $P_{\rm lin}(k)$ and clustering amplitude $\sigma_8(z)$ given by the best-fit *Planck* value [5]. The latter assumptions are necessary to compute the contributions of next-to-leading-order nonlinearities; we note that the signal-to-noise of the BOSS power spectra are not sufficient to break the $b-\sigma_8$ and $f-\sigma_8$ degeneracies, such that freeing the value of σ_8 in fits can lead to undesirable parameter-volume effects [102]. Finally, as in the official BOSS fits, and to be agnostic to anisotropies due to deviations of the expansion history from the fiducial cosmology, we perform our fits with the Alcock-Paczynski parameters $\alpha_{\parallel,\perp}$ as free parameters.

In order to marginalize over the nonlinearities of structure formation in our analysis we adopt the Lagrangian perturbation theory (LPT) model of Chen *et al.* [103,104], implemented in the publicly available code VELOCILEPTORS.[11] This model contains both higher-order nonlinear bias terms as well as effective-theory corrections, e.g. isotropic and anisotropic $k^2 P_{\rm lin}(k)$ counterterms and stochastic noise terms including a scale-dependent $k^2\mu^2$ contribution. It represents a more careful treatment of nonlinear galaxy clustering and redshift-space distortions compared to the perturbation theory models used in the official BOSS results, particularly in the consistent effective-theory treatment of small-scale physics beyond the reach of perturbation theory. As a result, our RSD constraints are slightly less tight than those from the official analysis, though we note that this is not a weakness of the measurement but rather a result of including the full theoretical uncertainty in our model to this order. We use the publicly available pipeline developed for Chen *et al.* [44] and adopt the parameters and priors given in Table 2 of that work.

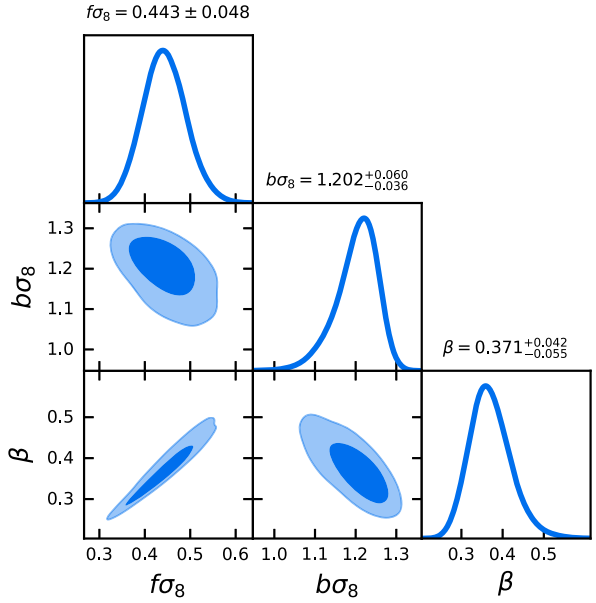

FIG. 7. RSD fit to the clustering signal of the CMASS data set. We show the corner-plot of the fitting parameters $f\sigma_8$, $b\sigma_8$ as well as the derived parameter $\beta = f\sigma_8/(b\sigma_8)$ that we use as part of our E_G estimate. For each parameter, we also show the mean value and 68.27% error. For LOWZ see Appendix C.

We use $k_{\rm max} = 0.2\ h/{\rm Mpc}$. We show the cornerplot for the constraints on $f\sigma_8, b\sigma_8$ and the derived parameter $\beta = f\sigma_8/(b\sigma_8)$ for CMASS in Fig. 7 using GetDist [105].[12] We also report the plot of the weighted mean value and one sigma errors for each parameter. For the same plot for the LOWZ sample see Appendix C.

For the CMASS sample we find $f\sigma_8 = 0.443 \pm 0.048$, $b\sigma_8 = 1.202^{+0.060}_{-0.0336}$. These are in good agreement with the values reported in Gil-Marín *et al.* [61] where for the original perturbation theory models of the original BOSS results and a slightly higher $k_{\rm max} = 0.24\ h/{\rm Mpc}$ they found $f\sigma_8 = 0.444 \pm 0.038$, $b\sigma_8 = 1.222 \pm 0.021$ for the CMASS sample.

For the LOWZ sample we find $f\sigma_8 = 0.426^{+0.071}_{-0.082}$, $b\sigma_8 = 1.271^{+0.080}_{-0.049}$ which is also in good agreement with the values from Gil-Marín *et al.* [61] which are $f\sigma_8 = 0.394 \pm 0.052$, $b\sigma_8 = 1.281 \pm 0.035$ for the LOWZ sample.

For the derived parameter β we find $\beta = 0.371^{+0.042}_{-0.055}$ for the CMASS sample and $\beta = 0.338^{+0.058}_{-0.081}$ for the LOWZ

[10]We thank Héctor Gil-Marín for providing us with the official BOSS measurements including window matrices.

[11]https://github.com/sfschen/velocileptors.

[12]https://github.com/cmbant/getdist.

sample. While less directly comparable our β results are also broadly consistent within one sigma with other RSD analyses of the SDSS BOSS data; Chuang *et al.* [62] do not use the optimal RSD weights and a slightly less restrictive upper redshift cut on CMASS finding $\beta = 0.435 \pm 0.070$ for CMASS and $\beta = 0.301 \pm 0.066$ for LOWZ. For the earlier BOSS DR11 data [34] report $\beta = 0.368 \pm 0.046$ for the CMASS sample. For our E_G measurement we use the full probability density function (PDF) of the β measurement as shown in the corner plots the process for which we describe in detail next.

VII. NON-GAUSSIAN UNCERTAINTY OF E_G MEASUREMENT

To estimate the significance of any E_G measurement, we need to be able to correctly calculate the PDF for the E_G statistic given the measurement covariances of the three constituent observables; $C_\ell^{\kappa g*}, C_\ell^{gg}$ and β. Even when these observables are Gaussian, since E_G involves a ratio of them, its uncertainty can be significantly skewed, and non-Gaussian [106]. In using the data across multiple bins, one also has to carefully consider how to combine them and the potential correlations between band powers.

We investigate the PDF of our estimator without assuming Gaussianity, describe how we report our constraints, and compare to assuming Gaussianity in Sec. VII A. Additionally, we build PTE tests to test our measurement for consistency with the ΛCDM-GR prediction (Sec. VII B) and for consistency with scale independence (Sec. VII C).

A. PDF of our $\hat{E}_G$ estimator

Let us first consider the ratio of the measured angular power spectra $\hat{C}_\ell^{\kappa g}$ and $\hat{C}_\ell^{gg}$. We include the Γ_ℓ factor and define $R \equiv \Gamma_\ell \hat{C}_\ell^{\kappa g*}/\hat{C}_\ell^{gg}$. The probability density for this ratio is given by the ratio distribution [107],

$$p_R = \int \mathrm{d}y |y| p_{XY}(Ry\hat{C}_\ell^{gg}, y\hat{C}_\ell^{gg}), \tag{60}$$

where R is the ratio we want to find and p_{XY} is the PDF of our observed angular power spectra which are correlated Gaussian random variables with our measurement as mean and the covariance of our observation,

$$XY \sim \mathcal{N}\left(\mu = \begin{pmatrix} \Gamma_\ell \hat{C}_\ell^{\kappa g*} \\ \hat{C}_\ell^{gg} \end{pmatrix}, \right.$$
$$\left. \mathrm{Cov}_{\ell\ell'} = \begin{bmatrix} \Gamma_\ell\Gamma_{\ell'}\mathrm{Cov}(C_\ell^{\kappa g*}, C_{\ell'}^{\kappa g*}), \Gamma_{\ell'}\mathrm{Cov}(C_\ell^{\kappa g*}, C_{\ell'}^{gg}) \\ \Gamma_\ell \mathrm{Cov}(C_\ell^{gg}, C_{\ell'}^{\kappa g*}), \mathrm{Cov}(C_\ell^{gg}, C_{\ell'}^{gg}) \end{bmatrix} \right). \tag{61}$$

With this approach, we can find the PDF for the ratio both for each bin individually (R_ℓ) and one overall constraint R over our cosmological range in ℓ. For the latter, we implicitly assume that E_G is scale invariant as predicted for GR. We also separately evaluate this prediction by testing for evidence of scale dependence (see Sec. VII C).

Given the ratio of the angular power spectra, we can then complete our calculation of our estimator [$\hat{E}_G$, Eq. (34)] by dividing R by β. To obtain the probability density function for our estimator $\hat{E}_G$ given the data we calculate the ratio distribution,

$$p_{\hat{E}_G}(\hat{E}_G|\hat{C}_\ell^{\kappa g}, \hat{C}_\ell^{gg}, \beta) = \int \mathrm{d}\beta' |\beta'| p_R(\hat{E}_G \cdot \beta') p_\beta(\beta'), \tag{62}$$

where p_β is the PDF of the β measurement. Here, and throughout the analysis, we assume no correlation between β and the angular power spectra. This correlation is expected to be negligible since the β measurement is based on the full 3D clustering of the galaxies while the angular power spectra only use a single projection onto the sky [34].

Crucially, this approach does not assume Gaussianity for multiplying or dividing our measurements. This ensures that we capture the asymmetric PDF for large uncertainties where the Gaussian approximation breaks down. We can also perform this approach per bin to measure the statistic for each bin $\hat{E}_G^\ell$.

Figure 8 shows the E_G PDF distribution for the fiducial cosmology and observational covariances when combining the correlation and β measurements across our cosmological range. For this test, β is treated simply as a Gaussian with a standard deviation equal to the mean of the upper and lower 68.27% errors.

We also show the distribution of $\hat{E}_G$ estimates for a large number of samples (100,000) from the fiducial cosmology and our observational covariance, showing that the PDF describes the distribution accurately. Note our PDF weights each binned measurement by their relative uncertainty but for the sampled points we take the mean of the binned estimates assuming equal weights since it is unclear how to estimate and define the relative uncertainty for each bin without assuming Gaussianity. Since the uncertainty of each of the binned measurements, for our specific log-spaced binning scheme, is of comparable size, the difference is negligible. We also confirmed that the estimated PDF for each of the individual bins accurately matches the distribution of a large number of samples.

The PDF for the $\hat{E}_G$ estimator is significantly skewed. The mean of the distribution is higher than the GR input value, the mode (peak) of the distribution is lower than the input value and the median of the distribution recovers the input value accurately. We quantify this for the example of a *Planck* + CMASS-like covariance at our fiducial cosmology shown in the top panel of Fig. 8; compared to the GR input value of $E_G = 0.401$, the mean is $\Delta E_G = 0.008$ (2.0%) high and the mode $\Delta E_G = -0.014$ (3.6%) low, however the median matches well with $|\Delta E_G| < 0.0005$

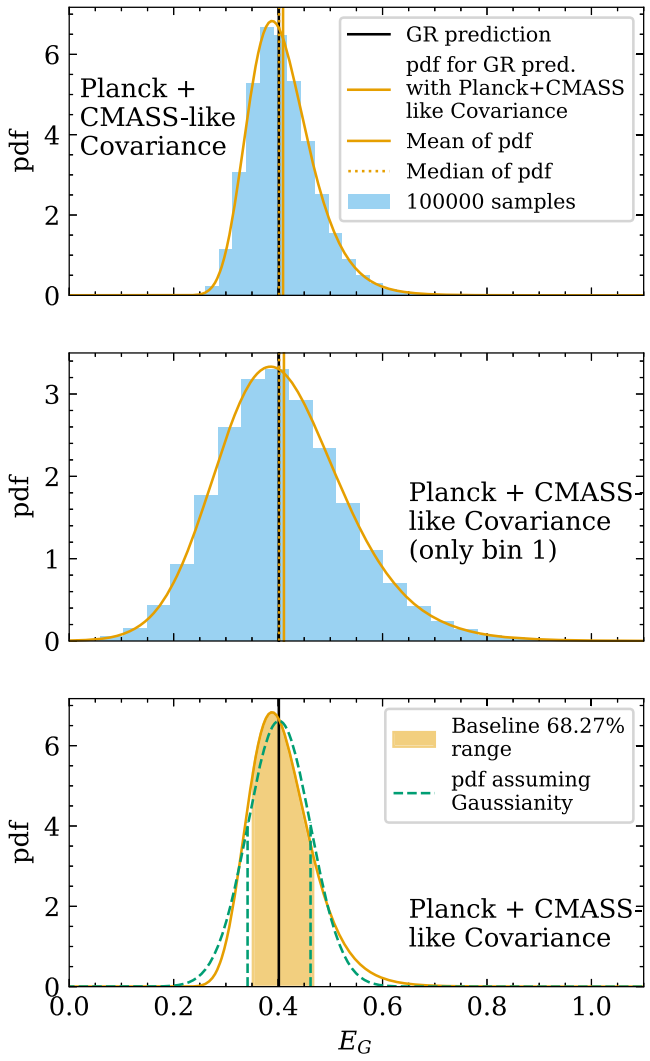

FIG. 8. Characterization of the non-Gaussian PDF for the $\hat{E}_G$ estimator for the fiducial cosmology and *Planck* + CMASS data covariance (orange curve). The fiducial value for E_G [vertical black line] is shown along with the mean of the PDF (vertical orange full line) and median (vertical orange dotted line). The distribution of 100,000 samples from the theory model of the data vector and covariance are also shown (blue shading). The top panel shows the case for combined measurement using all bins and the middle panel the case of only using the first bin. In the bottom panel, we show for comparison the result when assuming Gaussianity (green dashed) for the combined measurement.

(<0.2%). Given this, we use the median of the distribution of the PDF as the best estimate in the analysis. Consistently we use the middle 68.27% of the area under the PDF as the one sigma range.

