Cosmology without window functions: Quadratic estimators for the galaxy power spectrum

Oliver H. E. Philcox*

Department of Astrophysical Sciences, Princeton University, Princeton, New Jersey 08540, USA and School of Natural Sciences, Institute for Advanced Study, 1 Einstein Drive, Princeton, New Jersey 08540, USA

(Received 17 December 2020; accepted 5 April 2021; published 5 May 2021)

Conventional algorithms for galaxy power-spectrum estimation measure the true spectrum convolved with a survey window function, which, for parameter inference, must be compared with a similarly convolved theory model. In this work, we directly estimate the unwindowed power spectrum multipoles using quadratic estimators akin to those introduced in the late 1990s. Under Gaussian assumptions, these are optimal and free from the leading-order effects of pixelization and non-Poissonian shot noise. They may be straightforwardly computed given the survey dataset and a suite of simulations of known cosmology. We implement the pixel-based maximum-likelihood estimator and a simplification based on the Feldman-Kaiser-Peacock weighting scheme, both of which can be computed via FFTs and conjugate-gradient descent methods. Furthermore, the estimators allow direct computation of spectrum coefficients in an arbitrary linear compression scheme, without needing to first bin the statistic. Applying the technique to a subset of the BOSS DR12 galaxies, we find that the pixel-based quadratic estimators give statistically consistent power spectra, compressed coefficients, and cosmological parameters to those obtained with the usual windowed approaches. Because of the sample's low number density and compact window function, the optimal weighting scheme gives little improvement over the simplified form; this may change for dense surveys or those focusing on primordial non-Gaussianity. The technique is shown to be efficient and robust, and shows significant potential for measuring the windowless power spectrum and bispectrum in the presence of weak non-Gaussianity.

DOI: 10.1103/PhysRevD.103.103504

I. INTRODUCTION

The fundamental building block of spectroscopic surveys is the two-point correlator of galaxy positions in real or Fourier space. Under the assumption of Gaussianity, this fully specifies the deterministic part of the density field. Our Universe may be considered Gaussian on large scales; thus, this still encapsulates much of the available information. Measuring such a quantity, be it the two-point correlation function or the power spectrum, has thus become a central tenet of cosmological research throughout the past few decades.

Almost all current surveys (e.g., [1,2]) have opted to measure the galaxy power spectrum $P(\mathbf{k})$ using the "Feldman-Kaiser-Peacock" (FKP) approach first proposed in Ref. [3] and generalized to anisotropic spectra in Refs. [4–7], as well as light cone evolution in Ref. [8]. This estimates the galaxy power spectrum by performing a Fourier transform on the weighted (galaxy-minus-random) field, where the randoms are a set of artificial galaxies that are unclustered, yet encapsulate the survey geometry and selection function. A weight is applied to each galaxy individually, equal to $1/[1+\bar{n}(\mathbf{r})P_{\rm FKP}]$, where $\bar{n}(\mathbf{r})$ is the unclustered number density and $P_{\rm FKP}$ some constant. As shown in Ref. [9], this is the minimum variance solution for Gaussian fluctuations on small scales, where the spectrum is shot-noise dominated. On large scales, however, it is suboptimal, since it (a) does not account for the variations of $P(\mathbf{k})$ with respect to scale [assuming $P(\mathbf{k}) \approx P_{\rm FKP}$] and (b) ignores correlations *between* specific pixels, since the weighting is not applied in a pairwise fashion. Furthermore, as is well known, the FKP method estimates the true underlying galaxy power spectrum convolved with the square of the window function rather than the true spectrum itself.

In the late 1990s, a slew of work was devoted to avoiding the pitfalls of the FKP estimator, using a more general approach: the *quadratic estimators*. Developed originally for analysis of the cosmic microwave background (CMB) [10–14], these can be similarly applied to galaxy surveys [9,15,16] and allow one to compute *unwindowed* estimates of the underlying power spectra. At heart, a quadratic estimator is simply a quadratic function applied to the dataset, allowing for each pair of particles (or pixels) to be

*ohep2@cantab.ac.uk

 © 2021 American Physical Society

differently weighted. It can be shown that the *optimal* power-spectrum estimator is a member of this class (under Gaussian assumptions); thus, its application would be expected to give the smallest errors on cosmological parameters. The estimators can be further tuned to weight galaxies also according to their physical properties [17,18] and used to place competitive constraints on primordial non-Gaussianity [19].

