Toward testing gravity with LSST using E_G

C. D. Leonard,[1,*] S. Alam,[2] R. Mandelbaum,[3] M. M. Rau,[1] S. Singh,[3,4] and C. M. A. Zanoletti[1]

(LSST Dark Energy Science Collaboration)

[1]*School of Mathematics, Statistics and Physics, Newcastle University, Newcastle upon Tyne, NE1 7RU, United Kingdom*
[2]*Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, India*
[3]*McWilliams Center for Cosmology and Astrophysics, Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA*
[4]*Berkeley Center for Cosmological Physics, University of California, Berkeley, California 94720, USA*

(Received 24 November 2025; accepted 29 January 2026; published 25 February 2026)

E_G is a summary statistic that combines cosmological observables to achieve a test of gravity that is relatively model independent. Here, we consider the power of a measurement of E_G using galaxy-galaxy lensing and galaxy clustering with sources from the Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST), and lenses from the Dark Energy Spectroscopic Instrument (DESI). We first update the theoretical framework for the covariance of E_G to accommodate this Stage IV scenario. We then demonstrate that E_G offers in principle a model-agnostic test of gravity using only linear-scale information, with the caveat that a careful treatment of galaxy bias is required. We finally address the persistent issue of E_G's theoretical dependence on the measured value of $\Omega_{\rm M}^0$. We propose a framework that takes advantage of the posterior predictive test to consistently incorporate uncertainty on $\Omega_{\rm M}^0$ in tests of gravity with E_G, which should be of general use beyond the LSST + DESI scenario. Our forecasting study using this method shows that the prior information available for $\Omega_{\rm M}^0$ is instrumental in determining the power of E_G in the LSST + DESI context. For the full survey dataset, with priors on $\Omega_{\rm M}^0$ from existing cosmic microwave background data, we find that for some modified gravity scenarios considered, we are likely to be able to reject the general relativity null hypothesis.

DOI: 10.1103/49vk-h19p

I. INTRODUCTION AND MOTIVATION

Current and future optical and infrared photometric and spectroscopic surveys provide a powerful opportunity to test the nature of gravity on cosmological scales. Stage III photometric surveys such as the Kilo Degree Survey [1], the Dark Energy Survey (DES) [2] and Hyper Suprime-Cam [3] have begun to explore the potential of weak lensing and galaxy clustering to place constraints on gravity [4–6], and the volume of such data will increase by an order of magnitude with the Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST) [7], as well as space-based Stage IV surveys Euclid [8], and the Nancy Grace Roman Space Telescope [9].

A number of different approaches are available when testing gravity using cosmological data. On one hand, we can seek to constrain the parameters of specific individual theories of interest via Bayesian parameter inference methods (see e.g. Refs. [10–12] and many more). This approach is highly valuable in that it promises a clear physical interpretation of the resulting parameter constraints. However, the downside of this approach is that there are many alternative theories of interest. Developing analysis tools and running analyses for each individual theory is a challenging and time-consuming process; covering the entire theory space in this manner is not likely to be feasible. We are therefore also interested in parallel approaches to testing gravity that seek to be model agnostic.

One model-agnostic approach in common use is to consider a general parametrization which captures a substantial subset of viable alternative gravity theories on linear scales (e.g. Refs. [13,14]). We then place constraints on the parameters that quantify deviations from general relativity (GR). This approach offers the benefit of

*Contact author: danielle.leonard@newcastle.ac.uk

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

constraining relatively general deviations from GR in a single analysis. At the same time, these efficiency gains come at the expense of the specificity and physical interpretability of testing individual theories. Notably, these model-agnostic parametrizations can typically only claim to describe a substantial subset of modified gravity theories in the linear regime of structure growth. In Stage III work, this issue has largely been addressed by imposing scale cuts such that only scales that can be modeled as linear are included in the analysis. However, the incredible statistical power of LSST will severely limit the scales which we will be able to model using linear theory only without risking biases to our parameter inference. Methods to address this issue are under active development [15–18].

A third option, which is the focus of this work, is to take the approach of performing a consistency check of general relativity. The E_G statistic [19] proposes to do this by strategically combining observables to achieve a measurement which is sensitive to only the Weyl potential and the gravitational potential, with sensitivity to nongravitational degrees of freedom eliminated. Since its original proposal, two different practical estimators for E_G have been employed, which differ in whether they probe the Weyl potential via galaxy-galaxy weak lensing [5,20–26] or cosmic microwave background (CMB) lensing [25,27–31]. Alternate proposals to measure E_G via intensity mapping of neutral hydrogen [32,33], using direct measurements of the Weyl potential [34], and incorporating velocity information via the kinematic Sunyaev-Zel'dovic effect [35] also exist in the literature. These various methods share the feature that they aim for E_G to be insensitive to σ_8 (which describes the amplitude of the matter power spectrum), and to galaxy bias, both of which typically present significant degeneracies with gravitational parameters in other methods.

One issue which has become increasingly relevant to E_G tests of gravity is the fact that the GR prediction for E_G depends on the value of the matter density parameter today, $\Omega_{\rm M}^0$. As cosmological data have become more statistically powerful, we approach a regime where the measurement uncertainty on E_G as estimated from data is comparable to the theoretical uncertainty on the GR prediction for E_G due to uncertainty on our estimates of $\Omega_{\rm M}^0$ (see Ref. [23] for a discussion). A consistent statistical framework is required to support the robust interpretation of E_G measurements in the context of this theoretical uncertainty.

In this work, we explore what will be required to make a meaningful measurement of E_G using LSST weak lensing source galaxies together with a spectroscopic galaxy sample from the Dark Energy Spectroscopic Instrument (DESI). This paper is structured as follows: in Sec. II, we review the E_G statistic and its constituent observables as well as other required theoretical background, and we describe the assumed LSST + DESI setup which we consider. In Sec. III, we build on existing work in the Stage III context to develop the required covariance modeling framework. In Sec. IV we explore the sensitivity of E_G to nonlinear structure formation, including nonlinear galaxy bias, and discuss our method for calibrating the effect of the latter. We then introduce in Sec. V the posterior predictive test framework as applied to E_G with uncertain $\Omega_{\rm M}^0$, and determine the power of E_G to test gravity within an LSST + DESI context in several scenarios of true underlying deviations from GR. We discuss and conclude in Sec. VI.

II. THEORY AND SETUP

In this section, we first introduce the E_G statistic and provide theoretical expressions for its constituent observables in its galaxy lensing based formulation. We then review the required theoretical basis for the gravity theories we will consider in this work. Finally, we provide the detailed assumed configuration of the LSST and DESI survey setups that we consider.

For the purposes of this section and the remainder of this work, we use the scalar-perturbed Friedmann-LeMaître-Robertson-Walker metric with the following convention:

$$ds^2 = -(1+2\Psi)dt^2 + a(t)^2(1-2\Phi)\delta_{ij}dx_i dx_j. \quad (1)$$

A. E_G: Theoretical background

The original estimator for the E_G statistic was proposed in Ref. [19] as

$$E_G = \frac{C_{\kappa g}(l,\Delta l)}{3H_0^2 a^{-1}\sum_\alpha f_\alpha(l,\Delta l)P_{vg}^\alpha}, \quad (2)$$

where l is the magnitude of the two-dimensional on-sky Fourier-space wave number, $C_{\kappa g}(l,\Delta l)$ is the angular cross-spectrum of weak lensing convergence and galaxy positions in bins of Δl, P_{vg}^α is the cross-spectrum of velocities and galaxy positions between three-dimensional Fourier wave number k_α and $k_{\alpha+1}$, and f_α is a weighting function which converts P_{vg}^α to an angular power spectrum (not to be confused with f, the linear growth rate of structure). The value of the Hubble constant today is denoted by H_0 and a is the scale factor as in Eq. (1).

The theoretical expectation for E_G is then given by

$$E_G = \left(\frac{\nabla^2(\Psi+\Phi)}{3H_0^2 a^{-1} f\delta_{\rm M}}\right)_{k=l/\bar{\chi},\bar{z}}, \quad (3)$$

where in this case f is the linear growth rate of structure, $\bar{\chi}$ is the comoving distance corresponding to the weighted average redshift of the lens galaxy sample $\bar{z}$, $\delta_{\rm M}$ is the matter density contrast, and Φ and Ψ are the potentials which feature in the scalar-perturbed metric of Eq. (1).

In GR, and assuming scale-independent galaxy bias, the modeling of E_G can be simplified to [20]

$$E_G = \frac{\Omega_{\rm M}^0}{f(z)} \approx \frac{\Omega_{\rm M}^0}{(\Omega_{\rm M}(z))^\gamma}, \tag{4}$$

where $\gamma = 0.55$, $\Omega_{\rm M}(z)$ is the fractional energy density of matter, and $\Omega_{\rm M}^0$ is its value today. We can approximate $\Omega_{\rm M}(z)$ in a flat Universe at late times as

$$\Omega_{\rm M}(z) = \frac{\Omega_{\rm M}^0(1+z)^3}{\Omega_{\rm M}^0(1+z)^3 + (1-\Omega_{\rm M}^0)}. \tag{5}$$

In GR, E_G is thus expected to be a constant with respect to angular or projected radial separation, given by Eq. (4).

In seeking to practically apply E_G using galaxy-galaxy lensing and galaxy clustering data, Ref. [20] introduces an alternate estimator for E_G in real space:

$$E_G(r_p) = \frac{\Upsilon_{gm}(r_p)}{\beta \Upsilon_{gg}(r_p)}, \tag{6}$$

where r_p is projected radial separation, and β is an observable quantity to be extracted from redshift space clustering measurements in spectroscopic galaxy surveys (see e.g. Ref. [36]). From a modeling perspective, it is given by the ratio of the linear growth rate of structure f (at the redshift of the lens galaxy sample) to the linear galaxy bias (of the lens galaxy sample). $\Upsilon_{gm}(r_p)$ and $\Upsilon_{gg}(r_p)$ are annular differential surface density observables for galaxy-galaxy lensing and projected galaxy clustering, introduced in Ref. [37].

We now present in detail how we will compute these quantities in this work from a modeling perspective. We begin with $\Upsilon_{gm}(r_p)$, which is defined by convention as in Ref. [37] as

$$\Upsilon_{gm}(r_p) = \Delta\Sigma_{gm}(r_p) - \left(\frac{r_p^0}{r_p}\right)^2 \Delta\Sigma_{gm}(r_p^0), \tag{7}$$

where r_p^0 is selected to remove information from the smallest, most theoretically uncertain scales, and $\Delta\Sigma_{gm}(r_p)$ is the differential surface mass density at projected radius r_p. It is defined as

$$\Delta\Sigma_{gm}(r_p) = \bar{\Sigma}_{gm}(< r_p) - \Sigma_{gm}(r_p), \tag{8}$$

where $\Sigma_{gm}(r_p)$ is the projected surface mass density and $\bar{\Sigma}_{gm}(< r_p)$ is the same quantity averaged over scales less than or equal to r_p.

$\Delta\Sigma_{gm}(r_p)$ is related to the tangential shear (γ_t) of background source galaxies about lens galaxies. Within GR, and in the hypothetical and simplified case where we consider the tangential shear of a single source galaxy at z_s and a single lens galaxy at z_l associated with a surface mass distribution at the same redshift, we have

$$\Delta\Sigma_{gm}(r_p, z_l) = \gamma_t(r_p, z_l, z_s)\Sigma_{\rm c}(z_l, z_s), \tag{9}$$

where $\Sigma_{\rm c}$ is the so-called critical surface density. It is defined via

$$\Sigma_{\rm c}^{-1} = \frac{4\pi G}{c^2} \frac{(\chi_s - \chi_l)(1+z_l)\chi_l}{\chi_s}, \tag{10}$$

when $\chi_s \geq \chi_l$, and 0 otherwise. χ_s and χ_l are the comoving distances corresponding to the source and lens redshifts and c is the speed of light in a vacuum.

A typical, simplified estimator for $\Delta\Sigma_{gm}$ for a sample of lens and of source galaxies is given schematically by [38]

$$\Delta\Sigma_{gm} = \frac{\sum_{ls} w_{ls} \gamma_t^{ls} \Sigma_{c,ls}}{\sum_{ls} w_{ls}}, \tag{11}$$

estimated in bins in r_p, where sums are over lens-source pairs. The lens-source pair weights, w_{ls}, are typically chosen to downweight lens and source galaxies close in redshift space (see, e.g., Refs. [39,40]), as follows:

$$w_{ls} = \Sigma_{c,ls}^{-2}. \tag{12}$$

This adjusts Eq. (11) to

$$\Delta\Sigma_{gm} = \frac{\sum_{ls} \gamma_t^{ls} \Sigma_{c,ls}^{-1}}{\sum_{ls} \Sigma_{c,ls}^{-2}}. \tag{13}$$

In practice, weights will also include a per-source-galaxy factor related to measurement noise, and may also include terms related to the observational selection function of the lens sample. In this work, we make the simplifying assumption that these terms are uncorrelated with redshift and therefore that we can neglect them; including them would simply involve adding an additional redshift-dependent multiplier to Eq. (12).

We can then write a more theoretically oriented version of Eq. (13), to allow us to model $\Delta\Sigma_{gm}$ in terms of tangential shear and critical density:

$$\Delta\Sigma_{gm}(r_p) = \frac{1}{\bar{w}} \int dz_l \frac{dN_l}{dz_l} \int dz_s \frac{dN_s}{dz_s} \times \gamma_t(r_p, z_s, z_l)\Sigma_c(z_s, z_l)^{-1}, \tag{14}$$

where dN_s/dz_s and dN_l/dz_l are the redshift distributions of the source and lens galaxy samples respectively, and $\bar{w}$ is

$$\bar{w} = \int dz_l \frac{dN_l}{dz_l} \int dz_s \frac{dN_s}{dz_s} \Sigma_c^{-2}(z_s, z_l). \tag{15}$$

Equation (14) requires also an expression for $\gamma_t(r_p, z_s, z_l)$, which is given by (see e.g. Ref. [41])

$$\gamma_t(r_p, z_s, z_l) = -\frac{3H_0^2\Omega_{\rm M}^0}{2c^2}\int_0^{\chi_s} d\chi \frac{\chi(\chi_s-\chi)(1+z(\chi))}{\chi_s}\frac{d}{d\ln r_p}\int_0^{r_p}\frac{dr'_p r'_p}{r_p^2}\xi_{gm}\left(\sqrt{(r'_p)^2+(\chi-\chi_l)^2}, z_l\right), \tag{16}$$

where ξ_{gm} is the three-dimensional two-point position cross-correlation function between the galaxy and matter fields, and we have invoked the small-angle approximation.

We then propagate Eq. (14) for $\Delta\Sigma_{gm}(r_p)$ into an expression for measured $\Upsilon_{gm}(r_p)$. To do so, we follow the procedure of Ref. [41], defining $\Pi \equiv \chi - \chi_l$ and combining Eqs. (7), (14), (15), and (16). We find

$$\Upsilon_{gm}(r_p) = \frac{\rho_{\rm c}^0\Omega_M^0}{\bar{w}}\frac{4\pi G}{c^2}\int dz_l \frac{dN}{dz_l}\int d\Pi(\chi_l(z_l)+\Pi)\left[z(\Pi+\chi_l(z_l))+1\right]\int_{z_l+\Pi}^{z_s^{\rm max}} dz_s \frac{dN}{dz_s}\frac{(\chi_s(z_s)-\chi_l(z_l)-\Pi)}{\chi_s(z_s)}\Sigma_{\rm c}^{-1}(z_s,z_l)$$
$$\times\left[\frac{2}{r_p^2}\int_{r_p^0}^{r_p} r'_p dr'_p \xi_{gm}\left(\sqrt{(r'_p)^2+\Pi^2}, z_l\right) - \xi_{gm}\left(\sqrt{r_p^2+\Pi^2}, z_l\right)\right) + \left(\frac{r_p^0}{r_p}\right)^2 \xi_{gm}\left(\sqrt{(r_p^0)^2+\Pi^2}, z_l\right)\Bigg], \tag{17}$$

where $\rho_{\rm c}^0$ is the critical density today. Note that we have used the identity,

$$r_p\frac{d}{dr_p}\left(\frac{2}{r_p^2}\int_0^{r_p} dr'_p r'_p f(r'_p)\right) = 2\left(f(r_p) - \frac{2}{r_p^2}\int_0^{r_p} dr'_p r'_p . f(r'_p)\right). \tag{18}$$

The integral over Π in Eq. (17) is technically from $-\chi_l$ to $\chi_\infty - \chi_l$ (where χ_∞ is the comoving distance at $z=\infty$), however in practice any suitably long integration length to capture the correlation behavior around the lens sample will produce the correct value. We use a projection extent of $\{-\Pi_{\rm max}, \Pi_{\rm max}\}$ where $\Pi_{\rm max}$ is the line-of-sight comoving distance between $z=0.4$ and $z=1.0$, corresponding to the extent of our DESI lens sample as described in detail below.