To facilitate comparison with other results in the literature, which assumed Gaussian errors, we can compare our PDF with the result one would get when assuming Gaussianity and performing Gaussian error propagation. For reference, the Gaussian error propagation for the estimator is summarized in Appendix D. To get the covariance $\mathrm{Cov}(\hat{E}_G^\ell, \hat{E}_G^{\ell'})$ from the fractional covariance we get from the error propagation we scale by our estimate $\hat{E}_G$.[13] We can then under the assumption of Gaussianity describe the PDF for $\hat{E}_G$ as

$$p^{\mathrm{Gaussian}}_{\hat{E}_G}(\hat{E}_G) \propto \exp\left(\frac{-\chi^2_{\mathrm{Gaus}}}{2}\right) \tag{63}$$

with

$$\chi^2_{\mathrm{Gaus}} = \sum_{\ell,\ell'} (\hat{E}_G^\ell - \hat{E}_G)\mathrm{Cov}^{-1}(\hat{E}_G^\ell, \hat{E}_G^{\ell'})(\hat{E}_G^{\ell'} - \hat{E}_G), \tag{64}$$

where we apply Hartlap corrections to the inverse covariance as done throughout our analysis. This $p^{\mathrm{Gaussian}}_{\hat{E}_G}$ is equivalent to the approach in Pullen *et al.* [34] and Zhang *et al.* [35]. The best estimate given in their approaches is the peak of the distribution where χ^2_{Gaus} is minimized and the one sigma errors can be found from the $\chi^2_{\mathrm{Gaus}} = \chi^2_{\mathrm{Gaus,min}} + 1$ surface.

For the fiducial cosmology, this Gaussian approach gives the same best estimate as our PDF but the errors are different. When applying to measured data, the best-fit assuming Gaussianity for a single bin is the same as in our PDF, but the relative weighting of the measurements is different so the marginalized result across scales is not the same. The best-fit in the Gaussian case is equal to the weighted mean of the bin-wise measurements E_G^ℓ, weighted according to the inverse variance of each bin given by the Gaussian covariance. Since the bin-wise E_G^ℓ is high on average, the marginalized result will also be biased high.

The comparison of the Gaussian and full PDF in Fig. 8 shows that the upper error gets underestimated and the lower error gets overestimated. To estimate comparable uncertainty levels for the skewed distribution we use the central 68.27% confidence range which matches the one sigma error in the Gaussian limit. For the fiducial cosmology and for the *Planck* + CMASS covariance, the Gaussian estimate underestimates the upper error by 12% (0.060 for Gaussian vs 0.068 for full PDF) and overestimates the lower error by 11% (0.060 vs 0.054) compared to our PDF. For the 95.45% (2σ equivalent) confidence range, the difference between Gaussian and non-Gaussian PDF is even more pronounced; the Gaussian estimate of the upper error is 24% low (0.121 vs 0.16) and the lower error is 23% high (0.121 vs 0.098). This is relevant when reporting a low E_G measurement compared to expectation as done in

[13]We use the weighted average over the bins. To find this we first take the mean of the $\hat{E}_G^\ell$ and use it to calculate an initial covariance. Then use this initial covariance to calculate the weighted mean across the bins and use the weighted mean to find the covariance. In practice, the relative weighting makes little difference.

Pullen *et al.* [34] since the underestimated upper error leads to an overestimation of tension with the ΛCDM-GR prediction. The exact difference depends on the measurement since the PDF shape depends on the measurement and the best-fit values for the two approaches are generally not identical. As a simplified example to make the point, we can reduce the $C_\ell^{\kappa g}$ from our theory curve by a factor 0.7 so that according to Gaussian statistics the ΛCDM-GR prediction is outside of the central 99.07% range (equivalent of 2.6σ) matching the value reported in Pullen *et al.* [34], where we tuned the factor manually to get a similar disagreement. Fully accounting for the asymmetric PDF, the best-fit value is the same in this example, but the ΛCDM-GR value is only outside the central 95% range due to the asymmetric errors [based on $\mathrm{PTE_{GR}}$ as defined in Eq. (65); equivalent to 2.0σ for Gaussian statistics], which would only be a marginally significant result. We argue, therefore, that it is crucial to correctly account for the asymmetric PDF of the estimator in any interpretation of the measurement results.

B. PTE test for consistency with GR value

Given the PDF, we can test a measurement against the expectation value from ΛCDM-GR, E_G^{GR}. We calculate the PTE value for consistency with the ΛCDM-GR prediction as the distance from the median in the cumulative distribution given by

$$\mathrm{PTE_{GR}} = 1 - 2|0.5 - P(E_G < E_G^{\mathrm{GR}})|. \tag{65}$$

The theoretical ΛCDM-GR prediction itself carries uncertainty $\sigma_{E_G^{\mathrm{GR}}}$ because it relies on a measurement of $\Omega_{m,0}$ based on the expansion history. We account for the theory uncertainty by convolving our PDF for E_G with the normal distribution $\mathcal{N}(\mu = 0, \sigma^2_{E_G^{\mathrm{GR}}})$ but we will see that this makes very little difference in our case since $\sigma_{E_G^{\mathrm{GR}}}$ is much smaller than the measurement uncertainty for our data.

In Fig. 9 we show the $\mathrm{PTE_{GR}}$ values for $\hat{E}_G$ measured on our 480 simulated data vectors compared to the input E_G^{GR}. We here do not need to marginalize over the uncertainty of E_G^{GR} as we use our fiducial cosmology as an input to our simulations. We find that for all multipole bins combined, and individual bins, a flat distribution within the Poisson uncertainty showing that $\mathrm{PTE_{GR}}$ accurately tests for consistency with GR while accounting for the asymmetric PDF of our measurement.

C. PTE test for consistency with scale independence

Another key prediction for E_G in GR is scale independence. Above we implicitly assumed scale independence when measuring the combined constraint on E_G across multiple bins. To study whether the data are consistent with

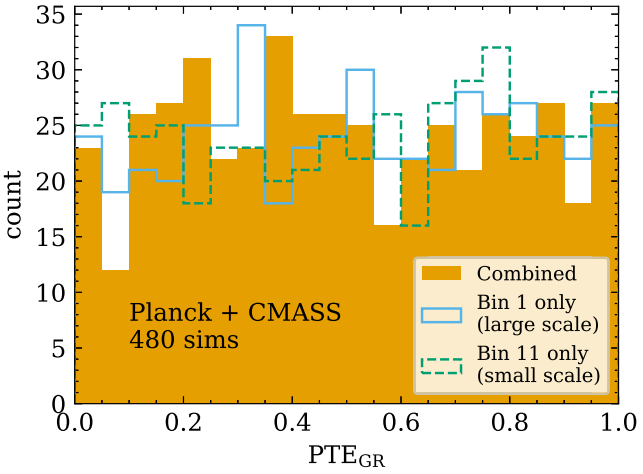

FIG. 9. Characterizing the statistical consistency test between the E_G measurement and the ΛCDM-GR prediction by applying it to simulations. For each of the set of 480 simulated *Planck* + CMASS-like data vectors drawn around the fiducial cosmology, we calculate the $\mathrm{PTE_{GR}}$ for consistency with the GR value. Results are shown when combining all multipole bins in the cosmological range [orange solid] and for two single band powers at the largest ($\ell = 53$) [blue line] and smallest scale ($\ell = 382$) [green dashed] of the cosmological analysis range. The $\mathrm{PTE_{GR}}$ distributions are consistent with the expected flat distribution for a valid test.

the assumption of scale independence we consider a distance measure, building on Miyatake *et al.* [108], defined as[14]

$$\chi^2(E_G) = \sum_{\ell\ell'} d_\ell \mathrm{Cov'}^{-1}_{\ell\ell'} d'_\ell, \tag{66}$$

where for the E_G statistic we have $d_\ell \equiv \Gamma_\ell \hat{C}_\ell^{\kappa g *} - E_G \beta \hat{C}_\ell^{gg}$. The covariance is given by

$$\begin{aligned}\mathrm{Cov'}_{\ell\ell'} \equiv\ & \Gamma_\ell \Gamma_{\ell'} \mathrm{Cov}(C_\ell^{\kappa g *}, C_{\ell'}^{\kappa g *}) \\ & - E_G \Gamma_\ell \beta [\mathrm{Cov}(C_\ell^{\kappa g *}, C_{\ell'}^{gg}) + \mathrm{Cov}(C_\ell^{gg}, C_{\ell'}^{\kappa g *})] \\ & + (E_G)^2 [\beta^2 \mathrm{Cov}(C_\ell^{gg}, C_{\ell'}^{gg}) + \hat{C}_\ell^{gg} \hat{C}_{\ell'}^{gg} \sigma_\beta^2].\end{aligned} \tag{67}$$

As before, we neglect the correlation between β and the angular power spectra and assume the β measurement is Gaussian. Note that the covariance Cov' depends on E_G.

The distribution of minima for this distance measure closely follows a χ^2 function with degrees of freedom equal to the number of bins minus one (since we find the minimum). Figure 10 shows the χ^2 distribution for the case of *Planck* + CMASS over the full cosmological range

[14]We note that the E_G value that minimizes this χ^2 distance measure is a biased estimator of the E_G statistic. We use this distance measure only to test for scale dependence, not to get an unbiased estimate of the ratio.

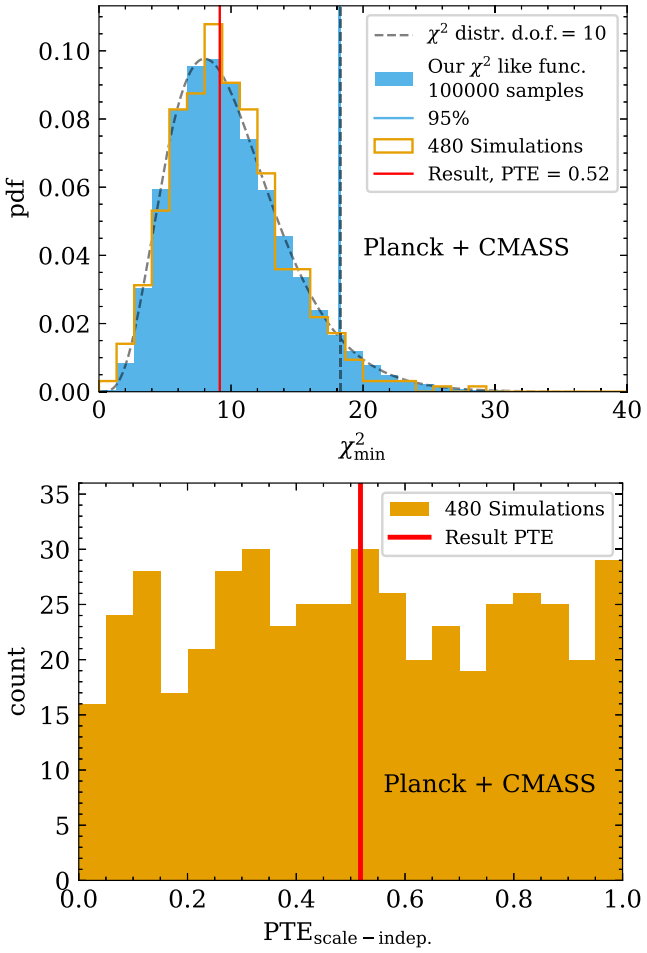

FIG. 10. Characterizing the statistical consistency test of the E_G measurement for scale independence for the case of *Planck* + CMASS. (Upper) The minimum value for the χ^2 distance measure. The χ^2 distribution for the degrees of freedom (gray dashed curve) is shown to be consistent with the results from the 480 simulations [orange] and the distribution for 100,000 samples around the fiducial cosmology and baseline (simulation-derived) covariance. The value of significance for the χ^2 distribution (gray dashed vertical) and for the 100,000 samples (95% of sampled distance measures fall below) (blue full vertical) are shown to be in good agreement. (Lower) The resulting $\rm PTE_{scale-indep}$ values for the 480 simulations. The $\rm PTE_{scale-indep}$ distribution is consistent with the expected flat distribution for a valid test. Our result for the *Planck* + CMASS $\hat{E}_G$ estimate, $\rm PTE_{scale-indep} = 0.52$, obtained in Sec. VIII is shown as a vertical red line. For *Planck* + LOWZ see Appendix C.

consisting of 11 bins spanning $48 \leq \ell < 420$. For comparison, we show a histogram of $\chi^2_{\rm min}$ values for 100,000 samples around the fiducial cosmology using the baseline (simulation-derived) covariance. Since the covariance is exact for these samples we do not apply Hartlap corrections. The plot demonstrates that the distribution of the samples and the line of significance, for which 95% of the $\chi^2_{\rm min}$ samples are below, are consistent with the analytical distribution.

The figure also shows the histogram of the $\chi^2_{\rm min}$ for the 480 simulated data vectors. In the case of the simulated data vectors as well as the observed data vector, the baseline covariance carries uncertainty. For an unbiased estimate of the inverse covariance used in the χ^2 calculation, we apply Hartlap corrections. We only have 480 simulations, therefore the Poisson scatter in the histogram is larger but overall the distribution matches well with the χ^2 distribution as well.

We use a Pearson's chi-square test to test our observation for consistency with scale independence, given as

$$\rm PTE_{scale-indep} = P_{\chi^2}(\chi^2 > \chi^2_{min}), \tag{68}$$

the probability of getting a higher χ^2 than for our measurement as based on the analytic χ^2 PDF for our degrees of freedom. A significant result ($\rm PTE_{scale-indep} < 0.05$) would indicate that the data is not well described by a constant ratio and therefore would be evidence for scale dependence. We however note a significant result could also be caused by underestimating the covariance of the measurement or from underlying scale-dependent systematic biases.

To test our approach we calculated the $\rm PTE_{scale-indep}$ value for the 480 simulated data vectors. As shown in Fig. 10, the distribution of PTE values is consistent with a flat distribution within the Poisson uncertainties of the counts indicating that our scale dependence test works correctly.