If quadratic estimators are so useful, why then have they not been adopted by recent surveys? Indeed, they were used extensively at the start of the 21st century (e.g., [20–23]), but have since fallen out of vogue. One reason for this lies in the computational cost; the optimal estimator is difficult to implement since it requires knowledge of the inverse covariance between any two galaxy positions (or pixels). Current surveys contain $\mathcal{O}(10^6)$ objects; thus, a brute-force computation of this inverse is unfeasible. Second, the FKP scheme is often found to perform well in practice. For a survey containing both uniform number density $\bar{n}$ and narrow bins, the FKP weights become optimal, just as on small scales. The gains from the optimal approach are expected to be most relevant on scales where the window function has significant power. Given that these are usually dominated by cosmic variance, their improved measurement is not expected to have great effect on cosmological analyses, unless one is searching for signatures only visible at low k (e.g., primordial non-Gaussianity).[1] Finally, the issue of windowing is not usually significant, since our theory models can be straightforwardly window convolved (e.g., [1]). That said, in the current era of "precision cosmology," our goal is to squeeze out every last drop of cosmological information; thus, any gains, even if small, are of great importance.[2] Furthermore, the ability to estimate unwindowed spectra will be of great use for higher-order statistics, whereupon window convolution is highly nontrivial.

In this work, we implement a variant of the quadratic power-spectrum estimators of old, considering both the optimal approach and a related scheme which allows for easier computation, akin to the FKP formalism. We provide a pedagogical discussion of their derivation and application, providing forms which can be efficiently applied, bypassing some of the usual computational caveats (following similar approaches to that done for the CMB in Refs. [14,26]). The scheme can be similarly used to measure coefficients of the power spectrum under some linear compression scheme, as in Ref. [27], and the accompanying subtleties of this are discussed. By formulating our estimators as a difference between quantities measured in the data and simulations, we are free from the leading-order effects of pixelization, binning, non-Poissonian shot noise, and fingers of God, and obtain power-spectrum estimates that are not convolved with the window function, with the integral constraint isolated to the first k bin. Indeed, we demonstrate that the resulting constraints on cosmological parameters from BOSS data analyzed with the quadratic estimators are statistically compatible with those obtained from the usual windowed approach.

While we do not expect significant gains from applying the optimal estimator to BOSS, this is a strongly survey-dependent statement. For dense small-volume datasets such as the DESI Bright Galaxy Survey [28], the improvements will be of greater importance, and the situation is similar for analyses constraining $f_{\rm NL}$ (demonstrated in Ref. [19] for the eBOSS quasar sample). The formalism additionally extends beyond the Gaussian limit and can be used to implement optimal estimators for the two- and three-point statistics in the presence of weak non-Gaussianity; our work thus provides a proof of concept for such applications. Computing bispectrum estimates free from the survey window function would be a significant achievement, given the inherent difficulty of performing window convolution in this case (e.g., [29]).

The structure of this paper is as follows. We begin by presenting a pedagogical introduction to the quadratic power-spectrum estimator in Sec. II, before we discuss its practical implementation in Sec. III. Section IV details the modifications required for the estimators to measure compressed subspace coefficients, following which, we present a demonstration of the approach in Sec. V, computing both spectra and corresponding cosmological parameter constraints. We conclude in Sec. VI, with Appendixes A–D providing useful mathematical results, consistency tests, and a discussion of the optimal estimators for the power spectrum beyond the Gaussian limit. Given the technical nature of this work, we provide a road map for two types of readers: Those who are principally interested in the quadratic estimator phenomenology and application to galaxy surveys are advised to read Secs. II–IV, while those concerned principally with its efficacy when applied to BOSS data (and beyond) should skip directly to Sec. V.

II. POWER-SPECTRUM ESTIMATORS

We begin with a discussion of quadratic power-spectrum estimators, discussing both their general and optimal form, before specializing to their application in spectroscopic surveys. Analogous estimators for the coefficients of the power spectrum following some linear compression scheme will be discussed in Sec. IV. Several of the derivations below follow Refs. [9,12,14,23], but we recapitulate them for clarity.

[1]See Ref. [24] for an extensive comparison of quadratic and FKP-type (pseudo-C_ℓ) estimators in the CMB context.

[2]For another notable work, see Ref. [25], which estimates the *initial* power spectrum via a Bayesian approach.