Having now an expression which can be used for theoretically computing $\Upsilon_{gm}(r_p)$, we move on to $\Upsilon_{gg}(r_p)$, which is the equivalent annular differential surface density for projected galaxy clustering. It is written as

$$\Upsilon_{gg}(r_p) = \rho_{\rm c}^0\left(\frac{2}{r_p^2}\int_{r_p^0}^{r_p} dr'_p r'_p w_{gg}(r'_p) - w_{gg}(r_p) + \left(\frac{r_p^0}{r_p}\right)^2 w_{gg}(r_p^0)\right), \tag{19}$$

where the projected galaxy clustering correlation function $w_{gg}(r_p)$ is defined as

$$w_{gg} = \int d\Pi \xi_{gg}(r_p, \Pi) \tag{20}$$

and $\xi_{gg}(r_p, \Pi)$ is the three-dimensional galaxy clustering position-position two-point correlation function. We project it over the line of sight here because we can only observe redshift space, where ξ_{gg} is anisotropic, and so write it in terms of the line-of-sight separation Π and the projected radial separation r_p. The integral over Π is theoretically over the full extent of the line of sight. In practice, for the cases considered in this work, we find that integrating over a range of $\Pi = \{-900, 900\}$ Mpc/h achieves numerical convergence.

For convenience, we define a function similar in character to $\Delta\Sigma_{gm}(r_p)$:

$$\Delta\Sigma_{gg}(r_p) = \rho_{\rm c}^0\left(\bar{w}_{gg}(< r_p) - w_{gg}(r_p)\right), \tag{21}$$

such that

$$\Upsilon_{gg}(r_p) = \Delta\Sigma_{gg}(r_p) - \left(\frac{r_p^0}{r_p}\right)^2 \Delta\Sigma_{gg}(r_p^0). \tag{22}$$

Note that in this work, we assume that lens galaxies have negligible redshift uncertainty due to being drawn from a spectroscopic sample.

The final component of E_G as defined in Eq. (6) is $\beta(z)$, defined as

$$\beta(z) = \frac{f(z)}{b}, \tag{23}$$

where b is the large scale galaxy bias and $f(z)$ is the linear growth rate of structure, defined as $f = \frac{d\ln D(a)}{d\ln a}$ where $D(a)$ is the linear growth factor of structure. In GR, $f(z)$ can be calculated theoretically by solving

$$\frac{{\rm d}^2 D(a)}{{\rm d}t^2} + 2H\frac{{\rm d}D(a)}{{\rm d}t} - \frac{3}{2}H^2\Omega_{\rm M}(a)D(a) = 0, \tag{24}$$

where t is physical time.

Note that observationally, β is extracted from measurements of the redshift-space two-point correlation function of galaxy positions. Redshift-space effects render this two-point function anisotropic and imprint information about

the growth rate of structure; β is then extracted via modeling of the anisotropic two-point function, accounting for these effects (see Ref. [42] for the linear-theory case). Although the models used for this purpose may not be fully independent of the theory of gravity (particularly when going beyond linear theory), it has been shown that large-scale signatures of alternative theories of gravity can nevertheless be detected via β [43].

B. Gravitational theories

In this work, we forecast the potential of the combination of LSST sources with DESI lenses to test gravity with E_G. To achieve this, we consider scenarios where an alternative theory in fact describes the Universe. We use two alternative theories of gravity for this purpose: Hu-Sawicki $f(R)$ gravity and normal-branch Dvali-Gabadadze-Porrati (nDGP) gravity. These are selected due to their common use as "workhorse" example theories in cosmology, representing as they do two different screening mechanisms and having a significant body of theoretical and numerical tools to compute quantities of interest. We now provide a review of the basic elements of each theory required in this work.

1. nDGP gravity

Dvali-Gabadadze-Porrati (DGP) gravity belongs to the (higher dimensional) braneworld class of modified gravity models. In DGP, our Universe is modeled as a lower-dimensional brane embedded in five-dimensional Minkowski space. The DGP action is [44]

$$\mathcal{S} = \frac{1}{16\pi G_{(5)}}\int \mathrm{d}^5x\sqrt{-g_{(5)}}R_{(5)} + \int \mathrm{d}^4x\sqrt{-g}\left(\frac{R}{16\pi G}+\mathcal{L}_m\right), \tag{25}$$

where the subscript (5) indicates the five-dimensional counterpart to the usual four-dimensional quantities. The crossover radius $r_c = \frac{G_{(5)}}{2G}$ is a transition scale between the large scales ($\geq r_c$) where five-dimensional dynamics dominate and the small scales ($\leq r_c$), where gravity looks four-dimensional. The strength of deviation from GR in this paper will be characterized by the parameter $\Omega_{\rm rc} \equiv \frac{1}{4H_0^2 r_c^2}$, with GR limit $\Omega_{\rm rc} \to 0$.

The DGP Hubble parameter, neglecting radiation, is given by [45,46]

$$H(a) = \sqrt{\Omega_{\rm M}^0 a^{-3} + \Omega_{rc}} \pm \sqrt{\Omega_{\rm rc}}. \tag{26}$$

The $(-)$ solution of the expression above is called nDGP—this is the solution of interest in this paper. Compared to its self-accelerating $(+)$ counterpart, it is free from ghost instabilities. However, in order to explain late-time acceleration, an additional dark energy component (and therefore an additional degree of freedom) has to be included in the background modeling. This fine-tuned component depends both on the modified gravity parameter and the cosmological constant, as follows:

$$\Omega_{\rm DE}(a) = \Omega_\Lambda^0 - 2\sqrt{\Omega_{\rm rc}(\Omega_{\rm M}^0 a^{-3} + \Omega_\Lambda^0)}. \tag{27}$$

This model provides a minimal extension to Λ cold dark matter (CDM) that displays Vainshtein screening and a scale-independent modification to linear structure growth. In the quasistatic regime, we can find the nDGP growth rate by solving the following equation:

$$\frac{\mathrm{d}^2D(a)}{\mathrm{d}t^2} + 2H(a)\frac{\mathrm{d}D(a)}{\mathrm{d}t} - \frac{3}{2}H(a)^2\Omega_{\rm M}(a)\mu(a)D(a) = 0, \tag{28}$$

with initial conditions $D(a) = a$ and $\frac{\mathrm{d}D(a)}{\mathrm{d}t} = 1$ during matter domination. Here G denotes Newton's constant, H is the time-evolving Hubble parameter, and $\mu(a)$ is defined as [47]

$$\mu(a) = 1 + \frac{1}{3\left[1+\frac{H(a)}{\Omega_{\rm rc}}\left(1+\frac{a}{3H(a)}\left(\frac{\mathrm{d}H(a)}{\mathrm{d}a}\right)\right)\right]}. \tag{29}$$

Comparing Eq. (28) with Eq. (23) in the GR case, we can easily see the change induced in the linear growth of structure.

We can find the three-dimensional linear matter power spectrum in nDGP gravity (under the continued assumption that we are in the linear and quasistatic regime) via

$$P_{\delta,\rm lin}^{\rm nDGP}(k,a) = \left(\frac{D_{\rm nDGP}(a)\times D_{\rm GR}(a_{\rm MD})}{D_{\rm nDGP}(a_{\rm MD})\times D_{\rm GR}(a)}\right)^2 P_{\delta,\rm lin}^{\rm GR}(k,a), \tag{30}$$

where $a_{\rm MD}$ is a value of the scale factor during matter domination.

To find the three-dimensional nonlinear matter power spectrum in nDGP gravity, we use the `COLA`-based `nDGPEmu` [48] emulator. Compared to `MG-AREPO` full N-body simulations, it is 3% accurate for $k = 5\ h/\mathrm{Mpc}$ and $0 \leq z \leq 2$ [49].

2. Hu-Sawicki $f(R)$ gravity

The $f(R)$ class of gravity theories introduces an action that is beyond-first order in the Ricci scalar R, as follows:

$$S_G = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{R+f(R)}{16\pi G}\right]. \tag{31}$$

The additional function $f(R)$ introduces a light scalar degree of freedom $f_R \equiv \mathrm{d}f/\mathrm{d}R$, which leads to a

long-range fifth force. We consider more specifically the Hu-Sawicki model [50], with characteristic function

$$f(R) = -m^2 \frac{c_1 (R/m^2)^n}{c_2 (R/m^2)^n + 1}, \tag{32}$$

where m is a mass-scale parameter and c_1, c_2 and n are dimensionless free parameters. m is taken to be

$$m^2 = \frac{\Omega_{\rm M}^0 H_0^2}{c^2}. \tag{33}$$

In the high-curvature limit ($R \gg m^2$), and fixing $\frac{c_1}{c_2} = 6\frac{\Omega_\Lambda^0}{\Omega_{\rm M}^0}$ where Ω_Λ^0 is the fractional energy density of dark energy today), we can write [51]

$$f(R) = -2\Lambda - \frac{f_{R0}}{n}\frac{R_0^{n+1}}{R^n}, \tag{34}$$

with $f_{R0} = f_R(R_0)$ and $R_0 = R(z=0)$.

In this paper, we will be looking at this version of the theory, with $n = 1$ and with a ΛCDM background. The strength of deviation from GR is then uniquely characterized by the parameter f_{R0}, with GR limit $f_{R0} \to 0$. Under the quasistatic approximation (QSA), with a similar approach as for nDGP, we can find the growth rate by solving Eq. (28), with

$$\mu(k,a) = 1 + \frac{(k/a)^2}{3\left((k/a)^2 + \frac{H_0^2(\Omega_{\rm M}^0 a^{-3} + 4(1-\Omega_{\rm M}^0))^3}{2f_{R0}(4-3\Omega_{\rm M}^0)^2}\right)}, \tag{35}$$

and similarly for the three-dimensional linear matter power spectrum under QSA,

$$P_{\delta,\rm lin}^{f(R)}(k,a) = \left(\frac{D_{f(R)}(k,a) \times D_{\rm GR}(k,a_{\rm MD})}{D_{f(R)}(k,a_{\rm MD}) \times D_{\rm GR}(k,a)}\right)^2 P_{\delta,\rm lin}^{\rm GR}(k,a). \tag{36}$$

We compute the three-dimensional nonlinear matter power spectrum for $f(R)$ gravity by using the `e-MANTIS` emulator [52]. When compared to full N-body simulations, it shows a maximum deviation of 3% for $k \approx 7$ h/Mpc and $0 \le z \le 2$ [51].

C. Assumed survey setup and other fiducial parameters

We now describe the assumed setup of the forecast LSST + DESI E_G measurements considered in this work. We consider an observational scenario in which the lensing source sample is from LSST (we consider both Year 1 and Year 10 scenarios), and the spectroscopic galaxy clustering and lens sample is the DESI-I Luminous Red Galaxy (LRG) sample, where DESI-I is the initial 5-year DESI program which commenced in 2021. We assume an overlap area of 5000 deg^2 for these two samples. This number falls between the "baseline" and the "larger extragalactic footprint" scenarios for LSST observing strategy as discussed in Ref. [53].

1. DESI LRGs

The DESI-I LRG sample is expected to have a comoving volume density of $\bar{n} = 5 \times 10^{-4}$ $(\rm Mpc/h)^{-3}$ over a redshift range of $z = 0.4$–0.8 [54]. We take the redshift distribution of the DESI-I LRG sample to be as given in Ref. [54], which we display in Fig. 1, making the approximation that the above number density holds across the full redshift range of the sample. The mean redshift of this sample is 0.72. As we will see below, for the covariance calculation we must set the line-of-sight extent of the lens sample; we approximate this as being the comoving distance between $z = 0.4$ and $z = 1.0$ for the DESI LRGs as this is where the

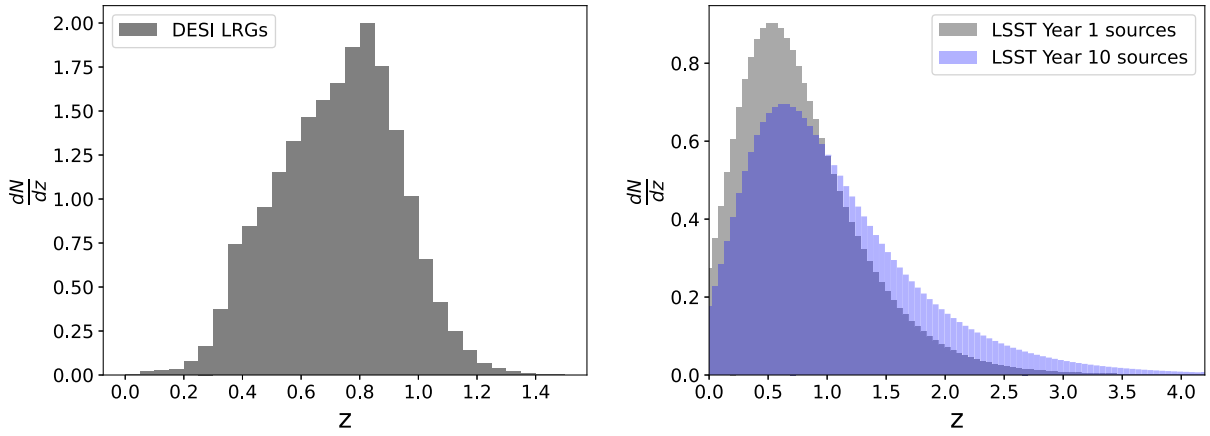


FIG. 1. Left: redshift distribution of DESI-I LRG sample, data points from Ref. [54]. Right: redshift distribution of LSST Year 1 and LSST Year 10 source sample. Numerical values of $\frac{dN}{dz}$ reflect in all cases the fact that distributions are normalized to unity.

majority of the number density falls. It is worth noting that Ref. [26] measures E_G using DESI LRG lenses with sources from Stage III lensing surveys; the DESI sample considered there does not perfectly map onto our case, as it includes DESI Data Release 1 only and has different lens-source sky overlap. Nevertheless we refer the reader to that work for an E_G measurement with DESI LRG lenses.

2. LSST source galaxies

The weak lensing source galaxies are assumed to correspond to the LSST Year 1 or Year 10 source sample as appropriate, defined in the LSST Dark Energy Science Collaboration (DESC) Science Requirements Document [55].

We follow the method of Ref. [55] and model the redshift distribution of the source galaxies by defining a reference distribution and convolving with a model which accounts for photometric redshift uncertainties. Specifically, the reference source galaxy redshift distribution is modeled with an analytic form given by

$$\frac{dN}{dz} \propto z^2 \exp\left(\frac{z}{z_0}\right)^{\alpha}, \tag{37}$$

with $z_0 = 0.13$ and $\alpha = 0.78$ for Year 1, and $z_0 = 0.11$, $\alpha = 0.68$ for Year 10. We model the effect of photometric redshift uncertainty by convolving with a Gaussian distribution with the form

$$p(z, \tilde{z}) \propto \exp\left(-\frac{1}{2}\left(\frac{z - (\tilde{z} + \Delta\bar{z})}{\sigma_z(1+z)}\right)^2\right), \tag{38}$$

where σ_z and $\Delta\bar{z}$ are parameters of the photo-z uncertainty model, which will in our analysis take fiducial values of $\sigma_z = 0.05$ and $\Delta\bar{z} = 0$ for both Year 1 and Year 10.