VIII. RESULTS FOR E_G WITH *PLANCK* AND SDSS BOSS

A. Baseline results for the E_G statistic

We now apply the E_G estimator presented in this work to the *Planck* PR4 CMB lensing data and SDSS BOSS galaxy samples using our measured angular power spectra and RSD parameter β.

Figure 11 shows the bin-wise estimate E_G^ℓ for *Planck* PR4 + CMASS and for *Planck* PR4 + LOWZ for the cosmological range in scales, with the ΛCDM-GR expectation value as a reference. For each, three different estimates of the 68.27% confidence ranges (the equivalent of one sigma error for Gaussian statistics) are shown; the baseline simulation-based covariance as well as the analytic and jackknife covariances. For individual bins, the estimator gives the same best-fit value independent of the error. Similar to the individual angular power spectra the different error techniques are also consistent with each other within expectation for the bin-wise E_G^ℓ constraints. Simulation-based errors and analytic errors match closely and the jackknife errors are a systematic overestimate of 20–30% and larger statistical variation consistent with expectations (see Sec. VI C).

The statistical test for scale-independence (Sec. VII C) evaluates if the measurement is statistically consistent with

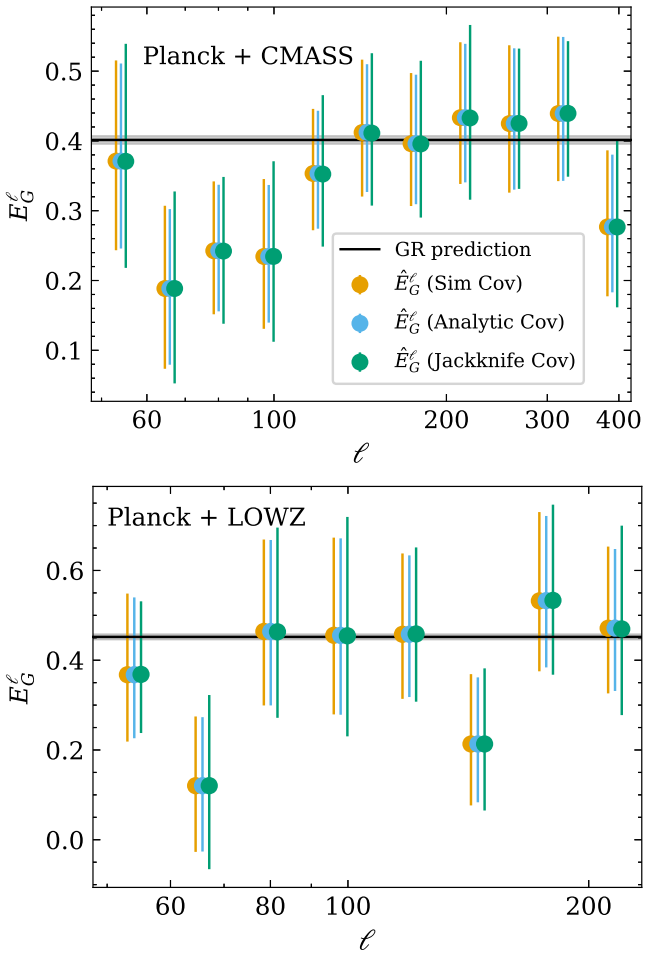

FIG. 11. Measurements of $\hat{E}_G^\ell$ for (Upper) *Planck* + CMASS and (Lower) *Planck* + LOWZ for each multipole bin. The median value is indicated as dots and the errors show 68.27% confidence ranges. The marginalized errors for all three of the covariance estimate approaches are given; simulations (black), analytic (blue), and jackknife (green) with the latter two shifted to the right for visibility. The GR prediction is shown (solid line) for reference with the one sigma uncertainty (gray shaded).

a constant across the cosmological range for each of the measurements. Scale independence is a direct prediction of GR that is violated in many models of modified gravity. Finding the data inconsistent with a constant could indicate tension with GR, but could also be caused by scale-dependent systematic effects. This test is insensitive to an overall normalization systematic in the analysis. We find that both the *Planck* + CMASS and *Planck* + LOWZ measurements are statistically consistent with a constant across the scales in the cosmological range. We find $\mathrm{PTE}_{\mathrm{scale-indep}} = 0.52$ for *Planck* + CMASS and also show the result in the context of the results on the set of simulations in Fig. 10. For *Planck* + LOWZ we find $\mathrm{PTE}_{\mathrm{scale-indep}} = 0.37$. Both cases indicate that the bin-wise measurements are within measurement uncertainty fully consistent with scale independence as predicted by GR. Since there is no evidence for scale dependence one can combine the measurements across the cosmological range to find an overall constraint on E_G.

It is worth noting that while GR predicts scale-independent growth in linear theory, we expect structure formation to introduce scale dependence on small scales due to nonlinearities. The transition between linear and nonlinear growth can be rigorously studied within perturbation theory, which should be valid over the range of scales and observables we fit [44]. In particular, our cumulative constraint on E_G in either sample is roughly 20%, which is wider than the expected correction due to nonlinearities, some of which further cancels in the ratio of the auto and cross-correlation, on the scales ($k < 0.2$ Mpc^{-1}) that we use in this work. However, we caution that this systematic will be more significant for future, more constraining data. These nonlinearities can also be taken into account by marginalizing over the effects of nonlinear bias as computed in perturbation theory; we leave such considerations to future work.

In Fig. 12 the overall constraints on $\hat{E}_G$ for *Planck* + CMASS and *Planck* + LOWZ are shown, giving the full PDF, along with the median value and the central 68.27% confidence range for the $\hat{E}_G$ estimator combining the constraints across the scales in the cosmological ranges. The figure also shows a vertical line indicating the ΛCDM-GR prediction for our fiducial cosmology with the one sigma uncertainty shown as a gray band.

For *Planck* + CMASS we find $E_G^{\mathrm{Planck+CMASS}} = 0.36^{+0.06}_{-0.05}(\mathrm{stat})$ where the statistical error is showing the 68.27% confidence range. The 95.45% confidence ranges are given by $^{+0.13}_{-0.10}$. At the effective redshift of the measurement the ΛCDM-GR prediction based on our fiducial cosmology is $E_G^{\mathrm{GR}}(z = 0.555) = 0.401 \pm 0.005$.

For *Planck* + LOWZ we find $\hat{E}_G^{\mathrm{Planck+LOWZ}} = 0.40^{+0.11}_{-0.09}(\mathrm{stat})$ where the statistical error is showing the 68.27% confidence interval. The 95.45% confidence ranges are given by $^{+0.26}_{-0.16}$. Note that the 95.45% confidence range is more than twice for the upper error and less than twice for the lower error compared to the 68.27% uncertainties due to the non-Gaussian shape of the uncertainty. At the effective redshift of the measurement, the ΛCDM-GR prediction based on the fiducial cosmology is $E_G^{\mathrm{GR}}(z = 0.316) = 0.452 \pm 0.005$.

For both the *Planck* + CMASS and *Planck* + LOWZ cases the ΛCDM-GR prediction lies within the 68.27% confidence range of the measurement. This agreement is quantified with the statistical test for consistency with GR (Sec. VII B). The test accounts for the statistical uncertainty of the measurement and the uncertainty of the ΛCDM-GR prediction. We find $\mathrm{PTE}_{\mathrm{GR}} = 0.46$ for *Planck* + CMASS and $\mathrm{PTE}_{\mathrm{GR}} = 0.64$ for *Planck* + LOWZ. Therefore, we find the marginalized E_G measurements in good agreement with the ΛCDM-GR prediction.

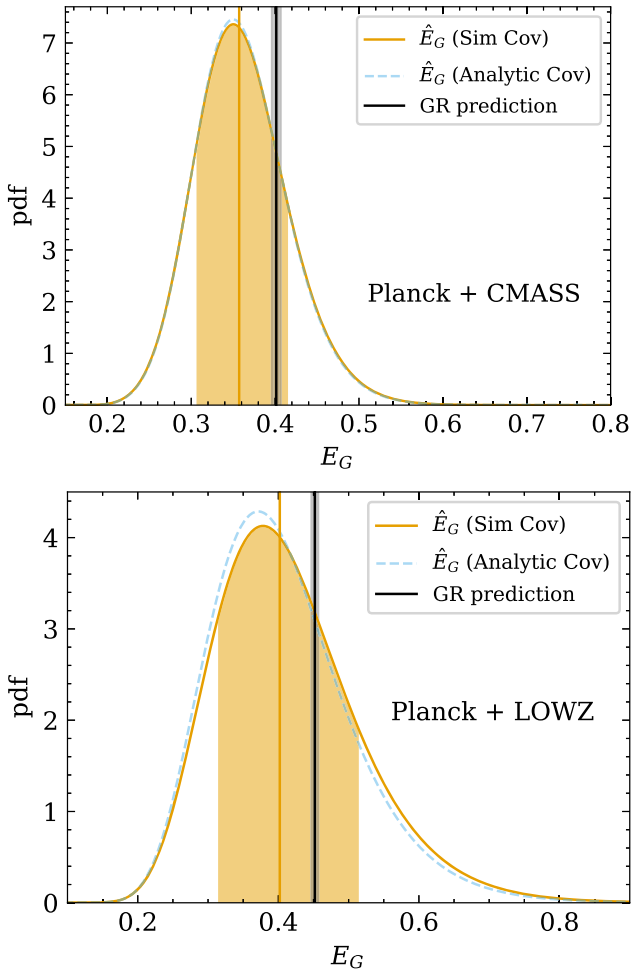

FIG. 12. The PDF of the measurement $\hat{E}_G$ marginalized over the cosmological range of scales for *Planck* + CMASS (upper panel) and *Planck* + LOWZ (lower panel). Shown is the PDF of the $\hat{E}_G$ measurement using the baseline simulation-based covariance (solid orange line) with the median [vertical orange] and central 68.27% confidence range (shaded orange) highlighted. The resulting PDF for the measurement when using the analytic covariance is also shown (dashed blue line). The ΛCDM-GR prediction is shown as a vertical black line with the one sigma uncertainty shaded. Our measurement is in good agreement with the GR prediction with $\mathrm{PTE_{GR}} = 0.46$ for *Planck* + CMASS and $\mathrm{PTE_{GR}} = 0.64$ for *Planck* + LOWZ.

The uncertainty of the theory prediction is largely negligible in comparison to the measurement uncertainty. Not accounting for the theory uncertainty by convolving our measurement PDF with the theory uncertainty before calculating the $\mathrm{PTE_{GR}}$ gives 0.44 and 0.64 for *Planck* + CMASS and *Planck* + LOWZ, respectively, which is only a marginal difference.

On top of the statistical uncertainty, the analysis can also be affected by systematic errors. Throughout the analysis, conservative estimates of the systematic error budget for the estimate are given. These include approximations made to derive the estimator, bias evolution, magnification bias (Sec. V C), and the level to which the pipeline is validated (Fig. 4). Overall, combining the estimates in quadrature and rounding up we have systematic error budgets of 2% for *Planck* + CMASS and 3% for *Planck* + LOWZ, well within the statistical uncertainties of the measurements, which are approximately $^{+17\%}_{-14\%}$ and $^{+28\%}_{-23\%}$, respectively. The subdominant systematic uncertainty is not accounted for in the test for consistency with the ΛCDM-GR value, since it has unknown statistical properties and is an upper estimate that can artificially inflate the level of agreement with GR. We can, for reference, conservatively assume the systematic error as one sigma of a Gaussian and convolve it with the statistical PDF as done for the ΛCDM-GR prediction uncertainty. The sigmas from the systematic error and theory error are added in quadrature. This widens the PDF, increasing the $\mathrm{PTE_{GR}}$ values to 0.46 for *Planck* + CMASS and 0.65 for *Planck* + LOWZ, an increase of 0.5 (hidden below the rounding) and 1 percentage point respectively. Even without accounting for the systematic uncertainty, the result agrees with ΛCDM-GR and a conservative upper estimate of the total systematic error only increases the level of the agreement and we find the difference is only marginal justifying neglecting the systematic error budget for our statistical test.

Here, we disclose how data was handled in this analysis to avoid biases in the statistical tests from fine-tuning the analysis choices. During the development of the angular power spectrum estimation, simulations at the fiducial cosmology and a preliminary galaxy bias of 2.0 were used, first Gaussian simulations and then the simulations discussed in Sec. VI B. The $\hat{E}_G$ estimator, all the corrections applied, the non-Gaussian error estimation, and all analysis choices including for example the fiducial cosmological range in scales and the use of the simulation-based covariance were finalized before revealing the results on data. After the analysis approach was frozen, we refined the galaxy bias, to make the covariance matrix more accurate, as described in Sec. VI B 2. We then calculated the PDFs for E_G and E_G^ℓ on data and reported the result. There were no significant changes to the results after the initial calculation. After the results were finalized we added the theory curves to the plots of the angular power spectra for comparison (e.g. Fig. 6) and investigated the sensitivity of our results to a range of analysis choices (Sec. VIII B). The data used in this analysis has been public and used in numerous analyses before this work, therefore the general insight that on large scales the cross-correlation between *Planck* CMB lensing and SDSS BOSS tends low is well-known. Furthermore, the E_G statistic has been estimated for previous iterations of the *Planck* CMB lensing and SDSS BOSS datasets with different approaches which were finalized and published before this analysis. After finalizing our results we compared them to previous analyses in detail (Sec. VIII C).

A significant value, defined in this work as $\mathrm{PTE} < 0.05$ will naturally occur on average 5% of the time even when

TABLE II. Summary of the results for the $\hat{E}_G$ estimator presented in this work applied to *Planck* PR4 and SDSS BOSS DR12 CMASS and LOWZ data. Shown is the effective redshift of the measurement, $z_{\rm eff}$, the ΛCDM-GR prediction at the effective redshift based on our fiducial cosmology, $E_G^{\rm GR}$, the measurement result, $\hat{E}_G$, and the PTE values for consistency of the measurement with scale independence as well as for consistency with the ΛCDM-GR prediction. Measurements for both datasets are statistically consistent with scale independence as predicted by GR and statistically consistent with the value predicted by ΛCDM-GR for our fiducial cosmology, they match within the 68.27% confidence range.