A. General quadratic estimators

Consider a vector $\mathbf{d}$ of observations, e.g., the pixelized overdensity for a galaxy survey in $N_{\rm pix}$ cells. We model this as the sum of two contributions: a signal $\mathbf{m}$ and a noise component $\mathbf{n}$ satisfying $\langle \mathbf{m}\mathbf{m}^T\rangle = \mathsf{S}$ and $\langle \mathbf{n}\mathbf{n}^T\rangle = \mathsf{N}$, respectively, for signal and noise covariances S and N, where $\langle \ldots \rangle$ indicates an average over both the underlying density fields and their Poisson realizations. Assuming the noise to be uncorrelated with the signal, the covariance of $\mathbf{d}$ is given by $\langle \mathbf{d}\mathbf{d}^T\rangle = \mathsf{C}_D = \mathsf{S} + \mathsf{N}$. This can be expressed as a function of band powers $\mathbf{p} \equiv \{p_\alpha\}$,[3] such that $\mathsf{C}_D \equiv \mathsf{C}(\mathbf{p}^{\rm true})$, where $\mathsf{C}(\mathbf{p})$ is the covariance defined at using an arbitrary set of band powers $\mathbf{p}$. If the covariance is a linear function of the band powers, i.e., $\mathsf{C}(\mathbf{p})$ is linear in $\mathbf{p}$, a general quadratic estimator for $\mathbf{p}$ is

$$\hat{q}_\alpha^{\rm QE} = \frac{1}{2}\mathbf{d}^T\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}\mathbf{d} \equiv \frac{1}{2}{\rm Tr}[(\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1})\mathbf{d}\mathbf{d}^T], \tag{1}$$

where $\mathsf{C}_{,\alpha} = \mathsf{S}_{,\alpha} \equiv \frac{\partial \mathsf{S}}{\partial p_\alpha}$ is the derivative of the signal with respect to the band powers, and H is some (positive-definite and symmetric) pixel weighting matrix. Both matrices have dimension $N_{\rm pix} \times N_{\rm pix}$, and their action is to weight the data by some linear function H^{-1} (which may or may not be diagonal), then to extract the piece depending on p_α via the derivative term.

$\hat{q}_\alpha$ is neither an unbiased nor optimal estimator of the band powers. In expectation, we obtain

$$\begin{aligned}\mathbb{E}[\hat{q}_\alpha^{\rm QE}] &= \frac{1}{2}{\rm Tr}[\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}\mathsf{C}_D] \\ &= \frac{1}{2}{\rm Tr}[\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}\mathsf{N}] + \frac{1}{2}\sum_\beta p_\beta^{\rm true}{\rm Tr}[\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}\mathsf{C}_{,\beta}],\end{aligned} \tag{2}$$

rewriting the data covariance C_D in terms of the noise covariance and signal derivatives for true band powers $\mathbf{p}^{\rm true}$.[4] To debias (1), we require both an additive term to subtract the first (noise-induced) contribution to $\hat{q}_\alpha^{\rm QE}$ and a multiplicative one to cancel the second trace term. Incorporating these leads to the general quadratic estimator defined by

$$\boxed{\hat{p}_\alpha^{\rm QE} = p_\alpha^{\rm fid} + \sum_\beta F^{-1,{\rm QE}}_{\alpha\beta}(\hat{q}_\beta^{\rm QE} - \bar{q}_\beta^{\rm QE}),} \tag{3}$$

where we have defined the Fisher and bias terms

$$\begin{aligned}F_{\alpha\beta}^{\rm QE} &= \frac{1}{2}{\rm Tr}[\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}\mathsf{C}_{,\beta}], \\ \bar{q}_\alpha^{\rm QE} &= \frac{1}{2}{\rm Tr}[\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}\mathsf{C}_{\rm fid}],\end{aligned} \tag{4}$$

again depending on the weighting matrix H. We have further introduced a set of *fiducial* band powers $\mathbf{p}^{\rm fid}$ and corresponding covariance $\mathsf{C}_{\rm fid} \equiv \mathsf{C}(\mathbf{p}^{\rm fid})$; while these could alternatively be absorbed into the bias term $\bar{\mathbf{q}}$, alongside a redefinition $\bar{q}_\beta \to \frac{1}{2}{\rm Tr}[\mathsf{H}^{-1}\mathsf{C}_{,\beta}\mathsf{H}^{-1}\mathsf{N}]$, we generally avoid this, as it makes the estimators more sensitive to unmeasured k bins, non-Poissonian shot noise, and gridding artifacts. Further discussion of this can be found in Sec. III D. The computation of each term in (3) will be elaborated upon in subsequent sections.