The resulting source galaxy distribution is then given by

$$\frac{dN}{d\tilde{z}} = \frac{\int dz \frac{dN}{dz} p(z, \tilde{z})}{\int d\tilde{z} \int dz \frac{dN}{dz} p(z, \tilde{z})} \tag{39}$$

and is displayed in Fig. 1.

We assume a surface density of source galaxies of $n_{\rm eff} = 10$ galaxies per arcmin^2 in Year 1 and $n_{\rm eff} = 27$ galaxies per arcmin^2 in Year 10. We take the shape-noise parameter $\sigma_\gamma = 0.26$ for both Year 1 and Year 10.

3. Fiducial cosmological and other parameters

We take our fiducial cosmological parameters to match those of the simulation suite which we use for estimating covariances, described in more detail in Sec. III B below [56]. Explicitly, our fiducial cosmological parameters are $\{h = 0.69,\ \Omega_{\rm B}^0 = 0.022h^2 = 0.04621,\ \Omega_{\rm M}^0 = 0.292,\ n_s = 0.965,\ A_s = 2.115 \times 10^{-9},\}$. This value of A_s is equivalent to $\sigma_8 = 0.82$ within ΛCDM + GR with these other fiducial cosmological parameters.

When we adopt a linear galaxy bias model below, we specifically make use of a single redshift-independent bias parameter for the DESI LRG lens sample, defined such that $\delta_{\rm g} = b\delta_{\rm m}$ where $\delta_{\rm g}$ is the galaxy overdensity and $\delta_{\rm m}$ is the overdensity for all matter. We take a fiducial linear bias value of $b = 2.33$ for the DESI LRG sample (following Ref. [57] as discussed further below in Sec. IV, together with details of higher-order bias parameters where relevant). This means that the fiducial linear bias differs slightly from the linear bias as computed from the simulations used to estimate the covariance ($b = 2.03$). It has been shown that using a covariance matrix computed at fixed parameter values somewhat discrepant from the true underlying parameter values has no significant effect in cosmological analyses [58], thus, we do not expect any significant impact on our results from this mild discrepancy in galaxy bias value.

We follow Refs. [5,20] in fixing $r_p^0 = 1.5$ Mpc/h, although note that Ref. [5] tested the impact of varying r_p^0 for appropriate small separations and found no effect.

III. COVARIANCE ESTIMATION AND PROPERTIES

Existing measurements of E_G via galaxy-galaxy lensing take a variety of approaches to covariance matrix estimation, including via mock galaxy catalogs, jackknife sampling, analytic methods, or some combination thereof.

A common theme in Stage III E_G covariance estimation is the assumption that the three composite probes of E_G are independent of each other [5,22,23,25,26]. For Υ_{gm} and Υ_{gg}, this is argued to be the case because the covariance of Υ_{gm} is dominant (and dominated itself by shape noise, which is independent of clustering). Additionally, for some measurements, the sky area covered by the source sample is only a small subset of that covered by the lens sample. In the cases of covariance both between Υ_{gm} and β, and between Υ_{gg} and β, covariance in the signals is expected to be small and is often neglected [30,32,33]. This is because $\Upsilon_{gm}(r_p)$ and $\Upsilon_{gg}(r_p)$ are sensitive to modes perpendicular to the line of sight, while β is sensitive to line-of-sight information as extracted from the redshift-space power spectrum [42].

Although these assumptions are well founded for the Stage III measurements cited, it is not *a priori* obvious that they will hold in the LSST + DESI case. Shape noise in a galaxy-galaxy lensing measurement is suppressed by the number of lens-source galaxy pairs (and similarly for shot noise in a clustering measurement); at the number of galaxies available in LSST and DESI, shape and shot noise may be subdominant to cosmic variance on some scales. Furthermore, the use of the same sample of lens galaxies to measure all three constituent probes may result in some non-negligible covariance between Υ_{gm} and β, or else

between Υ_{gg} and β, due to the fact that the same galaxy sample is being used within the measurement of all three observables. At the Stage IV level considered here, this may be an important covariance contribution.

In this section, we present methods for estimating the covariance between constituent probes of E_G based in both analytic methods and simulations, and discuss the extent to which the above assumptions can be carried forward in the Stage IV context explored here.

A. Analytic covariance estimation: Theoretical background

We first present analytic expressions for the covariance elements which arise between auto- and cross-combinations of $\Upsilon_{gm}(r_p)$ and $\Upsilon_{gg}(r_p)$. Although in this work we use the analytic covariances presented here only as a point of comparison for our covariance matrix estimated from simulations, we present the relevant expressions as theoretical background. We make the approximation of neglecting all non-Gaussian contributions from the analytic expressions in this section; it is, however, straightforward to extend the below to incorporate non-Gaussian terms given the ability to compute these for the standard 3×2 point multiprobe analysis case.

We begin with the covariances of $\Delta\Sigma_{gx}$ quantities which can then be propagated to Υ_{gx} (with x being m or g). We follow closely here the work of Ref. [59] and refer the reader also to Ref. [5] for a similar analytic treatment.

Considering first the covariance of $\Delta\Sigma_{gm}(r_p)$, we have (see Ref. [59] in the Gaussian limit)

$$\mathrm{Cov}\left[\Delta\Sigma_{gm}(r_p),\Delta\Sigma_{gm}(r'_p)\right]=\frac{(\overline{\Sigma_c^{-2}})^{-1}}{V}\int\frac{kdk}{2\pi}J_2(kr_p)J_2(kr'_p)\left[\left(P_{g\kappa}(k)\right)^2+\left(P_{gg}(k)+\frac{1}{n_l}\right)\left(P_{\kappa\kappa}(k)+\frac{\sigma_\gamma^2}{n_s}\right)\right],\tag{40}$$

where n_l is the volume density of lens galaxies, n_s is the effective surface density of source galaxies, σ_γ is the rms per-component shape noise, V is the volume of the lens galaxy sample and J_2 is the second-order spherical Bessel function. $\overline{\Sigma_c^{-2}}$ is defined as

$$\overline{\Sigma_c^{-2}}=\int dz_l\frac{dN}{dz_l}\int dz_{ph}\frac{dN}{dz_s}\Sigma_c^{-2}(z_l,z_s),\tag{41}$$

$P_{\kappa\kappa}(k)$ is given by

$$P_{\kappa\kappa}(k)=\int d\Pi\left(\int dz_s\frac{dN}{dz_s}\int dz_l\frac{dN}{dz_l}\times\frac{\rho_{\mathrm{m}}^0}{\Sigma_{\mathrm{c}}(\chi_s(z_s),\chi_l(z_l)+\Pi)}\right)^2P_\delta(k),\tag{42}$$

and similarly $(P_{g\kappa}(k))^2$ is

$$(P_{g\kappa}(k))^2=\int d\Pi\left(\int dz_s\frac{dN}{dz_s}\int dz_l\frac{dN}{dz_l}\times\frac{\rho_{\mathrm{m}}^0 b}{\Sigma_{\mathrm{c}}(\chi_s(z_s),\chi_l(z_l)+\Pi)}\right)^2P_\delta(k)^2,\tag{43}$$

where b is the galaxy bias, which could be redshift or scale dependent, and ρ_m^0 is the value of the matter density today. We note that the projection over Π is here over the entire line-of-sight distance relevant to the lensing measurement, reflecting the appropriate window function.

Similarly, for the covariance of $\Delta\Sigma_{gg}(r_p)$, we follow the derivation of Ref. [59] and take the Gaussian limit to obtain

$$\mathrm{Cov}\left[\Delta\Sigma_{gg}(r_p),\Delta\Sigma_{gg}(r'_p)\right]=\frac{2\Delta\Pi(\rho_{\mathrm{c}}^0)^2}{V}\int\frac{kdk}{2\pi}J_2(kr_p)\times J_2(kr'_p)\left(P_{gg}(k)+\frac{1}{n_l}\right)^2,\tag{44}$$

where $\Delta\Pi$ is the line-of-sight extent of the lens sample.

Finally, we once again follow an analogous derivation to those in Ref. [59] to obtain the covariance of $\Delta\Sigma_{gm}$ with $\Delta\Sigma_{gg}$, as follows:

$$\mathrm{Cov}\left[\Delta\Sigma_{gm}(r_p),\Delta\Sigma_{gg}(r'_p)\right]=\frac{2\rho_{\mathrm{c}}^0}{V}(\overline{\Sigma_c^{-1}})^{-1}\int\frac{kdk}{2\pi}J_2(kr_p)J_2(kr'_p)\bar{P}_{g\kappa}(k)\left(P_{gg}(k)+\frac{1}{n_l}\right),\tag{45}$$

where $\bar{P}_{g\kappa}$ differs from $P_{g\kappa}$ in Eq. (43) above in that it is projected over the window function suitable for one galaxy clustering measurement and one galaxy-galaxy lensing measurement, as follows:

$$\bar{P}_{g\kappa}(k)=\int_{\Pi_{\mathrm{min}}}^{\Pi_{\mathrm{max}}}d\Pi\left(\int dz_s\frac{dN}{dz_s}\int dz_l\frac{dN}{dz_l}\times\frac{\rho_{\mathrm{m}}^0 b}{\Sigma_{\mathrm{c}}(\chi_s(z_s),\chi_l(z_l)+\Pi)}\right)P_\delta(k),\tag{46}$$

where Π_{min} and Π_{max} are the limits of the lens sample along the line of sight.

Having now $\mathrm{Cov}[\Delta\Sigma_{gx}(r_p),\Delta\Sigma_{gy}(r'_p)]$ for each of x and y equal to g and m, we return to the definitions of $\Upsilon_{gm}(r_p)$

and $\Upsilon_{gg}(r_p)$ [Eqs. (7) and (19), respectively] to find

$$\mathrm{Cov}\big(\Upsilon_{gx}(r_p),\Upsilon_{gy}(r'_p)\big)=\mathrm{Cov}\big(\Delta\Sigma_{gx}(r_p),\Delta\Sigma_{gy}(r'_p)\big)-\left(\frac{r_p^0}{r_p}\right)^2\mathrm{Cov}\big(\Delta\Sigma_{gx}(r_p^0),\Delta\Sigma_{gy}(r'_p)\big)$$
$$-\left(\frac{r_p^0}{r'_p}\right)^2\mathrm{Cov}\big(\Delta\Sigma_{gx}(r_p),\Delta\Sigma_{gy}(r_p^0)\big)+\left(\frac{r_p^0}{r_p}\right)^2\left(\frac{r_p^0}{r'_p}\right)^2\mathrm{Cov}\big(\Delta\Sigma_{gx}(r_p^0),\Delta\Sigma_{gy}(r_p^0)\big),\quad(47)$$

where x and y are either g or m as appropriate. Explicit expressions for the averaging in r_p bins as required are found in Appendix A, along with a visualization of the correlation matrix estimated for this subset of terms using this analytic method for LSST sources and DESI LRG lenses (Fig. 11).

B. Covariance estimation via simulations

We can also estimate the covariance of the constituent probes of E_G via mock galaxy catalogues based on N-body simulations. In practice, we will use this method in this work. This is primarily due to the fact that the simulation-based method allows us to easily estimate the covariance between $\Upsilon_{gm}/\Upsilon_{gg}$ and β, for which an analytic formalism is not available in the literature. Where available, we make use of covariance matrices generated analytically as a point of comparison for those estimated from simulations (although we do not expect perfect agreement due to differences in the effects captured in each case).

To obtain the required mock galaxy catalogs, we use the N-body simulations described in Ref. [56] ("MedRes"), taking a redshift slice at $z=0.75$ to best match the DESI LRG sample redshift. We run the Rockstar [60] halo finder on this N-body snapshot and populate the resulting haloes with galaxies using a halo occupation distribution (HOD) model fit to the DESI LRG Buzzard simulations between redshift 0.7–0.8 (see Ref. [22] for the precise form of the HOD model). The result is a galaxy number density of 4.954×10^{-4} $(\mathrm{Mpc/h})^{-3}$, in excellent agreement with our assumed DESI LRG number density.

The simulation volume is subdivided into 100 jackknife regions. Given the mock galaxy catalogs, Υ_{gm} and Υ_{gg} are measured in r_p bins as surface mass densities using dark matter particles and galaxies. β is obtained by fitting a convolutional Lagrangian perturbation theory (LPT) model [61] between 30–80 Mpc/h. We then use standard expressions for jackknife covariance estimation to find a covariance matrix which captures the auto- and cross-variance between all three quantities, rescaling the resulting covariance to account for the difference in the simulation volume vs that expected for the LSST + DESI measurement under consideration.

We must then correct this matrix to account for the fact that although one can estimate Υ_{gm} easily from the simulations described as a cross-correlation between dark-matter particles and galaxies, this ignores a key feature of the practical observational estimate of Υ_{gm} via galaxy-galaxy lensing: shape noise. Fortunately, shape noise is independent of the other noise terms which are accounted for by the simulation. Therefore, we can construct an independent covariance matrix to capture this effect. We construct this shape-noise covariance matrix by considering only the final term in brackets of Cov $[\Delta\Sigma_{gm}(r_p),\Delta\Sigma_{gm}(r'_p)]$ [Eq. (40)], which yields

$$\mathrm{Cov}^{\mathrm{SN}}\left[\Delta\Sigma_{gm}(r_p),\Delta\Sigma_{gm}(r'_p)\right]=\frac{(\overline{\Sigma_c^{-2}})^{-1}\sigma_\gamma^2}{2\pi V n_l n_s}\frac{\delta^D(r_p-r'_p)}{r_p},\quad(48)$$

where δ^{D} indicates a Dirac δ function.

We propagate this expression to the Υ_{gm} case using Eq. (47) and average in r_p bins using Eq. (A2) given in Appendix A. Note that in practice, when computing this component, we assume that $\Delta\Sigma_{gx}(r_p^0)$ is measured as $\Delta\Sigma_{gx}$ in the r_p bin containing r_p^0. Because this noise component is independent of those captured by our jackknife-estimated covariance matrix, we can simply add the resulting contribution to the $\Upsilon_{gm}(r_p)$ autocorrelation block of the matrix.

Finally, note that because we use a covariance matrix estimate based on jackknife methods with a finite number of jackknife realizations, when inverting the covariance matrix, we must account for the fact that the inverse of a noisy covariance matrix is not an unbiased estimate for the true inverse covariance. This is accounted for by multiplying the resulting inverse covariance by the so-called Hartlap factor [62],

$$\mathrm{Cov}^{-1}=\frac{n-p-2}{n-1}\mathrm{Cov}_*^{-1},\quad(49)$$

where Cov_*^{-1} is the uncorrected inverted estimate of the covariance matrix, Cov^{-1} is the corrected version we will use, n is the number of jackknife realizations, and p is the length of the data vector.