	$z_{\rm eff}$	GR prediction $E_G^{\rm GR}$	Measurement $\hat{E}_G$	$\text{PTE}_{\text{scale-indep}}$	PTE_{GR}
Planck PR4 + CMASS	0.555	0.401 ± 0.005	$0.36^{+0.06}_{-0.05}(\text{stat})$	0.52	0.46
Planck PR4 + LOWZ	0.316	0.452 ± 0.005	$0.40^{+0.11}_{-0.09}(\text{stat})$	0.37	0.64

the underlying data is statistically consistent. Therefore to fully contextualize any significant tests we summarize the total number of PTE tests. In this work, we present a total of eight statistical PTE value tests for the data and pipeline. Given this number of tests, we could potentially find a significant value by chance (with a probability of around 33% when assuming independence). Indeed for our initial set of LOWZ simulations, the galaxy autocorrelation recovery test failed marginally as we disclose in Sec. VI D. Since we can create an independent set of simulations we reran the test which passed and we concluded that this was a chance failure. The other PTE value tests presented and in particular none of the four PTE tests that evaluate consistency with GR show a significant result. In summary, we find the values of the PTE tests presented consistent with a flat PTE distribution, i.e., no statistical tension that would indicate inaccuracies in our pipeline or disagreement with GR predictions.

The baseline results and tests are summarized in Table II. Next, we investigate the sensitivity of the results to analysis choices.

B. Consistency tests

The E_G estimate presented in this work consists of a complex analysis that has a range of steps where multiple potential approaches are valid. We investigate some of these analysis choices for their impact on the final constraints. We also quantify the impact of improved corrections and consideration of the uncertainty. The tests are summarized in Fig. 13.

Apart from the simulation-based approach we use for our baseline covariance, we could also use an analytically derived covariance. We find them to be in full agreement with each other (third row of Fig. 13).

We can consider a more conservative cut of $\ell_{\rm min} = 59$ (rather than the baseline $\ell_{\rm min} = 49$), cutting the first bin from the analysis, to more closely match the cosmological range used in Pullen *et al.* [34]. We find that the results with the restricted range agree well with the full range (fourth row). Additionally, we also present the result when cutting all the large-scale bins that show a low result, applying $\ell_{\rm min} = 106$. This shows that the statistically insignificant tendency of the measurement to be low compared to the ΛCDM-GR prediction is driven by the largest scales for both the CMASS and LOWZ measurements.

1. North vs South sample

In this work, we present combined constraints from the SDSS BOSS North and South maps. Previous analyses have noted and investigated differences between the North and South observations of the SDSS datasets (e.g. [44]).

We here investigate the difference in our results between North and South. While we only have the β measurement available for the full sample since we rely on the published data vector that combines the samples for the RSD analysis, we can split the angular power spectra and calculate them for the North and South maps separately. We recalculate the norm correction for the CMB lensing maps specific to the North and South samples and derive the covariance matrix from the same simulations split into the two patches. The results for E_G calculated using only the Northern and Southern patches are shown in Fig. 13. The separate results are in statistical agreement with each other, justifying combining them. The Northern patch contains the majority of the objects which results in a smaller uncertainty for that result, which also agrees closer with the combined sample.

2. Investigating Planck overlap vs full map

The *Planck* CMB lensing map covers most (about 96%) but not all of the SDSS data. We use the full galaxy map to calculate the autocorrelations. If the effective galaxy bias for the overlap region is different from the one for the full sample this could introduce a bias, since while we always use the full SDSS overdensity map the cross-correlation is additionally restricted by the overlap region with the *Planck* data. The LSS samples are designed to be uniform across the footprint through the use of weights. We do not expect the small difference to matter in the analysis, but we can investigate whether our results are sensitive to this difference by recalculating the E_G statistic with the autocorrelation restricted to the overlap region with the *Planck* CMB lensing map. Again we only have the β measurement for the full sample. The results are shown in Fig. 13,

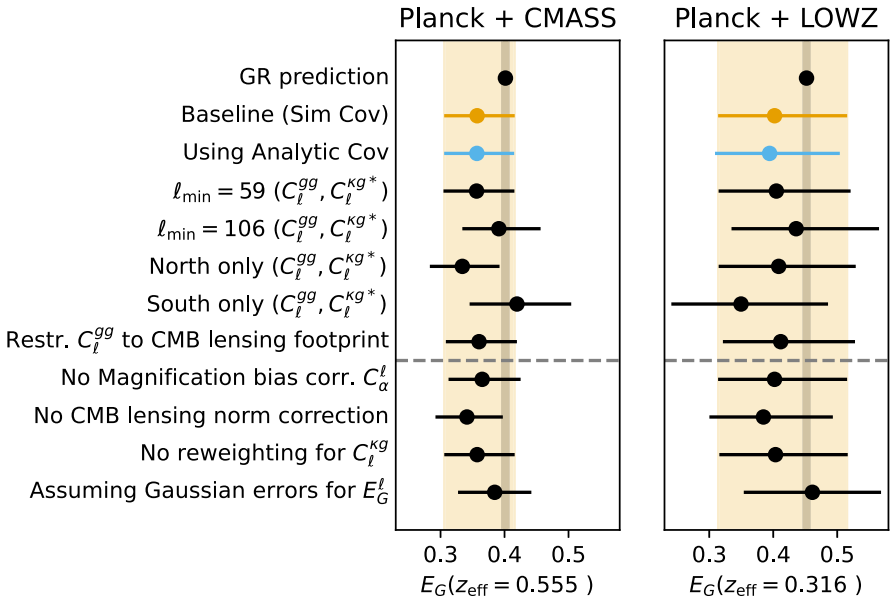

FIG. 13. The sensitivity of the baseline $\hat{E}_G$ results presented in this work to analysis choices and corrections applied. The ΛCDM-GR predicted value and one sigma error based on the fiducial cosmology (gray band) and the 68.27% confidence range for the baseline measurements from this work (orange band) are used as references. These are compared with the results using an analytic covariance instead of simulation-based, restricting to smaller scales with $\ell_{\rm min} = 59$ and $\ell_{\rm min} = 106$, analyzing the angular power spectra for the BOSS North and South patches only and restricting the autocorrelation to the same footprint as the cross-correlation. Additionally, below the gray dashed line, the impact of removing well-motivated analysis steps is shown; removing the magnification bias correction (negligible for *Planck* + LOWZ), not correcting the CMB lensing normalization, not reweighting the galaxy sample for the cross-correlation and not accounting for the non-Gaussianity of the estimator uncertainty.

indicating that our results are insensitive to the choice of restricting the autocorrelation to the overlap with the CMB lensing map.

3. Impact of analysis steps

In the analysis, we carefully account for the effect of magnification bias, correct the normalization of the CMB lensing maps, match the effective redshift of all observables, and account for the non-Gaussianity of the $\hat{E}_G$ estimator uncertainty. We can investigate the impact of these by removing them individually and comparing the results. The results for removing each of them individually are shown in Fig. 13. They all affect the result by only a fraction of the 68.27% (1σ) uncertainty. Removing any one of them, the results are still statistically consistent with GR and they do not all shift in the result in the same direction. The effect of magnification bias and the normalization of the CMB lensing maps approximately cancel each other out for the *Planck* + CMASS measurement. When not reweighting the galaxy sample for the angular cross-correlation the expected value of the estimator has a scale-dependent bias at the percent level due to the mismatch in effective redshift as we show in Appendix B. This is especially relevant when testing for scale dependence. When averaging over our cosmological range in scales the bias cancels partially and we find that the difference is small compared to the statistical uncertainty for both measurements.

We have motivated these analysis steps in this work and argue that they should be applied for an accurate estimate of E_G, nonetheless, these checks show that given the statistical uncertainty of the measurement, these analysis steps are not decisive for our overall conclusion that the measurements are statistically consistent with GR. As statistical constraining power increases with upcoming datasets, these analysis steps will become increasingly impactful.

C. Results in context

In Fig. 14 we plot the baseline marginalized E_G constraints presented in this work compared to other constraints from the literature using *Planck* CMB lensing and SDSS BOSS. The plot shows E_G as a function of redshift. The curve for ΛCDM-GR with our fiducial cosmology is shown as a solid black line with one sigma uncertainty shown as a gray band. The E_G measurements presented in this work are shown with an orange circle showing the result for *Planck* PR4 + CMASS, $E_G(z = 0.555) = 0.36^{+0.06}_{-0.05}(\text{stat})$, and a blue pentagon showing the result for *Planck* PR4 + LOWZ, $\hat{E}_G(z = 0.316) = 0.40^{+0.11}_{-0.09}(\text{stat})$.

Additional literature results using *Planck* and SDSS BOSS to measure E_G are also shown, specifically from Pullen *et al.* [34] who find $E_G(z = 0.57) = 0.243 \pm 0.060(\text{stat}) \pm 0.013(\text{sys})$ for *Planck* + CMASS and Singh *et al.* [32] who find $E_G(z = 0.27) = 0.46^{+0.082}_{-0.088}$ for

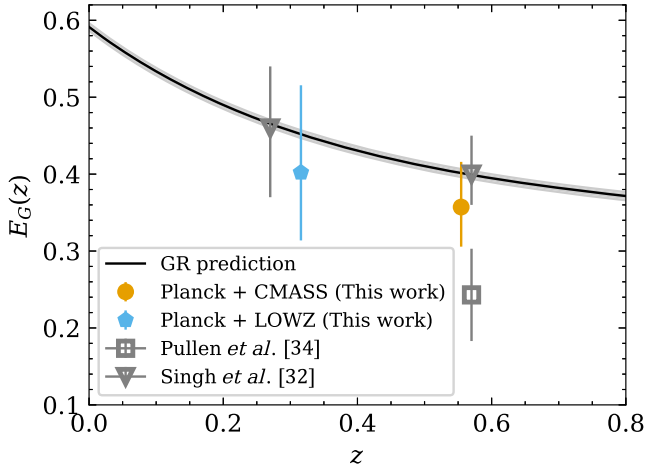

FIG. 14. Overview of the cosmological constraints based on the $\hat{E}_G$ estimator presented in this work applied to the *Planck* PR4 CMB lensing map and the SDSS BOSS DR12 CMASS and LOWZ galaxy samples. Shown are the measurements of E_G at their respective effective redshift for *Planck* + CMASS (orange dots) and *Planck* + LOWZ [blue pentagons]. The ΛCDM-GR expectation for the fiducial cosmology (assuming $\Omega_{m,0} = 0.3111 \pm 0.0056$) is shown as a black line with a gray one sigma uncertainty band. Literature results that use earlier versions of the *Planck* and SDSS data and different approaches to estimate the E_G statistic are shown in gray. Note that the redshifts of the different measurements are not shifted for visibility, other results using *Planck* and SDSS BOSS data quote different effective redshifts. The statistical 68.27% confidence ranges for each measurement are shown, systematic error budgets are not plotted, they are reported as subdominant for all cases.

Planck + CMASS and $E_G(z = 0.57) = 0.4^{+0.049}_{-0.051}$ for *Planck* + LOWZ with 3% systematic error budgets for both. Note the plot only shows the statistical uncertainty in each case. For an overview of additional E_G results based on BOSS quasars or with galaxy weak lensing as the lensing tracer, we refer the interested reader to Fig. 11 in Zhang *et al.* [35].

Compared to the results in Pullen *et al.* [34], we find similar measurement uncertainty, however, the best-fit value in this previous result was $\Delta E_G = 0.12$ lower, leading to a reported significant tension with ΛCDM-GR. Their estimate is also using a harmonic space estimator [21]. We use the same pseudo-Cl approach for the calculation of the auto-correlation as for the cross-correlation with careful consideration of the shot noise while the previous analysis used a quadratic minimum variance estimator for the autocorrelation. The Pullen *et al.* [34] measurement uses *Planck* PR2 [109] and SDSS CMASS DR11 while *Planck* PR4 and SDSS CMASS DR12 are used for our analysis. The CMASS DR12 catalog contains approximately 12% more galaxies compared to DR11. The catalog construction calculation of systematics weights is largely the same, with only minor improvements, including a redefinition of the close pair and redshift failure weight calculation affecting only small fractions of the sample and small changes to the stellar density map and seeing estimation for weights [46]. Pullen *et al.* [34] report the mean redshift of the galaxy sample instead of the effective redshift as defined in this work. Using the more accurate estimation of effective redshift presented in this work would assign a lower redshift to their result leading to an even stronger tension with ΛCDM-GR in their findings than was reported.

In the present work, we motivate, develop, and apply a significantly updated $\hat{E}_G$ estimator, that carefully considers and matches the effective redshift of all observables and does not require corrections based on simulations and an HOD, in contrast to the estimator used in the previous analysis [21]. We compare the estimators in detail in Appendix B, finding that under ideal assumptions the results with the previous estimator together with its significant correction terms match the estimator presented in this work. We argue that the removal of the reliance on large simulation-based corrections makes our approach more robust against potential mismatches between the simulations and the data. We find that the corrections required for the previous estimator are very sensitive to the redshift distribution of galaxies and the analysis assumptions, for example, whether redshift-dependent FKP weights are used or not. Additionally, we add corrections for magnification bias and the CMB lensing normalization that were not accounted for in Pullen *et al.* [34]. We find in our analysis that these two effects approximately cancel for *Planck* + CMASS and also find that restricting our analysis to $\ell \geq 59$, which is closer to $\ell \geq 62$ used in [34], does not significantly change our results (see Fig. 13). The Pullen *et al.* [34] analysis also assumes Gaussian errors for the bin-wise and marginalized estimates. As shown in Sec. VII A, this can significantly underestimate the upper error which for a low measurement overestimates the tension with ΛCDM-GR. Our result for β for CMASS DR12 of $\beta = 0.371^{+0.042}_{-0.055}$ is wholly consistent with the value found in Pullen *et al.* [34] using CMASS DR11 of $\beta = 0.386 \pm 0.046$.