The full correlation matrices for our $\{\Upsilon_{gm}(r_p),\Upsilon_{gg},\beta\}$ data vector, estimated via the method described in this subsection, are visualized in Fig. 2 for LSST Year 1 and Year 10 sources with DESI lenses. We see that, as we suggested may be the case at the beginning of this section,

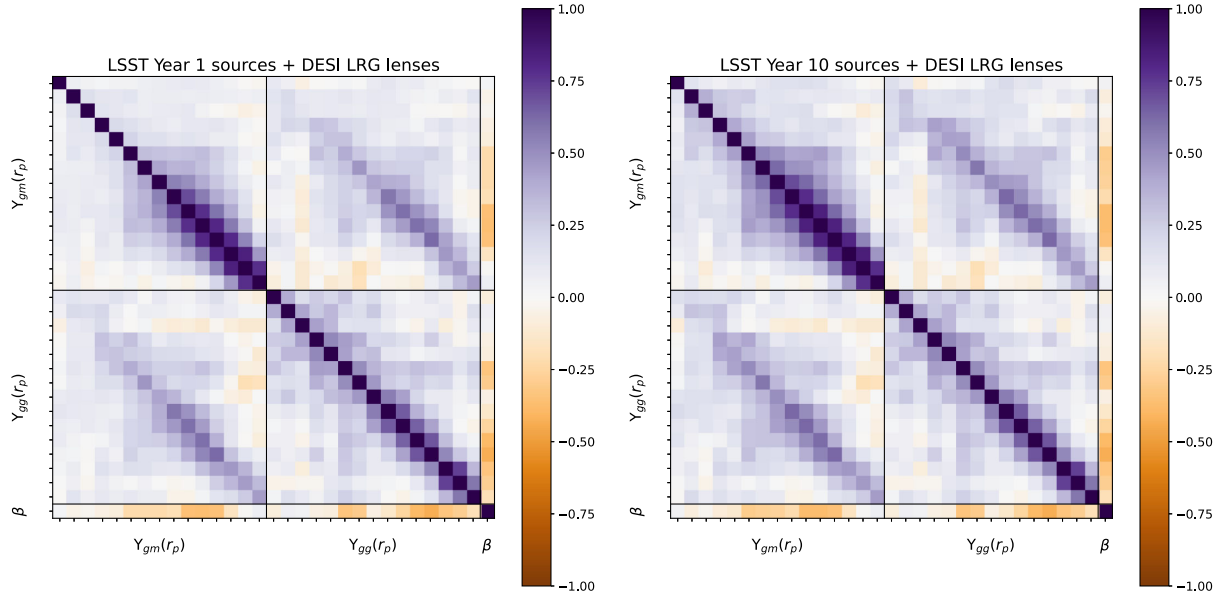


FIG. 2. Correlation matrices of the constituent probes of E_G. Within the $\Upsilon_{gm}(r_p)$ and $\Upsilon_{gg}(r_p)$ sub-blocks, bins increase logarithmically in r_p with bin centers ranging from 1.4 to 87.9 Mpc/h.

the matrix is not fully dominated by the shape noise associated with the source sample. Furthermore, contrary to conventional expectation for Stage III data and as alluded to above [30,32,33], we also find non-negligible $\mathrm{Cov}(\Upsilon_{gm},\beta)$ and $\mathrm{Cov}(\Upsilon_{gg},\beta)$.

C. $E_G(r_p)$ covariance

As can be readily seen from Eq. (3), $E_G(r_p)$ is a ratio of noisy quantities. In principle, the ratio of Gaussian-distributed quantities is not necessarily a Gaussian, but under some circumstances can be approximated as such. If the Gaussian approximation holds, we can carry on with the use of a covariance matrix for $E_G(r_p)$ to characterize an approximately Gaussian likelihood. We first therefore verify to what extent the Gaussian approximation holds for our case.

We construct a multivariate Gaussian using the covariance matrix of the constituent probes and the fiducial data vector. We then jointly draw samples of $\Upsilon_{gm}(r_p)$, $\Upsilon_{gg}(r_p)$, and β, then use these values to create a population of $E_G(r_p)$ samples. In each r_p bin (call this r_p^i), we examine the distribution of $E_G(r_p^i)$ and find that for all bins, the Gaussian approximation is sufficient for our purposes. Note that this diverges from recent work for the case of the CMB lensing E_G statistic, where it was found that E_G is not Gaussian-distributed [30]. The validity of the Gaussian likelihood approximation should be verified for each individual dataset as it will depend on the specific noise properties in question.

Turning again to our population of $E_G(r_p)$ samples, $\mathrm{Cov}(E_G(r_p),E_G(r_p'))$ can then be estimated using standard formulas. The correlation matrices for E_G in r_p bins are shown in Fig. 3 for LSST Year 1 and Year 10 sources with DESI LRG lenses. The correlation matrix demonstrates very significant off-diagonal covariance between r_p bins, which is further enhanced in the Year 10 case (where the shape noise is reduced).

IV. E_G AND NONLINEAR SCALES

A major issue facing Stage IV weak lensing and galaxy clustering tests of deviations from GR is the challenge of appropriately modeling nonlinear growth (and related effects such as nonlinear galaxy bias) outside of standard gravity. For constraints on parametrized modifications to GR in Stage III surveys, this has often been circumvented by restricting to only linear scales. In Stage IV, we expect this strategy to be highly suboptimal, as increased statistical precision of our measurements will make it much more restrictive (see, e.g. Ref. [17]). Strategies to overcome this issue are under development [15–18,63], but these do not map directly onto the E_G case, where the goal is not to constrain parameters but to perform a litmus test for GR.

In this section, we consider the robustness of an LSST + DESI E_G measurement to uncertainties in the modeling of nonlinear growth of structure. We examine the degree to which this depends on the underlying gravity theory and the effect of nonlinear galaxy bias.

We make use of the standard linear criterion (seen elsewhere in the literature in, for example, Refs. [4,6,64,65]) that scales are retained in an analysis for which it is true that

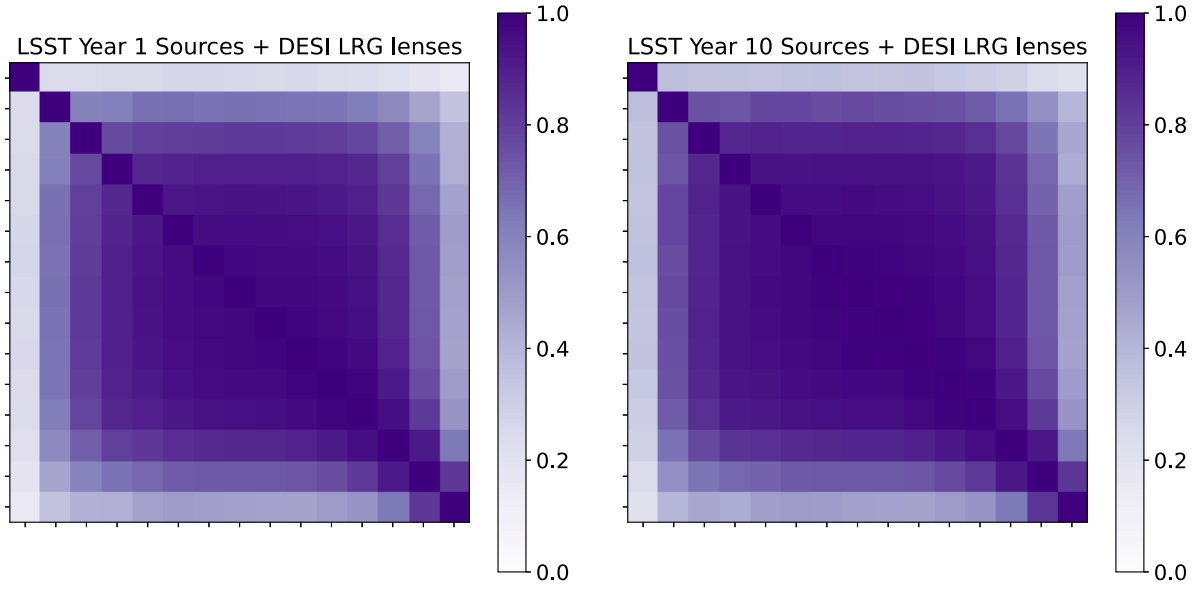


FIG. 3. Correlation matrices of $E_G(r_p)$. Bins increase logarithmically in r_p with bin centers ranging from 1.4 to 87.9 Mpc/h.

$$\Delta\chi^2_{\text{nonlin-lin}} = (\vec{D}_{\text{nonlin}} - \vec{D}_{\text{lin}})\text{Cov}^{-1}(\vec{D}_{\text{nonlin}} - \vec{D}_{\text{lin}}) \leq 1, \tag{50}$$

where $\vec{D}_{\text{nonlin}}$ is the fiducial theoretical data vector and $\vec{D}_{\text{lin}}$ is the fiducial data vector modeled with linear theory only. Cov is the data covariance matrix.

We follow literature convention in imposing the additional constraint that if a given data point is removed by the criteria of Eq. (50), so too should be all data points at smaller scales for the observable in question. This accounts for two main effects. First, at quasilinear scales, the nonlinear behavior of voids can lead to a "zero crossing" in $(\vec{D}_{\text{nonlin}} - \vec{D}_{\text{lin}})$ [66]. While this may technically mean that some intermediate-scale data points contribute less to $\Delta\chi^2$ than data points at higher and lower scales, the exact location of the zero-crossing is expected to be sensitive to nonlinear modeling, and so we cut these intermediate points. Second, there may be cases where higher variance and covariance at smaller scales mean that the criteria of Eq. (50) would allow us to retain data points at scales which are certainly subject to significant nonlinear effects. Although we could in principle retain these points, we opt to define our criteria such that they are removed to maximize interpretability.

Note also that before applying Eq. (50), following Refs. [22,5], we restrict our data vector to those points at $r_p \geq 5$ Mpc/h. We do so because in addition to uncertainties in the nonlinear growth of structure, we must also account for small-scale uncertainties due to baryonic physics. Although the use of Υ statistics attempts to remove the impact of scales below r_p^0, this is not expected to be sufficient to fully account for these small-scale baryonic uncertainties, and it is hence typical to restrict fits to $E_G(r_p)$ to a slightly higher range of r_p values.

A. Linear galaxy bias, general relativity

We initially consider the effect of imposing Eq. (50) in the simplest (and standard in the literature) case: with $\vec{D}_{\text{nonlin}}$ modeled as in GR with `halofit` [67], and assuming a linear galaxy bias model. Within our LSST + DESI setup, imposing Eq. (50) results in the same "linear-only" scale cuts for both LSST Year 1 and Year 10 source samples. We attribute this perhaps-counterintuitive agreement to the fact that as seen above in Fig. 2, the difference in shape noise between LSST Year 1 and Year 10 is overall a fairly small effect on the covariance structure. We illustrate this in Fig. 4. We see that we are able to retain the entire $E_G(r_p)$ data vector, as compared to its constituents $\Upsilon_{gm}(r_p)$ and $\Upsilon_{gg}(r_p)$ where we only retain a small number of larger-scale data points. This is because, as seen in Eq. (6), numerator and denominator of $E_G(r_p)$ both contain a single two-point function of the matter field. Assuming (as we have done here) a linear galaxy bias, this effectively means that nonlinear behavior in the matter field is "canceled out." This is both a curse and a blessing: it means that E_G is unlikely to be sensitive to canonical nonlinear smoking-gun behavior of alternative theories of gravity, like screening. On the other hand, it potentially offers a very clear way of honing in on linear-only physics in a test of modified gravity, which is a much sought-after ability.

B. Alternative theories of gravity

The validity of Eq. (50) as a linear-scale-cuts criterion relies on the nonlinear behavior being actually described by

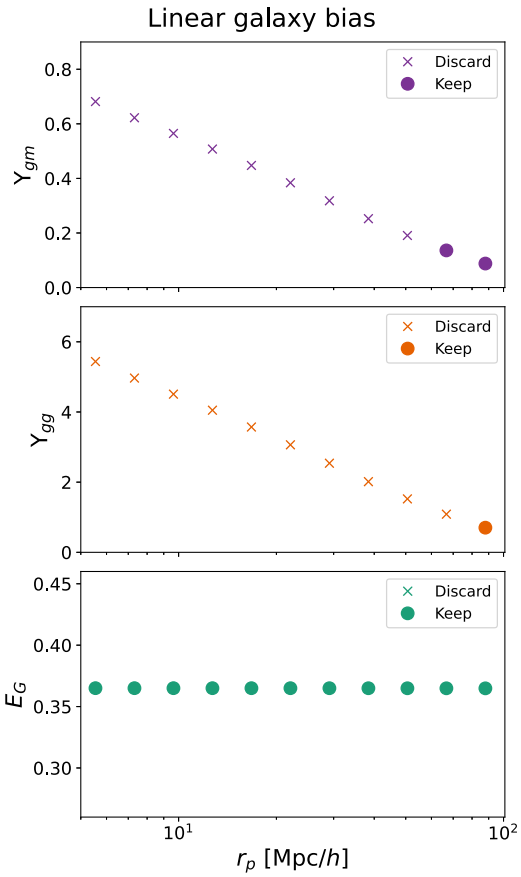


FIG. 4. Illustration of which points are kept in the linear-only scale cuts scenario for Υ_{gm} (top), Υ_{gg} (middle) and E_G (bottom), assuming a linear galaxy bias. The points plotted are for the fiducial data vector with nonlinear modeling from halofit (but linear galaxy bias), with LSST Year 1 sources. The r_p bins kept vs discarded are the same in both the LSST Year 1 and Year 10 source cases.

GR—or at minimum, the true nonlinear behavior becoming important at the same scales as in GR. To investigate whether allowing for nonlinear modeling in other theories of gravity alters the results of the previous subsection, we repeat the application of the criterion given in Eq. (50) using modeling of $\vec{D}_{\rm lin}$ and $\vec{D}_{\rm nonlin}$ in two different alternative theories of gravity (introduced in Sec. II B above):

(i) Hu-Sawicki $f(R)$ gravity: $f_{R0} = 10^{-4}$, $n = 1$.
(ii) nDGP gravity: $\Omega_{\rm rc} = 0.25$.

We select parameter values for Hu-Sawicki $f(R)$ and nDGP gravity which are significantly discrepant from GR (in the sense that the matter power spectrum under these theories and at these parameter values deviates from GR significantly) in order to amplify any possible impact. We will take a range of values for these parameters in the next section.

We find that when we apply Eq. (50) in both Hu-Sawicki $f(R)$ and nDGP, we retain precisely the same data points as we do in the GR case in the previous subsection. This holds across the Υ_{gm}, Υ_{gg} and E_G data vectors, as well as for both Year 1 and Year 10 LSST sources. Although these are only two examples of alternative gravity theories, we can cautiously state that the conclusions of Sec. IV A—that E_G offers a possible probe of gravity that is less sensitive to nonlinear effects than its constituent probes—appear to be robust to differences in the underlying true theory of gravity. We do, however, note that this test has not checked the impact of a theory where there is a scale-dependent difference between the lensing and Newtonian potentials.

C. Nonlinear galaxy bias

The definition of E_G is near-symmetric in the numerator and denominator with respect to the two-point correlation function of the matter field (modulo the lensing kernel), but it is not symmetric in galaxy bias. If scale-dependent galaxy bias is present, this effect will never cancel, as Υ_{gg} has two factors of this galaxy bias, Υ_{gm} has one, and β will only ever include linear galaxy bias behavior as it is definitionally measured at large scales. Measurements of E_G typically must correct for any such nonlinear galaxy bias; for a general discussion of the potential effect of leaving this effect uncorrected, see Ref. [41].

To investigate the potential impact of this on linear scale cuts in an E_G analysis, we repeat the procedure of Sec. IV A, but this time we allow the nonlinear data vector $\vec{D}_{\rm nonlin}$ to also include nonlinear galaxy bias, while the linear data vector continues to exhibit linear galaxy bias. Nonlinear matter clustering behavior is modeled in GR.

We are interested in the case of a nonlinear bias model which accurately reflects DESI LRGs. For this purpose, we take the LPT model with parameters as constrained in Ref. [57] for a DESI-LRG-like sample from the DECam Legacy Survey (DECaLS). This work finds best-fit nonlinear bias parameters $b_1 = 1.333$ and $b_2 = 0.514$ (with b_s and higher order terms fixed to 0; see e.g. Ref. [68] for definitions of these parameters). In practice, we convert these parameters to their Eulerian equivalents values using the relations found in Ref. [68] [Eq. (4.8)] and use the `FAST-PT` code [69] as linked through the Core Cosmology Library (`CCL`, [70]) to compute clustering statistics with nonlinear biases. The equivalent Eulerian parameters are given as $b_1 = 2.33$, $b_2 = 1.02$, and $b_s = -0.381$ (noticing b_s is nonzero in the Eulerian case).

The effect of this nonlinear galaxy bias on which data vector elements are retained or discarded under the criterion of Eq. (50) is displayed in Fig. 5, for LSST Year 1 sources. Note that we display the data vector under nonlinear modeling and hence we see a scale-dependence of E_G as a result of nonlinear galaxy bias. For the case of LSST Year 10 sources, the data points retained and cut are very similar, with the only difference being that one additional point is cut for E_G at $r_p \approx 30$ Mpc/h.

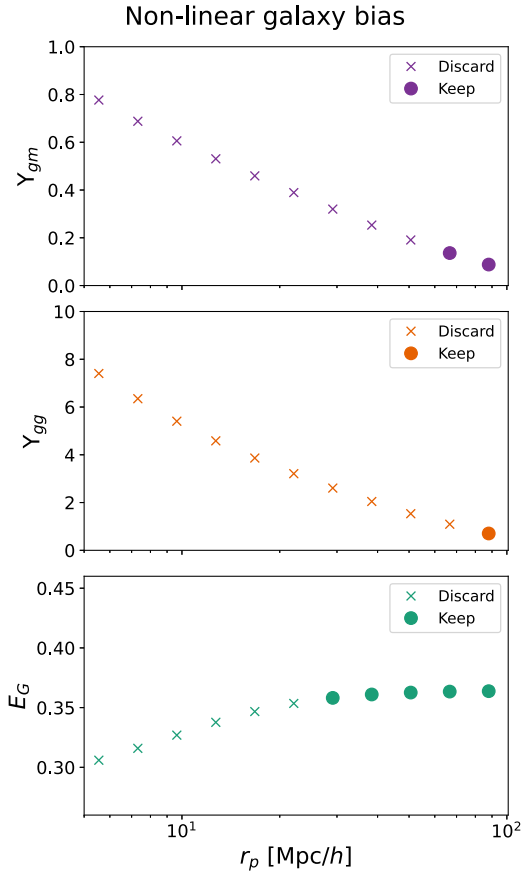


FIG. 5. Illustration of which points are kept in the linear-only scale cuts scenario for Υ_{gm} (top), Υ_{gg} (middle) and E_G (bottom), assuming a nonlinear galaxy bias as described in Sec. IV C. The data vector plotted assumes this nonlinear galaxy bias model. We display the case with LSST Year 1 source galaxies; for LSST Year 10 sources, we would discard one additional point in the E_G case, at $r_p \approx 30$ Mpc/h.

We see in Fig. 5 that the cuts to the individual Υ_{gm} and Υ_{gg} remain the same as in the linear galaxy bias case of Sec. IV A. More pertinently, though, a large number of small-scale data points are now cut from E_G. As expected, this is due to E_G being depressed at small scales due to the double factor of nonlinear bias in the denominator (vs the single factor in the numerator). Assuming our galaxy-bias model is reasonably close to what might be present in real DESI LRG data, it is clear that nonlinear galaxy bias must be accounted for.

D. Calibration with respect to nonlinear bias

As shown in the previous subsection, nonlinear galaxy bias results, for the case of an LSST + DESI sample, in stringent linear-only scale cuts on $E_G(r_p)$. We now explore the possibility of correcting our measured $E_G(r_p)$ for this nonlinear galaxy bias, recovering a version of $E_G(r_p)$ which is insensitive to nonlinear effects at smaller scales and which could thus in principle more reliably pick up scale-dependent signatures of alternative theories of gravity.

We follow Ref. [20], defining a similar correction factor to the one introduced in that work (for use in that case with mock galaxy catalogs), as follows:

$$C_b(r_p) \equiv \frac{\Omega_{\rm M}^0 \Upsilon_{gg}(r_p)}{b_1 \Upsilon_{gm}(r_p)}. \tag{51}$$

b_1 is the fiducial value of the linear bias parameter. $\Upsilon_{gm}(r_p)$ and $\Upsilon_{gg}(r_p)$ are computed from theory using Eqs. (7) and (19), respectively, in both cases crucially with nonlinear galaxy bias modeling (here, from Ref. [57] as in the previous subsection).

A corrected $E_G(r_p)$ is then defined as

$$\hat{E}_G(r_p) \equiv C_b(r_p) E_G(r_p). \tag{52}$$

To gain some intuition about this correction factor, consider that the ratio $\Upsilon_{gg}(r_p)/\Upsilon_{gm}(r_p)$ in Eq. (51) is equal to $(\beta E_G(r_p))^{-1}$ (computed theoretically using the nonlinear galaxy bias model). Recall that $\beta = f/b_1$ (where b_1 is the linear galaxy bias parameter in the notation of this section), and that as per Eq. (4), the theoretical prediction for E_G in GR is given by $f/\Omega_{\rm M}^0$. Putting these together, we see that C_b is designed to capture the scale-dependent behavior of $E_G(r_p)$ which is due to nonlinear galaxy bias, while returning to unity at large scales.

In the case of a real measurement, C_b should be calculated at the best-fit parameter values of the cosmology and nonlinear galaxy bias model. Note that a factor of $\Omega_{\rm M}^0$ will appear linearly in the theoretical prediction for $\Upsilon_{gm}(r_p)$, and as such applying the correction factor C_b does not introduce additional dependence on $\Omega_{\rm M}^0$; it in indeed for this exact purpose that we include $\Omega_{\rm M}^0$ in the definition of Eq. (51).

We now go further than Ref. [20] in seeking to incorporate uncertainty associated with estimation of C_b in our analysis. Particularly, we are concerned with the uncertainty which arises from the estimation of the nonlinear galaxy bias parameters. For the current case, we refer again to the nonlinear galaxy bias model fit to a DESI-LRG-like sample, from Ref. [57]. In addition to the best-fit values as given in Sec. IV C, that work found joint posteriors of the Lagrangian bias parameters b_1 and b_2. Taking the marginalized two-dimensional posterior distribution over these two parameters to be approximately Gaussian (in agreement with Ref. [57]), we determine a covariance matrix that reproduces their two-dimensional marginalized constraints. From the associated two-dimensional Gaussian joint posterior, we draw realizations of these two Lagrangian bias parameters and convert these to the associated Eulerian parameters at each realization.

At each set of bias parameter values, we then compute $C_b(r_p)$ using Eq. (51). From these realizations, we use the standard sample covariance relationships to compute $\mathrm{Cov}(C_b(r_p), C_b(r'_p))$. Note that we neglect the effect of the Hartlap correction arising from this finite number of samples when inverting the covariance as we expect it to be subdominant to that associated with the finite number of simulation realizations used in computing the covariance of the raw $E_G(r_p)$ (using the methodology of Sec. III).

Taking the assumption that $C_b(r_p)$ and $E_G(r_p)$ are uncorrelated, it holds that

$$\begin{aligned}&\mathrm{Cov}\big(\hat{E}_G(r_p), \hat{E}_G(r'_p)\big)\\&\quad= \langle C_b(r_p)\rangle\langle C_b(r'_p)\rangle \mathrm{Cov}\big(E_G(r_p), E_G(r'_p)\big)\\&\qquad+ \langle E_G(r_p)\rangle\langle E_G(r'_p)\rangle \mathrm{Cov}\big(C_b(r_p), C_b(r'_p)\big). \quad (53)\end{aligned}$$

We see from the form of this equation that if the covariance of the bias correction factor is small, this reduces to the previous version of the $E_G(r_p)$ covariance modulo factors of $C_b(r_p)$. We visualize the corrected correlation matrices for the cases of both LSST Year 1 and Year 10 sources in Fig. 6. Comparing with Fig. 3, we can see that the primary effect of incorporating this correction factor on the covariance is to increase the correlation in off-diagonal elements at small r_p. This can be understood by recalling that nonlinear bias has an impact primarily on smaller scales.

Taking the above covariance for $\hat{E}_G(r_p)$, we compute "linear" scale cuts on the corrected $\hat{E}_G(r_p)$ using the condition of Eq. (50), where the nonlinear data vector is now the corrected $\hat{E}_G(r_p)$. With this correction method, we find that we are once again not required to cut any of our $\hat{E}_G(r_p)$ data points. This holds for both LSST Year 1 and LSST Year 10 sources. We take this as our fiducial method of determining scales to be included in our analysis going forward.

It is worth noting briefly at this point that, if one's objective is to fit a model where E_G is a single r_p-independent parameter, and to compare this value with its scale-independent GR prediction, then the discarding of smaller-scale data points may not have much effect on the ultimate error bar on this constant E_G parameter. In this scenario, we are fitting a one-parameter model, and the inclusion of more data points at smaller scales does not add much information—if (crucially) those data points are in agreement with the constant model. The benefit of including smaller scales is more obvious in the case where the data show disagreement with an r_p independent model. If this scale-dependent signature could be reliably decoupled from other scale-dependent effects (and especially nonlinear galaxy bias effects), this would be a clear indication of a deviation from GR.

V. TESTING THE POWER OF AN LSST + DESI E_G MEASUREMENT

We are now in a position to explore in detail the benefits and potential pitfalls of using E_G to test gravity with LSST + DESI. We are interested in the question: given the true Universe is described by a theory of gravity other than GR, would we detect this?

Most literature which tests gravity using E_G takes a heuristic approach to this question, qualitatively comparing measurements with the predicted value in GR at the appropriate redshift (see, e.g., Refs. [5,20–22,24]). This

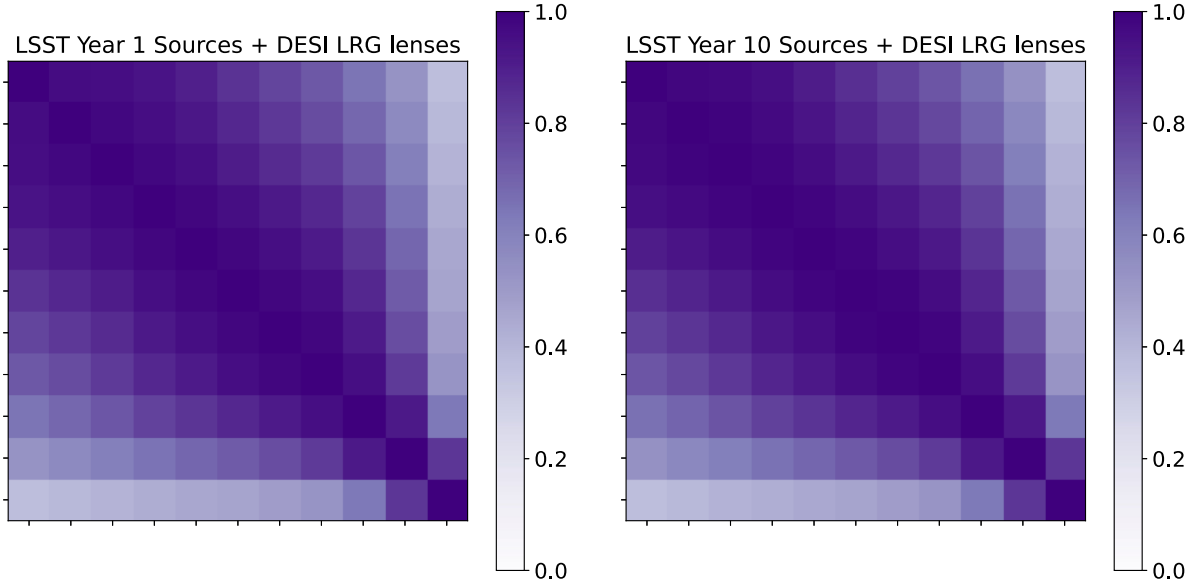


FIG. 6. Correlation matrices of $\hat{E}_G(r_p)$, incorporating a nonlinear bias correction factor as defined in Eqs. (51) and (53). Bins increase logarithmically in r_p with bin centers ranging from 1.4 to 87.9 Mpc/h.

approach offers valuable intuition and has led to the strong disfavoring of some theories of gravity with extreme behavior captured by E_G (e.g. most notably tensor-vector-scalar gravity (TeVeS) in Ref. [20]). However, the heuristic approach also has its pitfalls—notably, as discussed in Ref. [23], the fact that the waters are invariably muddied by the fact that the GR prediction itself depends on the value of $\Omega_{\rm M}^0$. A measurement of E_G which seems to disagree with the GR prediction could in fact therefore be instead indicative of the datasets in question preferring a different $\Omega_{\rm M}^0$ value than that assumed in constructing the GR prediction.

To address this, we propose a Bayesian approach to determine whether and in what circumstances an LSST + DESI measurement of E_G will detect true signatures of deviations from GR. Specifically, our approach makes use of the framework of *posterior predictive tests*, a Bayesian analog to hypothesis testing [71] which has found considerable interest in cosmology analysis in recent years (see e.g. Refs. [72–74]).

A. Posterior predictive test method

First, note that as in Sec. IV, we restrict to bins in $r_p \geq 5$ Mpc/h. Our method is then as follows.

Generate a simulated data realization: consider a hypothetical Universe that is described by an alternative theory of gravity with specific values of the gravity-theory parameters (e.g. f_{R0} in $f(R)$ gravity). We have:

(i) a model (M) for computing $E_G(r_p)$ under the alternative theory of gravity and at the specific gravity-theory parameters in question, as a function of cosmological parameters. In practice, the model is effectively insensitive to the value of any cosmological parameters other than $\Omega_{\rm M}^0$.
(ii) a covariance matrix for $E_G(r_p)$, from Sec. III. (This assumes that the covariance matrix as computed in GR is appropriate, which is a suitable approximation [58]).
(iii) a prior on $\Omega_{\rm M}^0$, $p(\Omega_{\rm M}^0)$.

Given these ingredients, we can generate a simulated data realization (D) of $E_G(r_p)$ in this Universe. We do so by sampling from the following distribution:

$$p(D|M,\Omega_{\rm M}^0)p(\Omega_{\rm M}^0). \tag{54}$$

Practically, we first draw from $p(\Omega_{\rm M}^0)$, and plug the resulting value of $\Omega_{\rm M}^0$ into the model M. This generates the mean of a multivariate Gaussian (with dimensionality equal to the number of r_p bins) from which we draw the realization of D.

Fit a constant model: we initially fit a constant model to the data realization drawn in the previous step, and check its goodness of fit via a χ^2 test. If the fit is bad, we do not proceed, and consider GR to be rejected for this data realization.

Find the posterior of the parameters of the GR model: assuming that the fit to a constant model is acceptable, we proceed to fit the parameters of the GR model for E_G to the simulated data realization. Recall that the GR model for E_G is given by Eq. (4), which means that the only parameter is $\Omega_{\rm M}^0$. Notationally we would like to separate this parameter $\Omega_{\rm M}^0$ from that which is used to generate the data; we call the current parameter $\Omega_{\rm M}^{0,\rm fit}$. In determining the posterior distribution $p(\Omega_{\rm M}^{0,\rm fit}|D)$, we incorporate a prior on $\Omega_{\rm M}^{0,\rm fit}$, which is identical to our data-generating prior on $\Omega_{\rm M}^0$ introduced above.

Generate replicated data: we draw samples from the posterior distribution on $\Omega_{\rm M}^{0,\rm fit}$. We use these sampled values of $\Omega_{\rm M}^{0,\rm fit}$ to compute corresponding instances of $E_G(r_p)$ under the GR model, $M_{\rm GR}$, which is given by Eq. (4) (and is constant across r_p bins). This set of E_G values is conventionally known as the "replicated data," $D_{\rm rep}$, within the posterior predictive framework.

Compare simulated data realization with distribution of replicated data: we now have samples $D_{\rm rep}$, as well as the original data realization D (which, recall, is drawn from a universe described by an alternate theory of gravity with specific parameters). We now want to determine where the original data realization lies with respect to the distribution of replicated data. In doing so, we apply a threshold at which we consider the null GR hypothesis to be rejected, in analogy to the p value thresholds applied in frequentist hypothesis testing. This serves only to summarize whether a given simulated analysis has accepted or rejected GR, and thus in what fraction of simulated analyses each outcome occurs. In frequentist hypothesis testing, a threshold of $p = 0.05$ is typical, and we follow that standard here by considering the GR hypothesis to be rejected if the true data fall outside the symmetric 95% probability density region of the distribution of replicated data. One could in theory apply a higher threshold to reject the null GR hypothesis, by replacing 95% with a larger value.

There is a slight complication due to the in-principle multidimensional nature of the data (having dimension equal to the number of bins in r_p). The posterior predictive test requires us to map the true (simulated) data realization as well as the samples of replicated data onto a scalar test statistic. Given that the replicated data will always be scalar-valued (due to the scale independence of E_G in GR), and we have already checked that a constant model is a good fit to the data realization, we take the constant value fit to the simulated data realization as this test statistic. An alternative option would have been to treat the measurement of $E_G(r_p)$ in each r_p bin individually from the beginning, repeating the analysis for each bin (skipping the "Fit to a constant model" step above). This would result in acceptance or rejection of GR given each data point, and would probe the question of scale dependence indirectly. However, the power of the posterior predictive test for each

TABLE I. Fraction of cases in which we reject the GR null hypothesis for LSST Year 1 + DESI.