In summary, we have outlined and assessed the sensitivity of the differences in the present analysis relative to those used in Pullen *et al.* [34] to different analysis approaches and differences in the datasets used. The additional corrections for magnification bias and transfer function as well as the slightly less conservative large-scale cutoff are not significant drivers for the differences in results. We further discussed in detail the consistency of the updated data products used in this analysis compared to the previous investigation. There are a range of small differences in the data that could induce differences in the results. We consider reprocessing previous data products, and the reconstruction of results in prior published work, as out of scope for this analysis. The consideration of the non-Gaussian error of the estimator in the present analysis is another contributor to the difference in result.

We can also compare to the measurements presented in Singh *et al.* [32]. The reported measurements are slightly

higher than the results we presented in this work but the disagreement is not as substantial as with Pullen *et al.* [34], and both results lead to the same overall conclusion of consistency with the ΛCDM-GR prediction. We discuss potential sources for the small difference. The Singh *et al.* [32] estimates are based on *Planck* PR2 and SDSS BOSS DR12 and use a real space estimator. The use of a real space estimator makes them less directly comparable to the harmonic estimator presented in this work. We quote the result values given in their Table I for the comparison. The analysis uses different redshift cuts for the galaxy samples. Most notably the upper cutoff for LOWZ is $z < 0.36$ instead of the $z < 0.43$ used in this work, explaining the lower effective redshift of their measurement. Since E_G decreases with redshift, this brings the two results to closer agreement (see Fig. 14). Additionally, the lower redshift cutoffs for both samples differ slightly, but this is unlikely to have a relevant effect. Singh *et al.* [32] note that for combining harmonic space constraints on E_G^ℓ one needs to account for the full covariance making the marginalized result susceptible to noisy covariances. They argue that the large Hartlap corrections needed when inverting the covariance could bias the result for harmonic space measurements of E_G when using jackknife errors. In this work, we do not use jackknife samples to estimate the full covariance since we argue the number of samples available is not sufficient to invert the covariance reliably. For our simulation-based baseline covariance, we have 480 samples, so for our marginalization requiring the full covariance the Hartlap corrections are only 4.8% for *Planck* + CMASS and 3.5% for *Planck* + LOWZ. Furthermore, we also confirmed that we recover consistent results when using an analytic covariance that does not suffer from these issues (see Figs. 12 and 13). Singh *et al.* [32] do discuss magnification bias but estimate it based on an approach that assumes that the SDSS BOSS samples are magnitude-limited, which significantly underestimates the effect (as demonstrated in Wenzl *et al.* [81]). Another difference is that we additionally correct for the CMB lensing norm correction. These two differences partially cancel for *Planck* + CMASS and are not large enough to explain the difference overall. In Singh *et al.* [32] the non-Gaussian uncertainty of the E_G estimator was not considered, therefore their approach is more closely comparable to an inference assuming Gaussianity. Indeed when assuming Gaussianity the results we find (see Fig. 13) match the results of Singh *et al.* [32] relatively closely.

IX. CONCLUSION

In this work, we have developed a significantly updated $\hat{E}_G$ estimator to constrain the properties of gravity using galaxy and CMB lensing data. We have applied this novel estimator to the *Planck* PR4 CMB lensing map and the SDSS BOSS DR12 CMASS and LOWZ galaxy catalogs, spanning two redshift ranges, and find that the measurements are statistically consistent with GR predictions.

The novel $\hat{E}_G$ estimator combines a reweighted angular cross-power spectrum between CMB lensing and galaxy clustering, the angular autopower spectrum of the galaxy sample, and the β parameter from an RSD analysis. The estimator is carefully derived so that the effective redshifts of all the measurements match, to minimize the overall error introduced by approximations and to avoid the need for simulation + HOD-based corrections.

A suite of detailed tests is performed to characterize the estimator's accuracy and its sensitivity to assumptions and choices in the analysis. The findings demonstrate that the estimator serves as an unbiased one for E_G in that statistical uncertainties are dominant over the major sources of potential systematic bias considered. For the CMASS and LOWZ samples, the overall bias of the estimator is within the 0.2% level before considering astrophysical systematics.

Galaxy bias evolution causes a percent level systematic to the estimator for these galaxy samples. It is difficult to correct for this redshift evolution of the galaxy bias since it can only be constrained by assuming GR. However, for the statistical uncertainty of the given sample, the effect is marginal, not restricting comparisons to the GR predictions. Furthermore for future measurements with larger statistical constraining power this systematic can be reduced by measuring E_G in narrower redshift bins. Another systematic effect is from magnification bias. This bias can be constrained from the underlying photometric selection of the galaxy sample [81], allowing careful corrections to account for the effect. The relevance of magnification bias grows with redshift: while it is negligible for a LOWZ-like sample, for CMASS we find a 2–3% correction and expect the impact to increase further with redshift.

In Sec. VII we studied the non-Gaussian uncertainty of the E_G estimator. Especially for individual measurements with large statistical uncertainty, either current datasets or future datasets split into narrow redshift bins, we found that the PDF for the E_G statistic is significantly asymmetric. This makes the 68.27% (the equivalent of 1σ for a Gaussian) confidence limits asymmetric around the median of the distribution and it can affect the best-fit value when combining multiple bins. We also derive two statistical tests that we use to test the measurement for consistency with scale independence and the E_G value predicted by GR.

We reported the constraints from the new $\hat{E}_G$ estimator with the *Planck* PR4 CMB lensing map and SDSS BOSS DR12 CMASS and LOWZ galaxy samples in Table II. Constraints from *Planck* + CMASS sample an effective redshift of $z_{\rm eff}^{\rm Planck+CMASS} = 0.555$ and constraints from *Planck* + CMASS sample $z_{\rm eff}^{\rm Planck+LOWZ} = 0.316$. We measured $\hat{E}_G^\ell$ as a function of angular scale covering a range of

approximately $k \in [0.02, 0.2]$ Mpc^{-1}. We tested the null hypothesis of GR that the E_G statistic is scale-independent, carefully accounting for the cross-covariance between different angular scales. We found $\mathrm{PTE}_{\mathrm{scale-indep}}$ values of 52% for *Planck* + CMASS and 37% for *Planck* + LOWZ, indicating full consistency of the measurement with scale independence as predicted by GR. This test only weakly depends on ΛCDM assumptions through the reweighting scheme and the magnification bias correction which each impacts the result well below the statistical uncertainty.

We combined the constraints across angular scales in the cosmological analysis range, assuming E_G is a constant as a function of scale. We found 68.27% confidence interval values of $\hat{E}_G^{\mathrm{Planck+CMASS}} = 0.36^{+0.06}_{-0.05}(\mathrm{stat})$ and $\hat{E}_G^{\mathrm{Planck+LOWZ}} = 0.40^{+0.11}_{-0.09}(\mathrm{stat})$ for *Planck* + CMASS and *Planck* + LOWZ, respectively, with additional subdominant systematic error budgets of 2% and 3% respectively.

The E_G statistic theory predictions for the ΛCDM cosmology at the effective redshifts for CMASS and LOWZ are $E_G^{\mathrm{GR}}(z = 0.555) = 0.401 \pm 0.005$ and $E_G^{\mathrm{GR}}(z = 0.316) = 0.452 \pm 0.005$. Comparing these to our measurements under careful consideration of the non-Gaussian uncertainty we found $\mathrm{PTE}_{\mathrm{GR}}$ values of 44% and 64% for *Planck* + CMASS and *Planck* + LOWZ, respectively, indicating that the measured $\hat{E}_G$ values are in good agreement with ΛCDM-GR predictions.

We overall found the datasets to be statistically fully consistent with GR predictions. We showed in detail in Fig. 13 that this conclusion is insensitive to a range of analysis choices and that the corrections derived in this work are not decisive for the overall conclusion of agreement with the GR prediction. In Sec. VIII A we additionally described how the data was handled to avoid unconscious bias in the results and discuss the full set of statistical tests performed.

In the context of previous E_G analysis on the *Planck* and SDSS BOSS datasets, we found that our harmonic space results are reasonably consistent with the real-space-based results presented in Singh *et al.* [32], especially when considering differences in the redshift cuts applied to the data and if we assume a Gaussian likelihood for the measurement. However, we do emphasize that the non-Gaussian uncertainty needs to be used to accurately determine constraints. We also discussed that we found a higher result than the one presented in Pullen *et al.* [34] for *Planck* + CMASS which had found results in significant tension with ΛCDM-GR. We discussed differences in the estimator and analysis in detail.

The estimator for the E_G statistic presented in this work includes redshift-dependent reweighing for the cross-correlation, to reduce the scale-dependent bias of the estimator, which requires accurate galaxy redshifts as provided in spectroscopic surveys. Applying this to potential future E_G measurements using photometric samples, such as from LSST [110], *Euclid* [111] and *Roman* [112] as suggested in Pullen *et al.* [21], will require careful consideration in future work.

In addition to testing consistency with GR, the E_G scale-independent and dependent measurements presented here also open up the potential for comparisons with predictions from alternative gravity models. This includes evidence of deviations of E_G from the GR-predicted magnitude and redshift evolution, as in the Chameleon model, and scale dependence, as in, for example, $f(R)$ models [20,21].

The E_G estimator presented in this work is readily applicable to upcoming cosmological datasets like CMB lensing measurements from ACT [36], SPT [38], SO [39] and CMB-S4 [40] as well as spectroscopic galaxy samples from DESI [41] and SPHEREx [42] which will allow interesting new constraints on gravity with this statistic.

ACKNOWLEDGMENTS

We thank Anthony Pullen for providing further information about the work presented in Pullen *et al.* [34] and helpful comments on a preliminary draft of this work. We thank Alex Krolewski for helpful discussions about the bias evolution with redshift for SDSS BOSS. We thank Héctor Gil-Marín for providing us with the official BOSS measurements including window matrices from [61] for the RSD analysis. We thank Zeyang Sun for helpful clarifications and discussions on ratio distributions presented in Sun *et al.* [106]. The work of L. W. and R. B. is supported by NSF Grant No. AST-2206088, NASA ATP Grant No. 80NSSC18K0695, and NASA ROSES Grant No. 12-EUCLID12-0004. M. M. acknowledges support from NSF Grants No. AST-2307727 and No. AST-2153201 and NASA Grant No. 21-ATP21-0145. G. S. F. acknowledges support through the Isaac Newton Studentship and the Helen Stone Scholarship at the University of Cambridge. G. S. F. and B. S. acknowledge support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 851274). S. C. acknowledges the support of the National Science Foundation at the Institute for Advanced Study. G. M. is part of Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics. N. S, acknowledges support from DOE Award No. DE-SC0020441. A. v. E. acknowledges support from NASA Grants No. 80NSSC23K0747 and No. 80NSSC23K0464. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231 using

NERSC Award No. HEP-ERCAPmp107. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

APPENDIX A: SDSS MASK MAKING

In this section, we describe how we construct the masks for the galaxy samples LOWZ and CMASS shown in Fig. 15. Accurate masks are crucial for an unbiased estimate of the normalization of the 2 pt correlation functions in harmonic space, while the original SDSS analysis primarily focused on real-space approaches. We need to estimate the fractional coverage of each pixel for our HEALPix maps using nside $= 1024$. To find the fractional coverage we need to combine information about the survey footprint and the local completeness, as well as account for veto regions. Veto regions were not included as part of the SDSS completeness calculation and therefore need to be accounted for separately.

The SDSS team provides completeness information per sector and veto areas as mangle polygon files.[15] Sectors are defined as unique combinations of points. We use PYMANGLE[16] to check individual sky locations against these polygon files. To accurately estimate the fractional coverage of each pixel we check 4096 sample points per pixel. Each point that is inside a veto region or outside the survey footprint is set to 0 and the other points are set to the local completeness value. The average of all points is then the coverage fraction of the pixel considered. This is equivalent to checking the centers of pixels in a nside $= 65,536$ map and then downsampling via averaging. For optimal memory efficiency, we do this in patches and for speed in parallel. With this, we estimate the coverage fraction accurate to $< 0.5\%$ (1 sigma). Note that this quoted accuracy neglects uncertainty on the original estimates of completeness and the definition of the veto regions presented in [46] which we presume to be accurate.

[15]https://data.sdss.org/sas/dr12/boss/lss/geometry/.
[16]https://github.com/esheldon/pymangle.

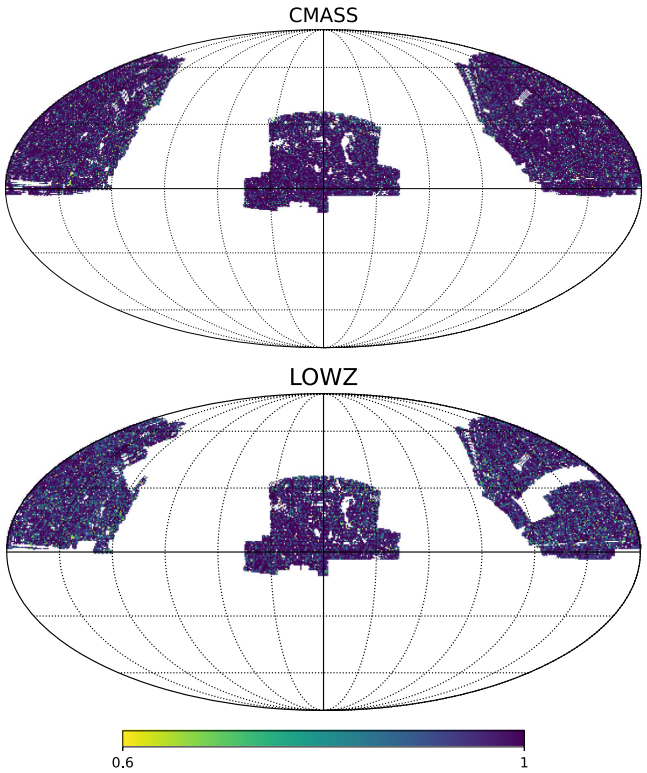

FIG. 15. Observational masks of the CMASS and LOWZ samples in equatorial coordinates and Mollweide projection. The values show the fractional coverage of each pixel combining survey footprint, local completeness, and accounting for veto regions. Shown are the pixels with coverage fraction above 60% which is the lower cutoff used in the analysis.