$\Omega_{\rm M}^0$ prior	$f_{R0} = 10^{-6}$	$f_{R0} = 10^{-5}$	$f_{R0} = 10^{-4}$	$\Omega_{\rm rc} = 0.05$	$\Omega_{\rm rc} = 0.25$	$\Omega_{\rm rc} = 0.5$
Stage III 3×2 pt	0.017	0.018	0.023	0.017	0.017	0.017
Planck	0.346	0.572	0.924	0.345	0.457	0.559

TABLE II. Fraction of cases in which we reject the GR null hypothesis for LSST Year 10 + DESI.

$\Omega_{\rm M}^0$ prior	$f_{R0} = 10^{-6}$	$f_{R0} = 10^{-5}$	$f_{R0} = 10^{-4}$	$\Omega_{\rm rc} = 0.05$	$\Omega_{\rm rc} = 0.25$	$\Omega_{\rm rc} = 0.5$
Stage III 3×2 pt	0.017	0.017	0.021	0.017	0.017	0.017
Planck	0.319	0.553	0.925	0.327	0.438	0.536

individual r_p bin would likely be reduced relative to the combined approach that we use here.

The above process provides the Bayesian analog to a two-sided p value, telling us whether to accept or reject GR given this specific data realization. In an analysis using real data, we would perform the steps discussed above just once. We would thus determine where the true measured value of E_G falls within the distribution of replicated data, and discuss the acceptance or rejection of the null GR hypothesis on this basis (assuming our measured $E_G(r_p)$ is consistent with scale independence). However, in this work, we are interested in assessing the power of an LSST + DESI measurement of E_G. We therefore repeat this process many times for different simulated data realizations (1000 times for each scenario considered). We ultimately report the fraction of such trials for which we reject GR, for a given scenario. This demonstrates the reliability that we will detect a true underlying deviation from GR with this method in the various scenarios considered.

B. Results

We will consider the following cases for the underlying "true" alternative gravity models:

(i) nDGP gravity with $\Omega_{\rm rc} = \{0.05, 0.25, 0.5\}$.
(ii) Hu-Sawicki $f(R)$ gravity with $f_{R0} = \{10^{-6}, 10^{-5}, 10^{-4}\}$ and $n = 1$.

We consider two options for the prior on $\Omega_{\rm M}^0$. In both cases, we use a Gaussian prior with mean shifted to agree with our fiducial parameters; the priors differ only by their variances.

(i) Stage III 3×2 point prior: $\sigma(\Omega_{\rm M}^0) = 0.03$, corresponding to the constraint on $\Omega_{\rm M}^0$ under the ΛCDM model from DES Year 3 [75].
(ii) CMB prior: $\sigma(\Omega_{\rm M}^0) = 0.0084$, corresponding to the constraint on $\Omega_{\rm M}^0$ under the ΛCDM model (temperature and E-mode polarization auto- and cross-spectra case) from *Planck* [76].

We find that in the case of the 3×2 point prior on $\Omega_{\rm M}^0$, for both LSST Year 1 and Year 10 sources, we have a very low fractional rate of rejecting the GR null hypothesis in the presence of true gravity described by any of our alternate models. However, when we move to the more restrictive CMB prior on $\Omega_{\rm M}^0$, the rate at which we reject the null hypothesis is considerably higher, particularly in the cases of true alternative gravity descriptions which deviate more from GR. Fractional rejection rates are given in Tables I and II for LSST Year 1 and Year 10, respectively. We see that, as discussed qualitatively in Ref. [23], the prior information available on $\Omega_{\rm M}^0$ is crucial in determining the power of E_G as diagnostic for modified gravity. The introduction of the tighter prior on $\Omega_{\rm M}^0$ from CMB data leads to significantly higher rejection rates of GR (recalling that in these simulated analyses, GR is never the true underlying theory of gravity). The case of the CMB prior on $\Omega_{\rm M}^0$ combined with a true gravity theory of $f(R)$ with $f_{R0} = 10^{-4}$ is particularly dramatic, with a Stage III 3×2 point prior offering only approximately 2% rejection of GR, whereas the CMB prior offer over 92% rejection rate in both LSST Year 1 and Year 10. Furthermore, the effect of the tighter vs wider $\Omega_{\rm M}^0$ prior far outweighs the effect of moving from LSST Year 1 to Year 10 sources.

To better understand what is happening, we consider in Figs. 7 and 8 two example data realizations under two different cases, and illustrate the steps of the posterior predictive method in each case. In both scenarios, the underlying true theory of gravity is $f(R)$ gravity with $f_{\rm R0} = 10^{-4}$, and the sources are from LSST Year 10. In Fig. 7, the prior on $\Omega_{\rm M}^0$ is from *Planck*, and the null hypothesis of GR is ultimately rejected, whereas in Fig. 8, the prior is from DES Year 3, and the null hypothesis is accepted. Recall first that both priors are Gaussians centered on the fiducial value of $\Omega_{\rm M}^0$ (here, 0.292) differing only in their variances. In examining particularly the middle panel of both figures, we see that in both cases, the non-GR data are pulling the posterior towards lower values of $\Omega_{\rm M}^{0,\rm fit}$. However, in Fig. 8, the less informative prior has less impact in pulling the posterior up to a higher value of $\Omega_{\rm M}^{0,\rm fit}$ than does the more informative prior in Fig. 7. The result is that in the case shown in Fig. 8 we accept GR, with the implicit corollary that we are allowed to be in a Universe with $\Omega_{\rm M}^0$ described by the middle panel

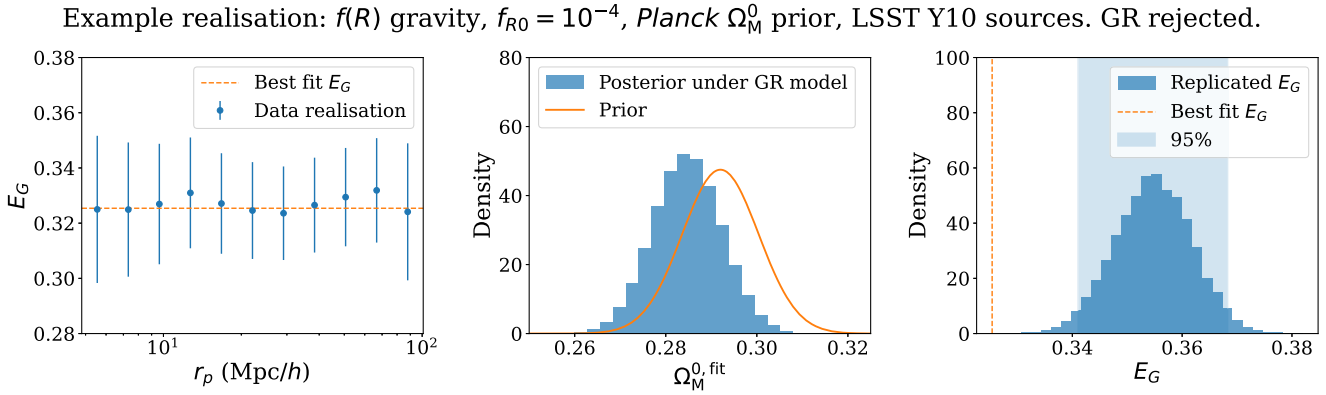


FIG. 7. An example where we reject the GR null hypothesis when $f(R)$ gravity is the underlying true gravity theory. We show an example data realization and the resulting steps of the posterior predictive method for the case of $f_{R0} = 10^{-4}$, LSST Year 10 sources, and a *Planck* CMB prior on Ω_M^0. Left: data realization and best-fit E_G constant value. Middle: posterior distribution on $\Omega_M^{0,\mathrm{fit}}$. Right: resulting replicated E_G values (D_{rep}), compared with the best-fit value of E_G with respect to the data realization and covariance. We see that in this example, we reject the GR null hypothesis.

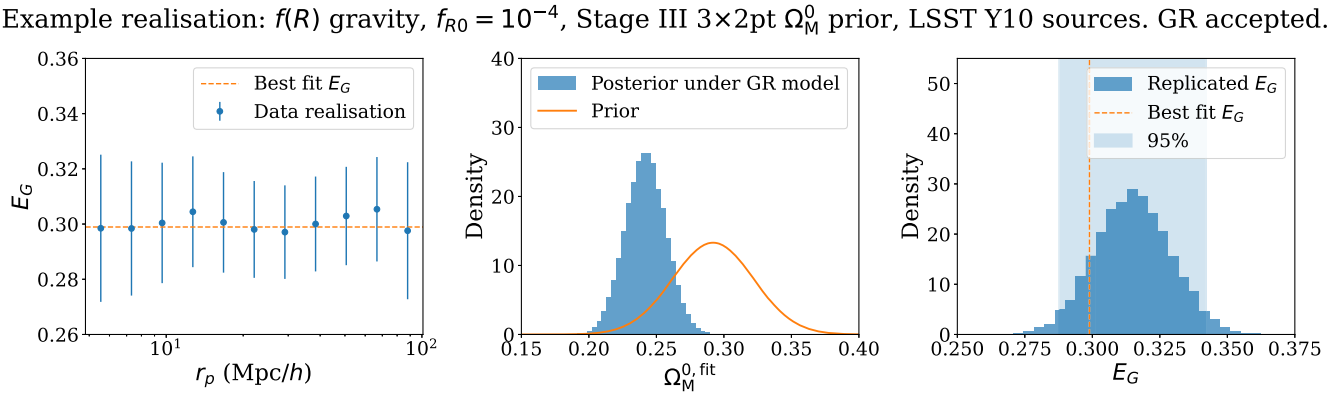


FIG. 8. An example where we accept the GR null hypothesis when $f(R)$ gravity is the underlying true gravity theory. We show an example data realization and the resulting steps of the posterior predictive method for the case of $f_{R0} = 10^{-4}$, LSST Year 10 sources, and a Stage III 3×2 point (DES Year 3) prior on Ω_M^0. Left: data realization and best-fit E_G constant value. Middle: posterior distribution on $\Omega_M^{0,\mathrm{fit}}$. Right: resulting replicated E_G values (D_{rep}), compared with the best-fit value of E_G with respect to the data realization and covariance. We see that in this example, we accept the GR null hypothesis.

of Fig. 8. On the other hand, in the case of a tighter *Planck* CMB prior as in Fig. 7, the prior drags the posterior on $\Omega_M^{0,\mathrm{fit}}$ closer to the fiducial value, with correspondingly higher E_G values in the replicated data distribution (right panel), which are inconsistent with the best-fit value to the E_G data realization in the non-GR Universe. As a result, it is this more strict prior information on Ω_M^0 that leads us to (correctly) reject the null GR hypothesis in this case. For completeness, we present equivalent cases under nDGP gravity in Figs. 9 and 10.

Zooming back out from these specific examples to the broad picture: for scenarios with the Stage III 3×2 point prior on Ω_M^0, the strong majority of cases of rejecting GR are due to the constant model being flagged as a bad fit by the reduced χ^2 test, rather than due to the posterior predictive test. Because the actual theoretical predictions from both $f(R)$ and nDGP are scale independent, we attribute these failures to fit a constant model instead to fluctuations in the data-generating draws. This is in contrast to the case of the CMB prior, where rejecting GR is nearly always due to the posterior predictive test.

We note also that even in the case of the CMB prior, differences between the rate of null hypothesis rejection in the case of LSST Year 1 vs Year 10 sources are minimal. In fact, we see that the LSST Year 10 case actually rejects GR slightly less frequently than for LSST Year 1 (although these differences are minor). We therefore conclude that the dominant effect for this data setup is that of the prior on Ω_M^0, rather than the shape noise of the source sample.

Finally, it is worth briefly noting that the priors we have considered on Ω_M^0 are taken from analyses performed under the assumption of GR. Analyses under the assumption that

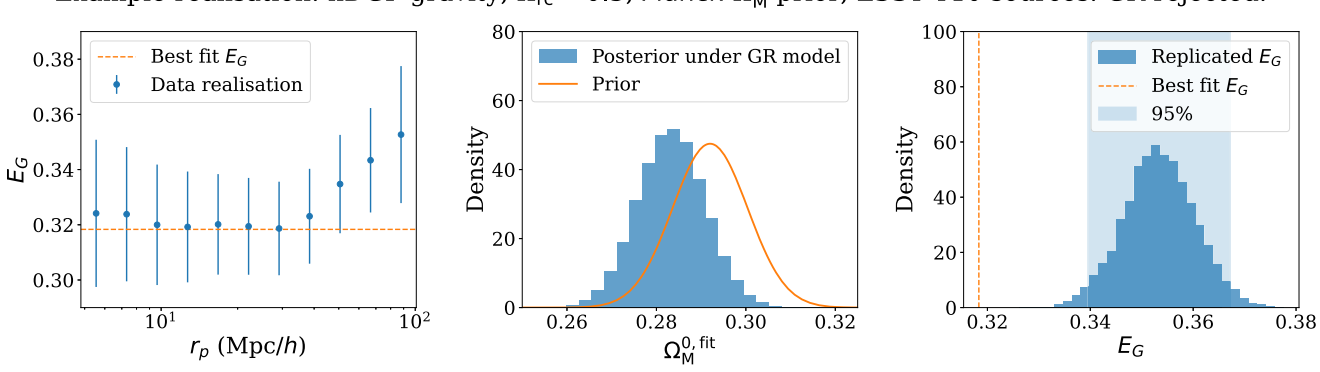


FIG. 9. An example where we reject the GR null hypothesis when nDGP gravity is the underlying true gravity theory. We show an example data realization and the resulting steps of the posterior predictive method for the case of $\Omega_{rc} = 0.5$, LSST Year 10 sources, and a *Planck* CMB prior on Ω_M^0. Left: data realization and best-fit E_G constant value. Middle: posterior distribution on $\Omega_M^{0,fit}$. Right: resulting replicated E_G values (D_{rep}), compared with the best-fit value of E_G with respect to the data realization and covariance. We see that in this example, we reject the GR null hypothesis.

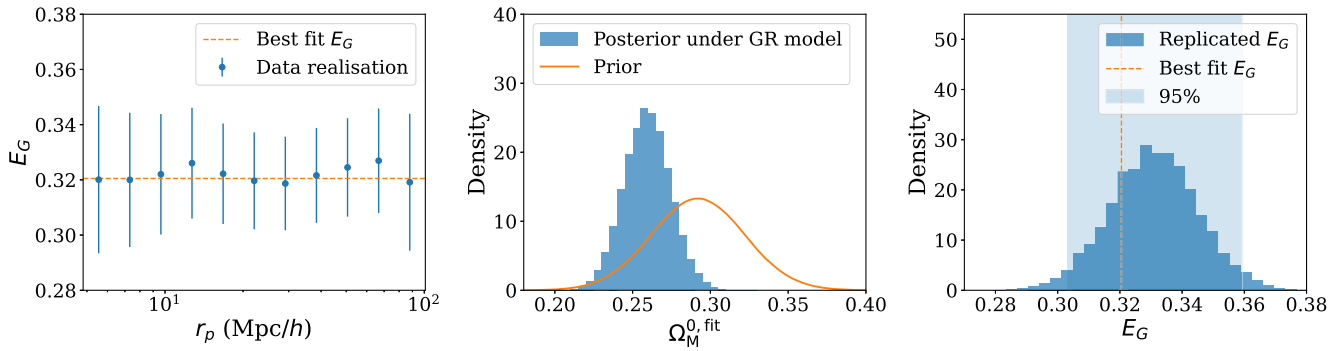


FIG. 10. An example where we accept the GR null hypothesis when nDGP gravity is the underlying true gravity theory. We show an example data realization and the resulting steps of the posterior predictive method for the case of $\Omega_{rc} = 0.5$, LSST Year 10 sources, and a Stage III 3×2 point (DES Year 3) prior on Ω_M^0. Left: data realization and best-fit E_G constant value. Middle: posterior distribution on $\Omega_M^{0,fit}$. Right: resulting replicated E_G values (D_{rep}), compared with the best-fit value of E_G with respect to the data realization and covariance. We see that in this example, we accept the GR null hypothesis.

a different theory of gravity holds would most likely have resulted in a broader constraints on Ω_M^0, due to their increased parameter space (see e.g. Refs. [12,77,78]). For real-data analyses, a conservative approach that takes the Ω_M^0 prior from analyses that allow for additional degrees of freedom in the gravitational sector should be considered.