This avoids potential inaccuracies from representing HEALPix pixels as polygons and is much more computationally efficient. It also adds less additional shot noise than using randoms to construct the masks, although, for the scales of interest in this analysis, the difference is mostly negligible.

APPENDIX B: COMPARING OUR E_G ESTIMATOR TO A PREVIOUS ESTIMATOR

The new estimator $\hat{E}_G$ we present in Eq. (34) contains notable changes compared to a previous harmonic space E_G estimator for CMB lensing and galaxy clustering presented in Pullen *et al.* [21,34] given by

$$\hat{E}_G^{\mathrm{Pullen}}(\bar{z}) = \Gamma^{\mathrm{Pullen}}(\bar{z}) \frac{C_\ell^{\kappa g}}{\beta C_\ell^{gg}}, \tag{B1}$$

$$\Gamma^{\mathrm{Pullen}}(\bar{z}) = \frac{2cH(\bar{z})}{3H_0^2} \frac{W_g(\bar{z})}{\hat{W}_\kappa(\bar{z})}, \tag{B2}$$

where $\bar{z}^{\mathrm{Pullen}} = 0.57$ is their quoted redshift of the measurement. The key improvements of our estimator are given as follows:

(1) Our final expression for Γ has a factor $\int \mathrm{d}z \frac{W_g^2(z)}{\hat{W}_\kappa(z)}$ instead of the $W_g(\bar{z})/\hat{W}_\kappa(\bar{z})$ factor in $\hat{E}_G^{\mathrm{Pullen}}$.
(2) We reweigh the galaxy samples for the cross-correlation to match the effective redshift of the measurement to the autocorrelation. So our $\hat{E}_G$ uses $C_\ell^{\kappa g*}$ compared to $C_\ell^{\kappa g}$ in $\hat{E}_G^{\mathrm{Pullen2015}}$.
(3) We choose to project our angular power spectra with the clustering kernel which defines our effective redshift. This choice is convenient since then the effective redshift of the RSD measurement matches that of the angular power spectra.
(4) These changes make the overall systematics of the estimator negligible when inputting a simulation with exactly linear bias and no other astrophysical systematics, S_Γ^ℓ [Eq. (36)], as opposed to in Pullen *et al.* [34] where a correction of approximately 5% based on simulations and assuming a HOD is used.
(5) We add a correction for the magnification bias of the galaxy samples (see Sec. V C 3).
(6) We account for the amplitude error in the CMB lensing reconstruction due to the mask (see Sec. VI B 1).
(7) We account for the non-Gaussian uncertainty of the estimator (see Sec. VII).

The first change could be directly applied to the $\hat{E}_G^{\mathrm{Pullen2016}}$ estimator without any other changes to partially resolve the large overall offset of the estimator. We first discuss the underlying reason for this large offset. The approximation $W_g(\chi) \approx \frac{W_g^2(\chi)}{W_g(\chi(\bar{z}))}$ used in Pullen *et al.* [21,34] would be close for galaxy samples where $\mathrm{d}N(z)/\mathrm{d}z$ is close to a constant in the observed redshift range. However, for realistic galaxy samples, this approximation can introduce a large offset, depending sensitively on the shape of the $\mathrm{d}N(z)/\mathrm{d}z$ distribution and the choice of redshift $\bar{z}$. We can quantify this with the overall error of the estimator S_Γ^ℓ for the case of SDSS CMASS. In Pullen *et al.* [34] FKP weights were not considered so in the following, to allow for comparison to their results, we also removed FKP weights for this discussion, to have a comparable galaxy sample. We include FKP weights in our analysis since they optimally weight the data for the RSD analysis minimizing the uncertainty on β. We note that the accuracy of the estimator presented in this work is similar with or without the use of FKP weights, but since these weights are redshift dependent they do sensitively affect the accuracy of the $\hat{E}_G^{\mathrm{Pullen2016}}$ estimator. Using our well-motivated effective redshift calculation, we find $z_{\mathrm{eff}} = 0.531$ for the CMASS sample without FKP weights, the offset S_Γ^ℓ in the final estimate using $\hat{E}_G^{\mathrm{Pullen}}(z_{\mathrm{eff}})$ is roughly 20%. For their choice of redshift of $\bar{z}^{\mathrm{Pullen}} = 0.57$ and using $\hat{E}_G^{\mathrm{Pullen}}(\bar{z}^{\mathrm{Pullen}} = 0.57)$ the offset is roughly 5% as discussed in their paper. This higher value of redshift was used based on the value reported in the BOSS DR11 release paper [63]. Note that one could also choose a lower redshift to similarly reduce the amount of correction needed. The correction only reduces because the approximation above is more accurate for redshifts off the peak for a CMASS-like redshift distribution.

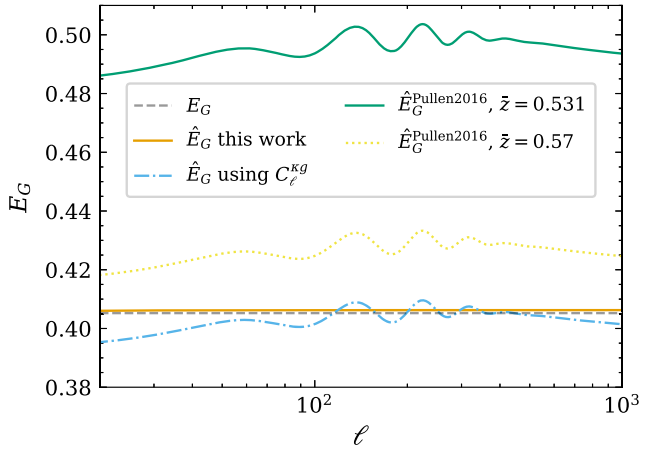

FIG. 16. A summary of how the improvements to the $\hat{E}_G$ estimator proposed in this work impact the overall accuracy of the estimator for our fiducial cosmology. Shown is the case of CMB lensing combined with a CMASS-like galaxy sample without considering FKP weights with $z_{\mathrm{eff}} = 0.531$. Shown are the analytic E_G value for the fiducial cosmology (black dashed), $\hat{E}_G$ using the estimator in this work [orange, full], our estimator but using $C_\ell^{\kappa g}$ rather than the reweighted $C_\ell^{\kappa g*}$ (blue, dot-dashed) and the estimator used in Pullen *et al.* [34] for both the redshift assumed in that work, $z = 0.57$, (yellow, dotted) and at the effective redshift of the sample, $z = 0.531$ [green, full]. See Appendix B for details.

Even after this change a scale-dependent (ringing) bias remains in the estimator that is caused by the mismatch of the effective redshift of the auto and cross-correlations sample. We resolve this by reweighing the galaxy sample for the cross-correlation and shifting the effective redshift to match the one for the autocorrelation.

In Fig. 16 we show the expectation value of the estimator for a CMASS-like sample,[17] when plugging in the analytic expressions for our fiducial cosmology and using the effective redshift of the sample as defined in this work ($z_{\mathrm{eff}} = 0.531$). This test does not model the magnification bias and uses a linear galaxy bias. Our final $\hat{E}_G$ estimator (solid orange line) matches the theory prediction E_G (black dashed line) very well ($|S_\Gamma^\ell - 1| < 0.3\%$). If we use $C_\ell^{\kappa g}$ instead of our reweighted $C_\ell^{\kappa g*}$, shown as a blue dot-dashed line, we see a scale-dependent bias of a few percent. If we additionally use $W_g(\bar{z})/\hat{W}_\kappa(\bar{z})$ instead of

[17]Without FKP weights to allow for direct comparison with Pullen *et al.* [34] but otherwise as described in Sec. IV. Note that the choice of using or not using FKP weights has no significant effect on the accuracy of the new estimator presented in this work.

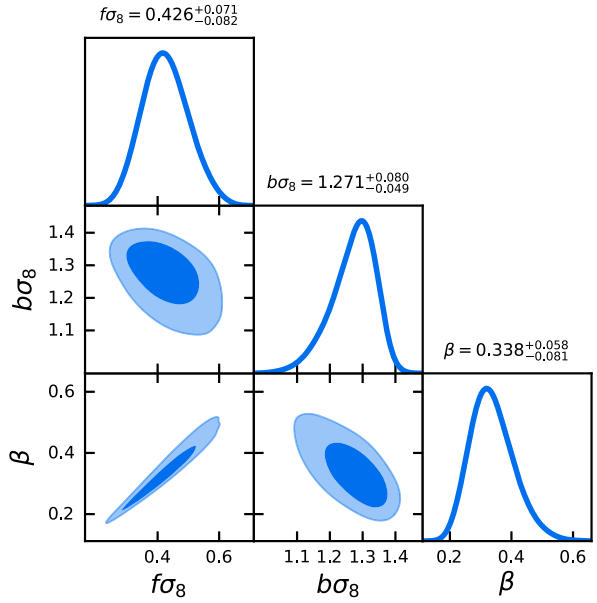

FIG. 17. RSD fit to the clustering signal of the LOWZ dataset. The corner plot shows the fitting parameters $f\sigma_8, b\sigma_8$ as well as the derived parameter $\beta = f\sigma_8/(b\sigma_8)$ used as part of our E_G estimate. For each parameter, we also show the mean value and one sigma error.

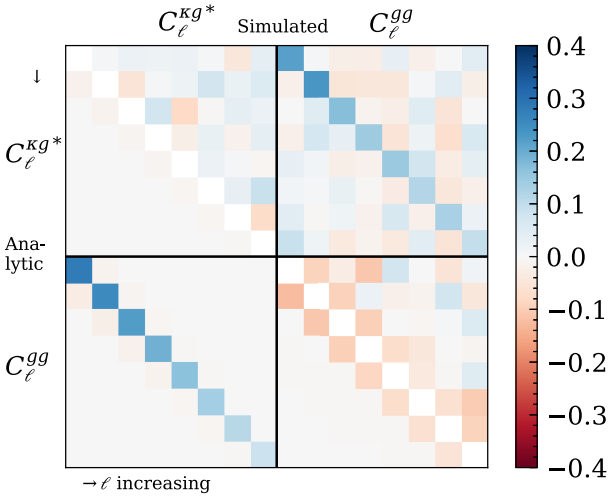

FIG. 18. Visualizing the off-diagonal correlation matrix between C_ℓ^{gg} and $C_\ell^{\kappa g *}$ for *Planck* and LOWZ. Shown is the correlation matrix defined as $\mathrm{Cov}(C_\ell^{XY}, C_\ell^{AB})/(\sigma_\ell^{XY}\sigma_\ell^{AB})$ with the diagonal removed for readability. The upper triangle shows the simulated covariance and the lower triangle the analytic covariance. Correlations between C_ℓ^{gg} and $C_\ell^{\kappa g *}$ are up to around 0.35 for our narrowest large-scale bins. The off-diagonals of the simulated covariance match the analytic covariance well within statistical uncertainty. For a comparison of the diagonals of the covariances see Fig. 19.

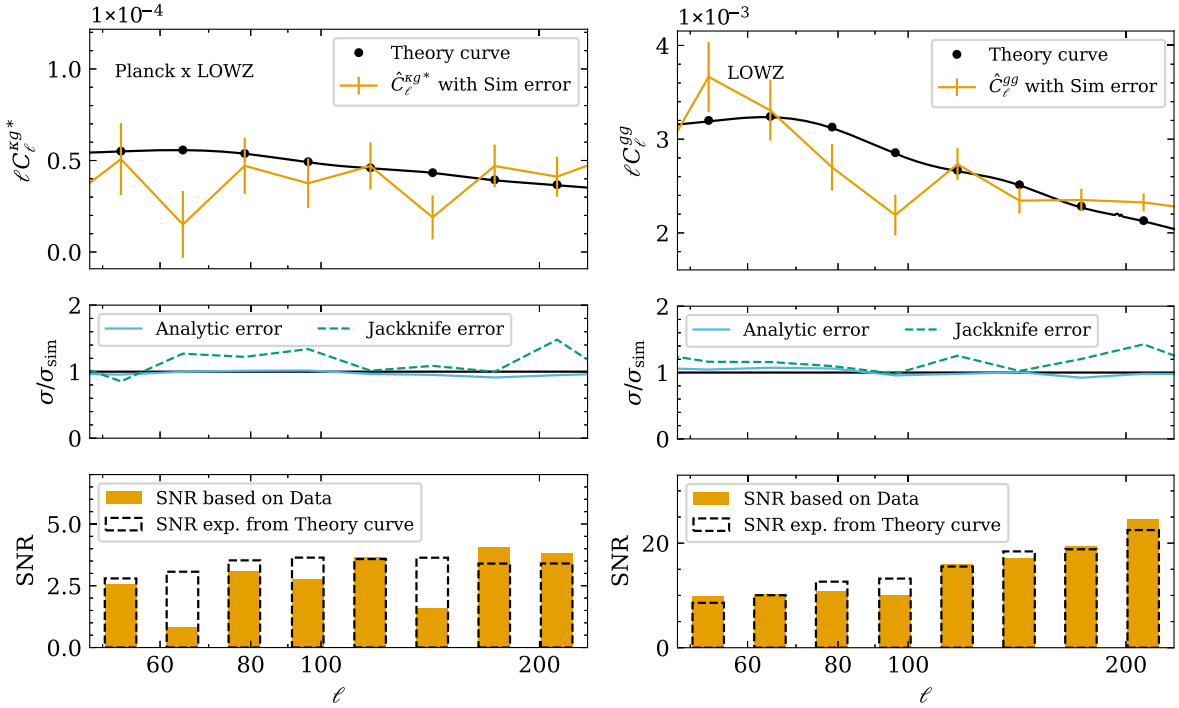

FIG. 19. Measurement of (Left) the angular cross-power spectrum, $C_\ell^{\kappa g *}$, for CMB lensing map based on *Planck* PR4 and the reweighted LOWZ galaxy sample and [Right] the angular auto-LOWZ galaxy power power spectrum (equivalent figure for CMASS shown in Fig. 6). (Top) The angular power spectrum with simulation-based errors as well as our theory curve. (Middle) Comparison of the analytic (solid blue line) and jackknife error estimates [dashed green] to the baseline marginalized errors from the simulations. (Lower) The signal-to-noise ratio, calculated as the ratio of measurement and measurement uncertainty is shown (filled orange). For reference, the ratio of the theory expectation and the measurement uncertainty is also shown (dashed black).

$\int dz \frac{W_g^2(z)}{W_\kappa(z)}$ we recover the functional form of the previous estimator $\hat{E}_G^{\text{Pullen2016}}$. For our effective redshift of the CMASS sample (shown as a green solid line), this estimator has an overall 20% bias that needs to be corrected for, as we already discussed by considering S_Γ^ℓ. For $\bar{z} = 0.57$ as assumed in Pullen *et al.* [34] (shown as a yellow dotted line) the bias is around 5% matching what is shown in their Fig. 3. Note that for the different $\bar{z} = 0.57$, the theory prediction and the other factors in Γ also shift slightly.

We want to stress that together with their C^Γ correction the estimator for E_G presented in Pullen *et al.* [21,34] is unbiased. However, the large correction relies on the accuracy of N-body simulations and the choice of HOD. We argue it is preferable to adapt our improved estimator to not require this correction for an unbiased result.

APPENDIX C: ADDITIONAL PLOTS FOR LOWZ SAMPLE

Here we show additional plots involving the LOWZ dataset. In Fig. 17 we show the RSD fit to measure beta for the LOWZ sample. In Fig. 18 we show the off-diagonal covariance matrix as a correlation matrix for the angular power spectrum measurements with *Planck* and LOWZ.

Figure 19 shows our cross-correlation measurement between *Planck* and LOWZ as well as the autocorrelation measurement of LOWZ.

In Fig. 20 we characterize the statistical test for scale independence of our measurement for the case of *Planck* + LOWZ.

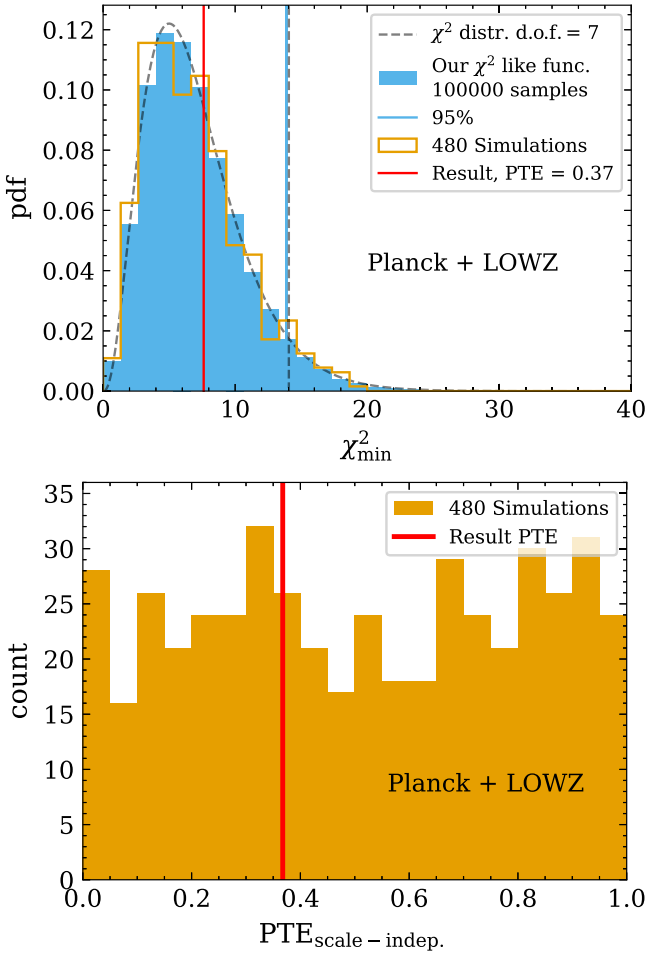

FIG. 20. Characterizing the statistical test for consistency of our measurement with scale independence for *Planck* + LOWZ. Content as described for CMASS in Fig. 10.

APPENDIX D: ERROR PROPOGATION FOR COVARIANCES

We, for convenience, discuss the error propagation for the covariances when combining multiple C_ℓ when assuming Gaussianity. If the two measurements are significantly correlated we need to account for the cross-covariance.

When multiplying or taking the ratio of two measurements, as is done for E_G it is convenient to work in fractional errors defining,

$$\text{CovN}(A_\ell, B_{\ell'}) \equiv \frac{\text{Cov}(A_\ell, B_{\ell'})}{A_\ell B_{\ell'}}. \tag{D1}$$

For multiplication or division $M_\ell \equiv C_\ell^X \cdot (C_\ell^Y)^{\pm 1}$ Gaussian error propagation gives,

$$\begin{aligned}\text{CovN}(M_\ell, M_{\ell'}) = {} & \text{CovN}(C_\ell^X, C_{\ell'}^X) + \text{CovN}(C_\ell^Y, C_{\ell'}^Y) \\ & \pm \text{CovN}(C_\ell^X, C_{\ell'}^Y) \pm \text{CovN}(C_\ell^Y, C_{\ell'}^X).\end{aligned} \tag{D2}$$

For the full E_G estimator as defined in Eq. (34) which depends on $C_\ell^{\kappa g}, C_\ell^{gg}$ and β one obtains

$$\begin{aligned}&\text{CovN}(E_G^\ell, E_G^{\ell'}) \\ &= \text{CovN}(C_\ell^{\kappa g}, C_{\ell'}^{\kappa g}) + \text{CovN}(C_\ell^{gg}, C_{\ell'}^{gg}) + \left(\frac{\sigma_\beta}{\beta}\right)^2 \\ &\quad - \text{CovN}(C_\ell^{\kappa g}, C_{\ell'}^{gg}) - \text{CovN}(C_\ell^{gg}, C_{\ell'}^{\kappa g}) \\ &\quad - \text{CovN}(\beta, C_{\ell'}^{\kappa g}) + \text{CovN}(\beta, C_{\ell'}^{gg}) \\ &\quad - \text{CovN}(C_\ell^{\kappa g}, \beta) + \text{CovN}(C_\ell^{gg}, \beta).\end{aligned} \tag{D3}$$

[1] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner *et al.*, Astron. J. **116**, 1009 (1998).

[2] S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom *et al.*, Astrophys. J. **517**, 565 (1999).

[3] P. J. Peebles and B. Ratra, Rev. Mod. Phys. **75**, 559 (2003).

[4] S. Alam, M. Ata, S. Bailey, F. Beutler, D. Bizyaev, J. A. Blazek, A. S. Bolton, J. R. Brownstein, A. Burden, C.-H. Chuang *et al.*, Mon. Not. R. Astron. Soc. **470**, 2617 (2017).

[5] N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo *et al.* (Planck Collaboration), Astron. Astrophys. **641**, A6 (2020).

[6] W. L. Freedman, B. F. Madore, D. Hatt, T. J. Hoyt, I. S. Jang, R. L. Beaton, C. R. Burns, M. G. Lee, A. J. Monson, J. R. Neeley *et al.*, Astrophys. J. **882**, 34 (2019).

[7] M. S. Madhavacheril, F. J. Qu, B. D. Sherwin, N. MacCrann, Y. Li, I. Abril-Cabezas, P. A. R. Ade, S. Aiola, T. Alford, M. Amiri *et al.*, Astrophys. J. **962**, 113 (2024).

[8] S. Weinberg, Rev. Mod. Phys. **61**, 1 (1989).

[9] S. M. Carroll, V. Duvvuri, M. Trodden, and M. S. Turner, Phys. Rev. D **70**, 043528 (2004).

[10] J. Khoury and A. Weltman, Phys. Rev. Lett. **93**, 171104 (2004).

[11] G. Dvali, G. Gabadadze, and M. Porrati, Phys. Lett. B **485**, 208 (2000).

[12] J. D. Bekenstein, Phys. Rev. D **70**, 083509 (2004).

[13] T. Clifton, P. G. Ferreira, A. Padilla, and C. Skordis, Phys. Rep. **513**, 1 (2012).

[14] B. Jain and J. Khoury, Ann. Phys. (Amsterdam) **325**, 1479 (2010).

[15] D. Huterer, D. Kirkby, R. Bean, A. Connolly, K. Dawson, S. Dodelson, A. Evrard, B. Jain, M. Jarvis, E. Linder *et al.*, Astropart. Phys. **63**, 23 (2015).

[16] J.-P. Uzan, Gen. Relativ. Gravit. **42**, 2219 (2010).

[17] B. Jain and P. Zhang, Phys. Rev. D **78**, 063503 (2008).

[18] M. Ishak, Living Rev. Relativity **22**, 1 (2019).

[19] J. Hou, J. Bautista, M. Berti, C. Cuesta-Lazaro, C. Hernández-Aguayo, T. Tröster, and J. Zheng, Universe **9**, 302 (2023).

[20] P. Zhang, M. Liguori, R. Bean, and S. Dodelson, Phys. Rev. Lett. **99**, 141302 (2007).

[21] A. R. Pullen, S. Alam, and S. Ho, Mon. Not. R. Astron. Soc. **449**, 4326 (2015).

[22] C. D. Leonard, P. G. Ferreira, and C. Heymans, J. Cosmol. Astropart. Phys. 12 (2015) 051.

[23] T. Giannantonio, P. Fosalba, R. Cawthon, Y. Omori, M. Crocce, F. Elsner, B. Leistedt, S. Dodelson, A. Benoit-Lévy, E. Gaztañaga *et al.*, Mon. Not. R. Astron. Soc. **456**, 3213 (2016).

[24] F. Bianchini and C. L. Reichardt, Astrophys. J. **862**, 81 (2018).

[25] Y. Omori, T. Giannantonio, A. Porredon, E. Baxter, C. Chang, M. Crocce, P. Fosalba, A. Alarcon, N. Banik, J. Blazek *et al.*, Phys. Rev. D **100**, 043501 (2019).

[26] G. A. Marques and A. Bernui, J. Cosmol. Astropart. Phys. 05 (2020) 052.

[27] R. Reyes, R. Mandelbaum, U. Seljak, T. Baldauf, J. E. Gunn, L. Lombriser, and R. E. Smith, Nature (London) **464**, 256 (2010).

[28] C. Blake, S. Joudaki, C. Heymans, A. Choi, T. Erben, J. Harnois-Deraps, H. Hildebrandt, B. Joachimi, R. Nakajima, L. van Waerbeke *et al.*, Mon. Not. R. Astron. Soc. **456**, 2806 (2016).

[29] S. de la Torre, E. Jullo, C. Giocoli, A. Pezzotta, J. Bel, B. R. Granett, L. Guzzo, B. Garilli, M. Scodeggio, M. Bolzonella *et al.*, Astron. Astrophys. **608**, A44 (2017).

[30] S. Alam, H. Miyatake, S. More, S. Ho, and R. Mandelbaum, Mon. Not. R. Astron. Soc. **465**, 4853 (2017).

[31] A. Amon, C. Blake, C. Heymans, C. D. Leonard, M. Asgari, M. Bilicki, A. Choi, T. Erben, K. Glazebrook, J. Harnois-Déraps *et al.*, Mon. Not. R. Astron. Soc. **479**, 3422 (2018).

[32] S. Singh, S. Alam, R. Mandelbaum, U. Seljak, S. Rodriguez-Torres, and S. Ho, Mon. Not. R. Astron. Soc. **482**, 785 (2019).

[33] C. Blake, A. Amon, M. Asgari, M. Bilicki, A. Dvornik, T. Erben, B. Giblin, K. Glazebrook, C. Heymans, H. Hildebrandt *et al.*, Astron. Astrophys. **642**, A158 (2020).

[34] A. R. Pullen, S. Alam, S. He, and S. Ho, Mon. Not. R. Astron. Soc. **460**, 4098 (2016).

[35] Y. Zhang, A. R. Pullen, S. Alam, S. Singh, E. Burtin, C.-H. Chuang, J. Hou, B. W. Lyke, A. D. Myers, R. Neveux *et al.*, Mon. Not. R. Astron. Soc. **501**, 1013 (2021).

[36] F. J. Qu, B. D. Sherwin, M. S. Madhavacheril, D. Han, K. T. Crowley, I. Abril-Cabezas, P. A. R. Ade, S. Aiola, T. Alford, M. Amiri *et al.*, Astrophys. J. **962**, 112 (2024).

[37] K. T. Story, D. Hanson, P. A. R. Ade, K. A. Aird, J. E. Austermann, J. A. Beall, A. N. Bender, B. A. Benson, L. E. Bleem, J. E. Carlstrom *et al.*, Astrophys. J. **810**, 50 (2015).