VI. DISCUSSION AND CONCLUSIONS

In this work, we have explored the potential of testing gravity with E_G using LSST weak lensing source galaxies and DESI LRG lenses. We construct the required covariance matrices and examine the sensitivity of E_G to nonlinear effects in structure formation and galaxy bias. We have introduced the idea of the posterior predictive test as a valuable tool for consistently incorporating theoretical uncertainties in E_G (from its model dependence on Ω_M^0).

Our work highlights in a quantitative manner the key role played by the prior information available on Ω_M^0 in the power of an LSST + DESI E_G measurement to reject the null GR hypothesis in the presence of non-GR gravitational behavior. We have seen that the effect of imposing a *Planck* prior on Ω_M^0 as opposed to a Stage III 3×2 point prior has a much more significant impact on our capacity to reject the null GR hypothesis than moving from an LSST Year 1 to an LSST Year 10 source sample, for the alternative theories of gravity considered here. This result highlights that which has been suggested in the literature previously [23]: that we are now well outside the regime where the uncertainty on Ω_M^0 can be disregarded when interpreting E_G measurements. We therefore strongly advocate the use of an analysis methodology which consistently incorporates this factor, such as the posterior predictive test as we have used here.

In the case of the stronger CMB priors on $\Omega_{\rm M}^0$, we reject GR in the strong majority of cases only for Hu-Sawicki gravity with $f_{R0} = 10^{-4}$ (out of the scenarios considered). As this theory and parameter combination is generally considered in the literature to be an unrealistically large deviation from GR, this is perhaps not terribly promising at face value. However, we emphasize that the theories under consideration here were selected not because they are considered representative of the most likely ways nature may deviate from general relativity. Hu-Sawicki $f(R)$ and nDGP are workhorse examples of alternative gravity theories. We selected these in part because they cover two major screening mechanisms prevalent amongst alternative theories of gravity, but also because of the availability of tools for computing cosmological observables in these theories. The apparent pessimism of our results should therefore be taken with a grain of salt, particularly in combination with our major finding, which is the extreme role of prior information on $\Omega_{\rm M}^0$ in determining our ability to reject the null GR hypothesis.

One further caveat to the selection of gravity theories considered here is that in neither of these theories do we expect a scale-dependent $E_G(r_p)$. To be clear, this scenario does capture many alternative gravity theories of interest. In all theories with a single scalar degree of freedom, for example, the linear modification to the growth and lensing potential exhibit scale dependence only subdominantly, becoming relevant only at near-horizon scales which are not relevant to us here [79]. In the original paper proposing E_G [19], a single theory (TeVeS [80]) was cited as being expected to exhibit a scale-dependent $E_G(r_p)$; this was strongly disfavored by the first measurement of E_G [20] not long after. However, for completeness we highlight that our results may not hold in the case where the true Universe is described by a theory which would induce strong scale-dependence in E_G; presumably, in such a case, an LSST + DESI measurement E_G would be more sensitive to deviations from GR than we have seen here.

In our exploration of the sensitivity of E_G to nonlinear behavior (Sec. IV), we found that if we assume a decoupling between the scale dependence of nonlinear structure formation at a dark matter only level, and that of galaxy bias (not a trivial assumption), E_G may offer a method for testing gravity uniquely insensitive to uncertainties in nonlinear structure formation. Although this means that E_G is unlikely to be sensitive to screening effects, it does open the door for the possibility of different ways of utilizing this combination of measurements—perhaps in combination with linear-only parametrizations of deviations from GR as a data vector used to constrain parameters.

While we have presented in this work a thorough study of the effect of nonlinear structure growth and galaxy bias on E_G for LSST + DESI, this is not an exhaustive consideration of potentially relevant systematic effects. Particularly, future work should consider to what extent unmodeled galaxy intrinsic alignment effects in Υ_{gm} could induce spurious detections of deviations from GR with E_G. This was considered in Ref. [26] for DESI Data Release 1 LRG lenses and Stage III weak lensing sources, where it was found that accounting for intrinsic alignment has a very minor impact on results, however, this may not carry over to the case of LSST sources. Similarly, although we expect E_G to be largely insensitive to the details of the source galaxy redshift distribution, verifying this explicitly for realistic LSST photometric redshift uncertainties would be valuable. Should either intrinsic alignment or photometric redshift uncertainties prove to be important systematic effects for E_G, it is likely that careful selection of lens and source tomographic bins (minimizing the number of lens-source pairs which are close along the line of sight) could mitigate their impacts.

LSST promises to deliver an unprecedented wealth of high-quality data, making it a uniquely powerful tool for testing gravity on cosmological scales. Consistency-check analyses, of which E_G is a major example, uniquely offer a way of testing gravity that does not require the assumption of a parametrization or particular theory of interest within our analysis. Our results provide a clear indication that for E_G to play a powerful role as a consistency check with LSST and DESI data requires precision prior knowledge on $\Omega_{\rm M}^0$.

ACKNOWLEDGMENTS

This paper has undergone internal review in the LSST Dark Energy Science Collaboration. The internal reviewers were Jessie Muir and Luisa Jaime. C. D. L. is supported by the Science and Technology Facilities Council (STFC) [Grant No. UKRI1172]. R. M. is supported by the Department of Energy Cosmic Frontier program, Grant No. DE-SC0010118. The DESC acknowledges ongoing support from the Institut National de Physique Nucléaire et de Physique des Particules in France; the Science & Technology Facilities Council in the United Kingdom; and the Department of Energy and the LSST Discovery Alliance in the United States. DESC uses resources of the IN2P3 Computing Center (CC-IN2P3–Lyon/Villeurbanne-France) funded by the Centre National de la Recherche Scientifique; the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231; STFC DiRAC HPC Facilities, funded by UK BEIS National E-infrastructure capital grants; and the UK particle physics grid, supported by the GridPP Collaboration. This work was performed in part under DOE Contract No. DE-AC02-76SF00515. We acknowledge the use of NumPy [81], SciPy [82] and Matplotlib [83]. The analysis code for this paper is publicly available at

Ref. [84]. We provide this code publicly in the spirit of open science and make no guarantees of code maintenance.

C. D. L. wrote code, performed the main analysis, wrote most of the paper text, and drove the project. S. A. wrote code and performed analysis to enable the calculation of simulation-based covariance matrices. R. M. participated in formulating the project idea, provided scientific guidance, and participated in paper revision. M. M. R. participated in formulating the posterior predictive methodology. S. S. wrote code to compute the analytic covariance matrices and provided scientific input. C. Z. wrote code to compute the power spectrum under alternative theories of gravity, wrote the section of the paper describing alternative gravity theories and participated in paper revision.

DATA AVAILABILITY

The data that support the findings of this article are openly available [84].

APPENDIX: ADDITIONAL COVARIANCE-RELATED EXPRESSIONS AND FIGURES

In this appendix, we provide for completeness some additional expressions and figures for helpful quantities in the calculation of covariance discussed in Sec. III.

The average in r_p and r'_p bins of $\mathrm{Cov}[\Delta\Sigma_{gx}(r_p), \Delta\Sigma_{gy}(r'_p)]$ (where x and y are given by either g or m) is given by

$$\mathrm{Cov}\left[\Delta\Sigma_{gx}(r_p^i), \Delta\Sigma_{gy}(r_p^j)\right] = \frac{4}{\left((r_p^{i,h})^2 - (r_p^{i,l})^2\right)\left((r_p^{j,h})^2 - (r_p^{j,l})^2\right)} \int_{r_p^{i,l}}^{r_p^{i,h}} dr_p r_p \int_{r_p^{j,l}}^{r_p^{j,h}} dr'_p r'_p \mathrm{Cov}\left[\Delta\Sigma_{gx}(r_p), \Delta\Sigma_{gy}(r'_p)\right], \quad \text{(A1)}$$

where $r_p^{i,h}$ and $r_p^{i,l}$ represent the high and low edges of bin r_p^i respectively (and similarly for bin r_p^j). The equivalent averaged quantity for $\mathrm{Cov}[\Upsilon_{gx}(r_p), \Upsilon_{gy}(r'_p)]$ is then

$$\begin{aligned}\mathrm{Cov}\left(\Upsilon_{gx}(r_p^i), \Upsilon_{gy}(r_p^j)\right) = {} & \mathrm{Cov}\left(\Delta\Sigma_{gx}(r_p^i), \Delta\Sigma_{gy}(r_p^j)\right) - \frac{2(r_p^0)^2}{(r_p^{i,h})^2 - (r_p^{i_l})^2} \ln\left(\frac{r_p^{i,h}}{r_p^{i,l}}\right) \mathrm{Cov}\left(\Delta\Sigma_{gx}(r_p^0), \Delta\Sigma_{gy}(r_p^j)\right) \\ & - \frac{2(r_p^0)^2}{(r_p^{j,h})^2 - (r_p^{j_l})^2} \ln\left(\frac{r_p^{j,h}}{r_p^{j,l}}\right) \mathrm{Cov}\left(\Delta\Sigma_{gx}(r_p^i), \Delta\Sigma_{gy}(r_p^0)\right) \\ & + \frac{2(r_p^0)^2}{(r_p^{i,h})^2 - (r_p^{i_l})^2} \ln\left(\frac{r_p^{i,h}}{r_p^{i,l}}\right) \frac{2(r_p^0)^2}{(r_p^{j,h})^2 - (r_p^{j_l})^2} \ln\left(\frac{r_p^{j,h}}{r_p^{j,l}}\right) \mathrm{Cov}\left(\Delta\Sigma_{gx}(r_p^0), \Delta\Sigma_{gy}(r_p^0)\right). \end{aligned} \quad \text{(A2)}$$

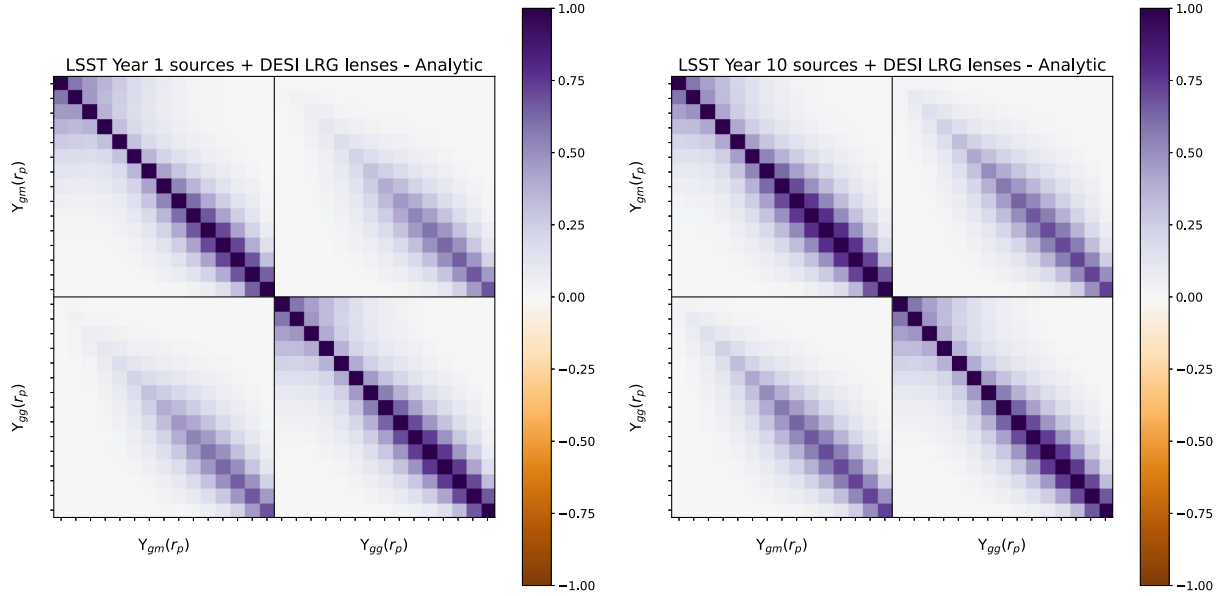


FIG. 11. Correlation matrix of Υ_{gm} and Υ_{gg} for LSST Year 1 (left) and Year 10 (right) sources with DESI LRG lenses, estimated using the analytic method of Sec. III A.

Specifically, in the case of considering the shape-noise-only term of $\mathrm{Cov}(\Upsilon_{gm}(r_p^i), \Upsilon_{gm}(r_p^j))$ for the purpose of combining with a simulation-estimated covariance that excludes this effect, we have

$$\begin{aligned}\mathrm{Cov}^{\mathrm{SN}}\left(\Upsilon_{gm}(r_p^i), \Upsilon_{gm}(r_p^j)\right) = \frac{(\overline{\Sigma_c^{-2}})^{-1}\sigma_\gamma^2}{2\pi V n_l n_s}\Bigg(&\frac{2}{(r_p^{i,h})^2-(r_p^{i,l})^2} - \frac{4(r_p^0)^2}{\left((r_p^{i,h})^2-(r_p^{i,l})^2\right)\left((r_p^{0,h})^2-(r_p^{0,l})^2\right)}\ln\left(\frac{r_p^{i,h}}{r_p^{i,l}}\right)\delta^K_{i,0}\\ &- \frac{4(r_p^0)^2}{\left((r_p^{j,h})^2-(r_p^{j,l})^2\right)\left((r_p^{0,h})^2-(r_p^{0,l})^2\right)}\ln\left(\frac{r_p^{j,h}}{r_p^{j,l}}\right)\delta^K_{j,0}\\ &+ \frac{8(r_p^0)^4}{\left((r_p^{i,h})^2-(r_p^{i,l})^2\right)\left((r_p^{j,h})^2-(r_p^{j,l})^2\right)\left((r_p^{0,h})^2-(r_p^{0,l})^2\right)}\ln\left(\frac{r_p^{i,h}}{r_p^{i,l}}\right)\ln\left(\frac{r_p^{j,h}}{r_p^{j,l}}\right)\Bigg),\end{aligned} \tag{A3}$$

where $\delta^K_{i,j}$ is a Kronecker δ on i, j. The analytic correlation matrices between $\Upsilon_{gm}(r_p)$ and $\Upsilon_{gg}(r_p)$ (auto and cross) as calculated using the expressions in Sec. III is displayed in Fig. 11 for LSST Year 1 and Year 10 sources.

[1] A. H. Wright, K. Kuijken, H. Hildebrandt *et al.*, The fifth data release of the Kilo Degree Survey: Multi-epoch optical/NIR imaging covering wide and legacy-calibration fields, Astron. Astrophys. **686**, A170 (2024).

[2] DES Collaboration, T. Abbott, M. Adamów *et al.*, The dark energy survey data release 2, Astrophys. J. Suppl. Ser. **255**, 20 (2021).

[3] H. Aihara, Y. AlSayyad, M. Ando *et al.*, Third data release of the Hyper Suprime-Cam Subaru strategic program, Publ. Astron. Soc. Jpn. **74**, 247 (2022).

[4] S. Joudaki, A. Mead, C. Blake *et al.*, KiDS-450: Testing extensions to the standard cosmological model, Mon. Not. R. Astron. Soc. **471**, 1259 (2017).

[5] C. Blake, A. Amon, M. Asgari *et al.*, Testing gravity using galaxy-galaxy lensing and clustering amplitudes in KiDS-1000, BOSS, and 2dFLenS, Astron. Astrophys. **642**, A158 (2020).

[6] T. Abbott, M. Aguena, A. Alarcon *et al.*, Dark energy survey year 3 results: Constraints on extensions to λCDM with weak lensing and galaxy clustering, Phys. Rev. D **107**, 083504 (2023).

[7] Ž. Ivezić *et al.*, LSST: From science drivers to reference design and anticipated data products, Astrophys. J. **873**, 111 (2019).

[8] R. Scaramella *et al.*, Euclid preparation-I. The Euclid Wide Survey, Astron. Astrophys. **662**, A112 (2022).

[9] R. Akeson *et al.*, The wide field infrared survey telescope: 100 Hubbles for the 2020s, arXiv:1902.05569.