[38] Z. Pan, F. Bianchini, W. L. K. Wu, P. A. R. Ade, Z. Ahmed, E. Anderes, A. J. Anderson, B. Ansarinejad, M. Archipley, K. Aylor *et al.*, Phys. Rev. D **108**, 122005 (2023).

[39] P. Ade, J. Aguirre, Z. Ahmed, S. Aiola, A. Ali, D. Alonso, M. A. Alvarez, K. Arnold, P. Ashton, J. Austermann *et al.*, J. Cosmol. Astropart. Phys. 02 (2019) 056.

[40] K. N. Abazajian, P. Adshead, Z. Ahmed, S. W. Allen, D. Alonso, K. S. Arnold, C. Baccigalupi, J. G. Bartlett, N. Battaglia, B. A. Benson *et al.*, arXiv:1610.02743.

[41] A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, L. E. Allen, C. Allende Prieto, J. Annis, S. Bailey, C. Balland *et al.* (DESI Collaboration), arXiv:1611.00036.

[42] O. Doré, J. Bock, M. Ashby, P. Capak, A. Cooray, R. de Putter, T. Eifler, N. Flagey, Y. Gong, S. Habib *et al.*, arXiv:1412.4872.

[43] C. Doux, M. Penna-Lima, S. D. P. Vitenti, J. Tréguer, E. Aubourg, and K. Ganga, Mon. Not. R. Astron. Soc. **480**, 5386 (2018).

[44] S.-F. Chen, M. White, J. DeRose, and N. Kokron, J. Cosmol. Astropart. Phys. 07 (2022) 041.

[45] J. Carron, M. Mirmelstein, and A. Lewis, J. Cosmol. Astropart. Phys. 09 (2022) 039.

[46] B. Reid, S. Ho, N. Padmanabhan, W. J. Percival, J. Tinker, R. Tojeiro, M. White, D. J. Eisenstein, C. Maraston, A. J. Ross *et al.*, Mon. Not. R. Astron. Soc. **455**, 1553 (2016).

[47] C.-P. Ma and E. Bertschinger, Astrophys. J. **455**, 7 (1995).

[48] A. Silvestri, L. Pogosian, and R. V. Buniy, Phys. Rev. D **87**, 104015 (2013).
[49] T. Baker, P. G. Ferreira, C. D. Leonard, and M. Motta, Phys. Rev. D **90**, 124030 (2014).
[50] L. Wenzl, C. Doux, C. Heinrich, R. Bean, B. Jain, O. Doré, T. Eifler, and X. Fang, Mon. Not. R. Astron. Soc. **512**, 5311 (2022).
[51] E. V. Linder, Astropart. Phys. **29**, 336 (2008).
[52] N. Aghanim, Y. Akrami, F. Arroja, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro *et al.* (Planck Collaboration), Astron. Astrophys. **641**, A1 (2020).
[53] N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo *et al.* (Planck Collaboration), Astron. Astrophys. **641**, A8 (2020).
[54] Y. Akrami, K. J. Andersen, M. Ashdown, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak *et al.* (Planck Collaboration), Astron. Astrophys. **643**, A42 (2020).
[55] E. Rosenberg, S. Gratton, and G. Efstathiou, Mon. Not. R. Astron. Soc. **517**, 4620 (2022).
[56] M. Tristram, A. J. Banday, M. Douspis, X. Garrido, K. M. Górski, S. Henrot-Versillé, L. T. Hergt, S. Ilić, R. Keskitalo, G. Lagache *et al.*, Astron. Astrophys. **682**, A37 (2024).
[57] A. Krolewski, S. Ferraro, E. F. Schlafly, and M. White, J. Cosmol. Astropart. Phys. 05 (2020) 047.
[58] M. White, R. Zhou, J. DeRose, S. Ferraro, S.-F. Chen, N. Kokron, S. Bailey, D. Brooks, J. García-Bellido, J. Guy *et al.*, J. Cosmol. Astropart. Phys. 02 (2022) 007.
[59] K. S. Dawson, D. J. Schlegel, C. P. Ahn, S. F. Anderson, É. Aubourg, S. Bailey, R. H. Barkhouser, J. E. Bautista, A. Beifiori, A. A. Berlind *et al.*, Astron. J. **145**, 10 (2013).
[60] D. J. Eisenstein, D. H. Weinberg, E. Agol, H. Aihara, C. Allende Prieto, S. F. Anderson, J. A. Arns, É. Aubourg, S. Bailey, E. Balbinot *et al.*, Astron. J. **142**, 72 (2011).
[61] H. Gil-Marín, W. J. Percival, J. R. Brownstein, C.-H. Chuang, J. N. Grieb, S. Ho, F.-S. Kitaura, C. Maraston, F. Prada, S. Rodríguez-Torres *et al.*, Mon. Not. R. Astron. Soc. **460**, 4188 (2016).
[62] C.-H. Chuang, M. Pellejero-Ibanez, S. Rodríguez-Torres, A. J. Ross, G.-b. Zhao, Y. Wang, A. J. Cuesta, J. A. Rubiño-Martín, F. Prada, S. Alam *et al.*, Mon. Not. R. Astron. Soc. **471**, 2370 (2017).
[63] L. Anderson, É. Aubourg, S. Bailey, F. Beutler, V. Bhardwaj, M. Blanton, A. S. Bolton, J. Brinkmann, J. R. Brownstein, A. Burden *et al.*, Mon. Not. R. Astron. Soc. **441**, 24 (2014).
[64] A. Nicola, C. García-García, D. Alonso, J. Dunkley, P. G. Ferreira, A. Slosar, and D. N. Spergel, J. Cosmol. Astropart. Phys. 03 (2021) 067.
[65] G. A. Marques, M. S. Madhavacheril, O. Darwish, S. Shaikh, M. Aguena, O. Alves, S. Avila, D. Bacon, E. J. Baxter, K. Bechtol *et al.*, J. Cosmol. Astropart. Phys. 01 (2024) 033.
[66] N. E. Chisari, D. Alonso, E. Krause, C. D. Leonard, P. Bull, J. Neveu, A. S. Villarreal, S. Singh, T. McClintock, J. Ellison *et al.*, Astrophys. J. Suppl. Ser. **242**, 2 (2019).
[67] A. Lewis, A. Challinor, and A. Lasenby, Astrophys. J. **538**, 473 (2000).
[68] C. Howlett, A. Lewis, A. Hall, and A. Challinor, J. Cosmol. Astropart. Phys. 04 (2012) 027.
[69] G. S. Farren, A. Krolewski, N. MacCrann, S. Ferraro, I. Abril-Cabezas, R. An, Z. Atkins, N. Battaglia, J. R. Bond, E. Calabrese *et al.*, arXiv:2309.05659.
[70] H. Hildebrandt, L. van Waerbeke, and T. Erben, Astron. Astrophys. **507**, 683 (2009).
[71] F. Schmidt, A. Leauthaud, R. Massey, J. Rhodes, M. R. George, A. M. Koekemoer, A. Finoguenov, and M. Tanaka, Astrophys. J. Lett. **744**, L22 (2012).
[72] C. Duncan, B. Joachimi, A. Heavens, C. Heymans, and H. Hildebrandt, Mon. Not. R. Astron. Soc. **437**, 2471 (2014).
[73] H. Hildebrandt, Mon. Not. R. Astron. Soc. **455**, 3943 (2016).
[74] L. Thiele, C. A. J. Duncan, and D. Alonso, Mon. Not. R. Astron. Soc. **491**, 1746 (2020).
[75] S. Unruh, P. Schneider, S. Hilbert, P. Simon, S. Martin, and J. C. Puertas, Astron. Astrophys. **638**, A96 (2020).
[76] R. Maartens, J. Fonseca, S. Camera, S. Jolicoeur, J.-A. Viljoen, and C. Clarkson, J. Cosmol. Astropart. Phys. 12 (2021) 009.
[77] M. von Wietersheim-Kramsta, B. Joachimi, J. L. van den Busch, C. Heymans, H. Hildebrandt, M. Asgari, T. Tröster, S. Unruh, and A. H. Wright, Mon. Not. R. Astron. Soc. **504**, 1452 (2021).
[78] C. A. J. Duncan, J. Harnois-Déraps, L. Miller, and A. Langedijk, Mon. Not. R. Astron. Soc. **515**, 1130 (2022).
[79] F. Lepori, I. Tutusaus, C. Viglione, C. Bonvin, S. Camera, F. J. Castander, R. Durrer, P. Fosalba, G. Jelic-Cizmek *et al.* (Euclid Collaboration), Astron. Astrophys. **662**, A93 (2022).
[80] J. Elvin-Poole, N. MacCrann, S. Everett, J. Prat, E. S. Rykoff, J. De Vicente, B. Yanny, K. Herner, A. Ferté, E. Di Valentino *et al.*, Mon. Not. R. Astron. Soc. **523**, 3649 (2023).
[81] L. Wenzl, S.-F. Chen, and R. Bean, Mon. Not. R. Astron. Soc. **527**, 1760 (2023).
[82] A. Moradinezhad Dizgah and R. Durrer, J. Cosmol. Astropart. Phys. 09 (2016) 035.
[83] S. Yang and A. R. Pullen, Mon. Not. R. Astron. Soc. **481**, 1441 (2018).
[84] B. Ghosh and R. Durrer, J. Cosmol. Astropart. Phys. 06 (2019) 010.
[85] D. Alonso, J. Sanchez, A. Slosar (LSST Dark Energy Science Collaboration), Mon. Not. R. Astron. Soc. **484**, 4127 (2019).
[86] A. Benoit-Lévy, T. Déchelette, K. Benabed, J. F. Cardoso, D. Hanson, and S. Prunet, Astron. Astrophys. **555**, A37 (2013).
[87] J. Carron, J. Cosmol. Astropart. Phys. 02 (2023) 057.
[88] A. Krolewski, W. J. Percival, S. Ferraro, E. Chaussidon, M. Rezaie, J. N. Aguilar, S. Ahlen, D. Brooks, K. Dawson, A. de la Macorra *et al.*, J. Cosmol. Astropart. Phys. 03 (2024) 021.
[89] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, Astrophys. J. **622**, 759 (2005).

[90] A. Zonca, L. Singer, D. Lenz, M. Reinecke, C. Rosset, E. Hivon, and K. Gorski, J. Open Source Software **4**, 1298 (2019).
[91] M. White, M. Blanton, A. Bolton, D. Schlegel, J. Tinker, A. Berlind, L. da Costa, E. Kazin, Y. T. Lin, M. Maia *et al.*, Astrophys. J. **728**, 126 (2011).
[92] J. K. Parejko, T. Sunayama, N. Padmanabhan, D. A. Wake, A. A. Berlind, D. Bizyaev, M. Blanton, A. S. Bolton, F. van den Bosch, J. Brinkmann *et al.*, Mon. Not. R. Astron. Soc. **429**, 98 (2013).
[93] M. Kamionkowski, A. Kosowsky, and A. Stebbins, Phys. Rev. D **55**, 7368 (1997).
[94] J. Hartlap, P. Simon, and P. Schneider, Astron. Astrophys. **464**, 399 (2007).
[95] G. Efstathiou, Mon. Not. R. Astron. Soc. **349**, 603 (2004).
[96] F. Couchot, S. Henrot-Versillé, O. Perdereau, S. Plaszczynski, B. Rouillé d'Orfeuil, M. Spinelli, and M. Tristram, Astron. Astrophys. **602**, A41 (2017).
[97] C. García-García, D. Alonso, and E. Bellini, J. Cosmol. Astropart. Phys. 11 (2019) 043.
[98] P. Norberg, C. M. Baugh, E. Gaztañaga, and D. J. Croton, Mon. Not. R. Astron. Soc. **396**, 19 (2009).
[99] G. A. Marques, J. Liu, K. M. Huffenberger, and J. Colin Hill, Astrophys. J. **904**, 182 (2020).
[100] F. G. Mohammad and W. J. Percival, Mon. Not. R. Astron. Soc. **514**, 1289 (2022).
[101] A. Nicola, D. Alonso, J. Sánchez, A. Slosar, H. Awan, A. Broussard, J. Dunkley, E. Gawiser, Z. Gomes, R. Mandelbaum *et al.*, J. Cosmol. Astropart. Phys. 03 (2020) 044.
[102] M. Maus, S.-F. Chen, and M. White, J. Cosmol. Astropart. Phys. 06 (2023) 005.
[103] S.-F. Chen, Z. Vlah, and M. White, J. Cosmol. Astropart. Phys. 07 (2020) 062.
[104] S.-F. Chen, Z. Vlah, E. Castorina, and M. White, J. Cosmol. Astropart. Phys. 03 (2021) 100.
[105] A. Lewis, arXiv:1910.13970.
[106] Z. Sun, P. Zhang, F. Dong, J. Yao, H. Shan, E. Jullo, J.-P. Kneib, and B. Yin, Astrophys. J. Suppl. Ser. **267**, 21 (2023).
[107] J. H. Curtiss, Ann. Math. Stat. **12**, 409 (1941).
[108] H. Miyatake, M. S. Madhavacheril, N. Sehgal, A. Slosar, D. N. Spergel, B. Sherwin, and A. van Engelen, Phys. Rev. Lett. **118**, 161301 (2017).
[109] P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett *et al.* (Planck Collaboration), Astron. Astrophys. **594**, A13 (2016).
[110] Ž. Ivezić, S. M. Kahn, J. A. Tyson, B. Abel, E. Acosta, R. Allsman, D. Alonso, Y. AlSayyad, S. F. Anderson, J. Andrew *et al.*, Astrophys. J. **873**, 111 (2019).
[111] R. Laureijs, J. Amiaux, S. Arduini, J. L. Augueres, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet *et al.*, arXiv:1110.3193.
[112] D. Spergel, N. Gehrels, C. Baltay, D. Bennett, J. Breckinridge, M. Donahue, A. Dressler, B. S. Gaudi, T. Greene, O. Guyon *et al.*, arXiv:1503.03757.