[10] S. Joudaki, P. G. Ferreira, N. A. Lima, and H. A. Winther, Testing gravity on cosmic scales: A case study of Jordan-Brans-Dicke theory, Phys. Rev. D **105**, 043522 (2022).

[11] R. Liu, G. Valogiannis, N. Battaglia, and R. Bean, Constraints on f(R) and normal-branch Dvali-Gabadadze-Porrati modified gravity model parameters with cluster abundances and galaxy clustering, Phys. Rev. D **104**, 103519 (2021).

[12] L. Piga, M. Marinucci, G. d'Amico, M. Pietroni, F. Vernizzi, and B. S. Wright, Constraints on modified gravity from the BOSS galaxy survey, J. Cosmol. Astropart. Phys. 04 (2023) 038.

[13] A. Silvestri, L. Pogosian, and R. V. Buniy, Practical approach to cosmological perturbations in modified gravity, Phys. Rev. D **87**, 104015 (2013).

[14] E. Bellini and I. Sawicki, Maximal freedom at minimum cost: Linear large-scale structure in general modifications of gravity, J. Cosmol. Astropart. Phys. 07 (2014) 050.

[15] D. B. Thomas, Cosmological gravity on all scales: Simple equations, required conditions, and a framework for modified gravity, Phys. Rev. D **101**, 123517 (2020).

[16] S. Srinivasan, D. B. Thomas, F. Pace, and R. Battye, Cosmological gravity on all scales. Part II. Model independent modified gravity N-body simulations, J. Cosmol. Astropart. Phys. 06 (2021) 016.

[17] C. Zanoletti and C. D. Leonard, Principal components for model-agnostic modified gravity with 3×2 pt data, Phys. Rev. D **112**, 063547 (2025).

[18] M. Tsedrik, B. Bose, P. Carrilho, A. Pourtsidou, S. Pamuk, S. Casas, and J. Lesgourgues, Stage-IV cosmic shear with modified gravity and model-independent screening, J. Cosmol. Astropart. Phys. 10 (2024) 099.

[19] P. Zhang, M. Liguori, R. Bean, and S. Dodelson, Probing gravity at cosmological scales by measurements which test the relationship between gravitational lensing and matter overdensity, Phys. Rev. Lett. **99**, 141302 (2007).

[20] R. Reyes, R. Mandelbaum, U. Seljak, T. Baldauf, J. E. Gunn, L. Lombriser, and R. E. Smith, Confirmation of general relativity on large scales from weak lensing and galaxy velocities, Nature (London) **464**, 256 (2010).

[21] C. Blake, S. Joudaki, C. Heymans *et al.*, RCSLenS: Testing gravitational physics through the cross-correlation of weak lensing and large-scale structure, Mon. Not. R. Astron. Soc. **456**, 2806 (2016).

[22] S. Alam, H. Miyatake, S. More, S. Ho, and R. Mandelbaum, Testing gravity on large scales by combining weak lensing

with galaxy clustering using CFHTLenS and BOSS CMASS, Mon. Not. R. Astron. Soc. **465**, 4853 (2017).
[23] A. Amon, C. Blake, C. Heymans *et al.*, KiDS+ 2dFLenS+ GAMA: Testing the cosmological model with the EG statistic, Mon. Not. R. Astron. Soc. **479**, 3422 (2018).
[24] E. Jullo, S. De La Torre, M.-C. Cousinou, S. Escoffier, C. Giocoli, R. B. Metcalf, J. Comparat, H.-Y. Shan, M. Makler, J.-P. Kneib *et al.*, Testing gravity with galaxy-galaxy lensing and redshift-space distortions using CFHT-Stripe 82, CFHTLenS, and BOSS CMASS datasets, Astron. Astrophys. **627**, A137 (2019).
[25] S. Singh, S. Alam, R. Mandelbaum, U. Seljak, S. Rodriguez-Torres, and S. Ho, Probing gravity with a joint analysis of galaxy and CMB lensing and SDSS spectroscopy, Mon. Not. R. Astron. Soc. **482**, 785 (2019).
[26] S. Rauhut, C. Blake, U. Andrade, H. Noriega, J. Aguilar, S. Ahlen, S. BenZvi, D. Bianchi, D. Brooks, T. Claybaugh *et al.*, Testing gravitational physics by combining DESI DR1 and weak lensing datasets using the E_G estimator, Open J. Astrophys. **8**, 145197 (2025).
[27] A. R. Pullen, S. Alam, and S. Ho, Probing gravity at large scales through CMB lensing, Mon. Not. R. Astron. Soc. **449**, 4326 (2015).
[28] A. R. Pullen, S. Alam, S. He, and S. Ho, Constraining gravity at the largest scales through CMB lensing and galaxy velocities, Mon. Not. R. Astron. Soc. **460**, 4098 (2016).
[29] Y. Zhang, A. R. Pullen, S. Alam *et al.*, Testing general relativity on cosmological scales at redshift $z \approx 1.5$ with quasar and CMB lensing, Mon. Not. R. Astron. Soc. **501**, 1013 (2021).
[30] L. Wenzl, R. Bean, S.-F. Chen, G. S. Farren, M. S. Madhavacheril, G. A. Marques, F. J. Qu, N. Sehgal, B. D. Sherwin, and A. van Engelen, Constraining gravity with a new precision EG estimator using Planck + SDSS BOSS data, Phys. Rev. D **109**, 083540 (2024).
[31] L. Wenzl, R. An, N. Battaglia, R. Bean, E. Calabrese, S.-F. Chen, S. K. Choi, O. Darwish, J. Dunkley, G. S. Farren *et al.*, Atacama cosmology telescope: DR6 gravitational lensing and SDSS BOSS cross-correlation measurement and constraints on gravity with the EG statistic, Phys. Rev. D **111**, 043535 (2025).
[32] A. Pourtsidou, Testing gravity at large scales with H I intensity mapping, Mon. Not. R. Astron. Soc. **461**, 1457 (2016).
[33] M. M. Abidi, C. Bonvin, M. Jalilvand, and M. Kunz, Model-independent test for gravity using intensity mapping and galaxy clustering, Phys. Rev. D **107**, 063514 (2023).
[34] N. Grimm, C. Bonvin, and I. Tutusaus, Testing general relativity through the E_G statistic using the Weyl potential and galaxy velocities, Phys. Rev. Lett. **133**, 211004 (2024).
[35] R. Patki, N. Battaglia, and R. Bean, Probing gravity at large scales with KSZ-reconstructed velocities and CMB lensing, Phys. Rev. D **113**, 023527 (2026).
[36] W. J. Percival, L. Samushia, A. J. Ross, C. Shapiro, and A. Raccanelli, Redshift-space distortions, Phil. Trans. R. Soc. A **369**, 5058 (2011).
[37] T. Baldauf, R. E. Smith, U. Seljak, and R. Mandelbaum, Algorithm for the direct reconstruction of the dark matter correlation function from weak lensing and galaxy clustering, Phys. Rev. D **81**, 063531 (2010).
[38] R. Mandelbaum, C. M. Hirata, U. Seljak, U. Seljak, J. Guzik, N. Padmanabhan, C. Blake, M. R. Blanton, R. Lupton, and J. Brinkmann, Systematic errors in weak lensing: Application to SDSS galaxy-galalxy weak lensing, Mon. Not. R. Astron. Soc. **361**, 1287 (2005).
[39] S. More, S. Sugiyama, H. Miyatake *et al.*, Hyper Suprime-Cam Year 3 results: Measurements of clustering of SDSS-BOSS galaxies, galaxy-galaxy lensing, and cosmic shear, Phys. Rev. D **108**, 123520 (2023).
[40] A. Dvornik *et al.*, KiDS-1000: Combined halo-model cosmology constraints from galaxy abundance, galaxy clustering, and galaxy-galaxy lensing, Astron. Astrophys. **675**, A189 (2023).
[41] C. D. Leonard, P. G. Ferreira, and C. Heymans, Testing gravity with EG: Mapping theory onto observations, J. Cosmol. Astropart. Phys. 12 (2015) 051.
[42] N. Kaiser, Clustering in real space and in redshift space, Mon. Not. R. Astron. Soc. **227**, 1 (1987).
[43] C. Hernández-Aguayo, J. Hou, B. Li, C. M. Baugh, and A. G. Sánchez, Large-scale redshift space distortions in modified gravity theories, Mon. Not. R. Astron. Soc. **485**, 2194 (2019).
[44] G. Dvali, G. Gabadadze, and M. Porrati, 4D gravity on a brane in 5D Minkowski space, Phys. Lett. B **485**, 208 (2000).
[45] C. Deffayet, Cosmology on a brane in Minkowski bulk, Phys. Lett. B **502**, 199 (2001).
[46] F. Schmidt, Self-consistent cosmological simulations of DGP braneworld gravity, Phys. Rev. D **80**, 043001 (2009).
[47] F. Schmidt, Self-consistent cosmological simulations of DGP braneworld gravity, Phys. Rev. D **80**, 043001 (2009).
[48] Https://github.com/BartolomeoF/nDGPemu.
[49] B. Fiorini, K. Koyama, and T. Baker, Fast production of cosmological emulators in modified gravity: The matter power spectrum, J. Cosmol. Astropart. Phys. 12 (2023) 045.
[50] W. Hu and I. Sawicki, Models of f(R) cosmic acceleration that evade solar system tests, Phys. Rev. D **76**, 064004 (2007).
[51] I. Sáez-Casares, Y. Rasera, and B. Li, The e-MANTIS emulator: Fast predictions of the non-linear matter power spectrum in $f(r)$cdm cosmology, Mon. Not. R. Astron. Soc. **527**, 7242 (2024).
[52] Https://pypi.org/project/emantis/.
[53] M. Lochner, D. Scolnic, H. Almoubayyed *et al.*, The impact of observing strategy on cosmological constraints with LSST, Astrophys. J. Suppl. Ser. **259**, 58 (2022).
[54] R. Zhou, B. Dey, J. A. Newman *et al.*, Target selection and validation of DESI luminous red galaxies, Astron. J. **165**, 58 (2023).
[55] LSST DESC, R. Mandelbaum, T. Eifler, R. Hložek, T. Collett, E. Gawiser, D. Scolnic, D. Alonso, H. Awan, R. Biswas, J. Blazek *et al.*, The LSST Dark Energy Science Collaboration (DESC) science requirements document, arXiv:1809.01669.
[56] B. A. Reid, H.-J. Seo, A. Leauthaud, J. L. Tinker, and M. White, A 2.5 per cent measurement of the growth rate from small-scale redshift space clustering of SDSS-III CMASS galaxies, Mon. Not. R. Astron. Soc. **444**, 476 (2014).
[57] E. Kitanidis and M. White, Cross-correlation of Planck CMB lensing with DESI-like LRGs, Mon. Not. R. Astron. Soc. **501**, 6181 (2021).

[58] D. Kodwani, P. G. Ferreira, and D. Alonso, The effect on cosmological parameter estimation of a parameter dependent covariance matrix, Open J. Astrophys. **2** (2019).
[59] S. Singh, R. Mandelbaum, U. Seljak, A. Slosar, and J. Vazquez Gonzalez, Galaxy-galaxy lensing estimators and their covariance properties, Mon. Not. R. Astron. Soc. **471**, 3827 (2017).
[60] P. S. Behroozi, R. H. Wechsler, and H.-Y. Wu, The ROCKSTAR phase-space temporal halo finder and the velocity offsets of cluster cores, Astrophys. J. **762**, 109 (2012).
[61] L. Wang, B. Reid, and M. White, An analytic model for redshift-space distortions, Mon. Not. R. Astron. Soc. **437**, 588 (2014).
[62] J. Hartlap, P. Simon, and P. Schneider, Why your model parameter confidences might be too optimistic—unbiased estimation of the inverse covariance matrix, Astron. Astrophys. **464**, 399 (2007).
[63] B. Bose, M. Cataneo, T. Tröster, Q. Xia, C. Heymans, and L. Lombriser, On the road to per cent accuracy IV: ReACT–computing the non-linear power spectrum beyond ΛCDM, Mon. Not. R. Astron. Soc. **498**, 4650 (2020).
[64] DES Collaboration, T. M. C. Abbott, F. B. Abdalla, S. Avila *et al.*, Dark Energy Survey Year 1 Results: Constraints on extended cosmological models from galaxy clustering and weak lensing, Phys. Rev. D **99**, 123505 (2019).
[65] P. A. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. Banday, R. Barreiro, N. Bartolo, E. Battaner *et al.*, Planck 2015 results-XIV. Dark energy and modified gravity, Astron. Astrophys. **594**, A14 (2016).
[66] A. Mead, S. Brieden, T. Tröster, and C. Heymans, HMcode-2020: Improved modelling of non-linear cosmological power spectra with baryonic feedback, Mon. Not. R. Astron. Soc. **502**, 1401 (2021).
[67] R. Takahashi, M. Sato, T. Nishimichi, A. Taruya, and M. Oguri, Revising the halofit model for the nonlinear matter power spectrum, Astrophys. J. **761**, 152 (2012).
[68] S.-F. Chen, Z. Vlah, and M. White, Consistent modeling of velocity statistics and redshift-space distortions in one-loop perturbation theory, J. Cosmol. Astropart. Phys. 07 (2020) 062.
[69] J. E. McEwen, X. Fang, C. M. Hirata, and J. A. Blazek, FAST-PT: A novel algorithm to calculate convolution integrals in cosmological perturbation theory, J. Cosmol. Astropart. Phys. 09 (2016) 015.
[70] N. E. Chisari, D. Alonso, E. Krause, C. D. Leonard, P. Bull, J. Neveu, A. Villarreal, S. Singh, T. McClintock, J. Ellison *et al.*, Core cosmology library: Precision cosmological predictions for LSST, Astrophys. J. Suppl. Ser. **242**, 2 (2019).
[71] A. Gelman, X.-L. Meng, and H. Stern, Posterior predictive assessment of model fitness via realized discrepancies, Statistica Sinica **6**, 733 (1996).
[72] A. Nicola, A. Amara, and A. Refregier, Consistency tests in cosmology using relative entropy, J. Cosmol. Astropart. Phys. 01 (2019) 011.
[73] N. Porqueres, A. Heavens, D. Mortlock, and G. Lavaux, Bayesian forward modelling of cosmic shear data, Mon. Not. R. Astron. Soc. **502**, 3035 (2021).
[74] C. Doux, E. Baxter, P. Lemos *et al.*, Dark energy survey internal consistency tests of the joint cosmological probes analysis with posterior predictive distributions, Mon. Not. R. Astron. Soc. **503**, 2688 (2021).
[75] T. M. Abbott, M. Aguena, A. Alarcon *et al.*, Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing, Phys. Rev. D **105**, 023520 (2022).
[76] Planck Collaboration, N. Aghanim, Y. Akrami *et al.*, Planck 2018 results. VI. Cosmological parameters, Astron. Astrophys. **641**, A6 (2020).
[77] B. Hu, M. Raveri, M. Rizzato, and A. Silvestri, Testing Hu–Sawicki f(R) gravity with the effective field theory approach, Mon. Not. R. Astron. Soc. **459**, 3880 (2016).
[78] D. Wang, Can f (R) gravity relieve H 0 and σ8 tensions?, Eur. Phys. J. C **81**, 482 (2021).
[79] T. Baker, P. G. Ferreira, C. D. Leonard, and M. Motta, New gravitational scales in cosmological surveys, Phys. Rev. D **90**, 124030 (2014).
[80] C. Skordis, The tensor-vector-scalar theory and its cosmology, Classical Quantum Gravity **26**, 143001 (2009).
[81] C. R. Harris, K. J. Millman, S. J. van der Walt *et al.*, Array programming with NumPy, Nature (London) **585**, 357 (2020).
[82] P. Virtanen, R. Gommers, T. E. Oliphant *et al.*, SciPy 1.0: Fundamental algorithms for scientific computing in Python, Nat. Methods **17**, 261 (2020).
[83] J. D. Hunter, Matplotlib: A 2D graphics environment, Comput. Sci. Eng. **9**, 90 (2007).
[84] https://github.com/c-d-leonard/EG_forecasts.