As shown in Appendix A, the general quadratic estimator is unbiased for all choices of symmetric and invertible weighting matrix H (assuming that the band powers $\mathbf{p}$ fully define the signal covariance S), such that $\mathbb{E}[\hat{p}_\alpha^{\rm QE}] = p_\alpha^{\rm true}$. This remains valid even when the underlying density field is non-Gaussian, in which case not all information is encoded within the two-point covariance C_D. In the Gaussian limit, the estimator variance is given by

$$\begin{aligned}{\rm cov}(\hat{p}_\alpha^{\rm QE}, \hat{p}_\beta^{\rm QE}) = \frac{1}{2}\sum_{\gamma\delta} F^{-1,{\rm QE}}_{\alpha\gamma}F^{-1,{\rm QE}}_{\beta\delta} \\ \times {\rm Tr}[\mathsf{H}^{-1}\mathsf{C}_{,\gamma}\mathsf{H}^{-1}\mathsf{C}_D\mathsf{H}^{-1}\mathsf{C}_{,\delta}\mathsf{H}^{-1}\mathsf{C}_D]\end{aligned} \tag{5}$$

(Appendix A), which depends on our choice of H. Clearly, one must carefully choose the pixel weight matrix to obtain an estimator with low variance; in subsequent sections, we will consider both the minimum variance solution and a simpler approximation.

B. The maximum-likelihood estimator

Under Gaussian assumptions, the maximum-likelihood (ML) estimator for the band powers $\mathbf{p}$ is just a special case of (3), with $\mathsf{H} = \mathsf{C}_D$. To demonstrate this, we begin by constructing a log-likelihood for the data as

$$\begin{aligned}\mathcal{L}[\mathbf{d}](\mathbf{p}) &= -2\log L[\mathbf{d}](\mathbf{p}) \\ &= \mathbf{d}^T\mathsf{C}^{-1}(\mathbf{p})\mathbf{d} + {\rm Tr}\log\mathsf{C}(\mathbf{p}) + {\rm const}\end{aligned} \tag{6}$$

assuming that $\langle \mathbf{d}\rangle = \mathbf{0}$ (i.e., that $\mathbf{d}$ is mean subtracted) and that the likelihood is Gaussian (valid if the number of modes is large and the underlying density field is Gaussian). Note that the band powers enter only through the data covariance $\mathsf{C}(\mathbf{p})$, with $\mathsf{C}_D \equiv \mathsf{C}(\mathbf{p}^{\rm true})$, as before.

[3] We use greek letters $\alpha, \beta, \ldots$ to index band powers and latin indices $i, j, \ldots$ to denote components of $\mathbf{d}$, i.e., pixels. We will later define the band powers as estimates of some power-spectrum multipole integrated over a finite k bin.

[4] Strictly, this is only true if $\mathbf{p}$ contains all possible band powers; i.e., it extends over all wave vectors. In practice, truncating to a finite range of wave numbers is a valid approximation; though, for this reason, one may wish to estimate a slightly larger number of band powers than used in the final analysis. This point is discussed in more detail in Sec. IV B.

To obtain a ML estimator for $\mathbf{p}$, we must extremize (6). Practically, this is difficult since the band powers do not enter the log-likelihood linearly; thus, we first expand $\mathcal{L}(\mathbf{p})$ around some fiducial spectrum $\mathbf{p}^{\rm fid}$,

$$\mathcal{L}(\mathbf{p}^{\rm fid}+\delta\mathbf{p}) \approx \mathcal{L}(\mathbf{p}^{\rm fid})+\delta\mathbf{p}^T\nabla_{\mathbf{p}}\log\mathcal{L}+\frac{1}{2}\delta\mathbf{p}^T(\nabla_{\mathbf{p}}\nabla_{\mathbf{p}'}\log\mathcal{L})\delta\mathbf{p}', \quad (7)$$

where $\delta\mathbf{p}=\mathbf{p}-\mathbf{p}^{\rm fid}$ and the gradients are evaluated at $\mathbf{p}^{\rm fid}$. Assuming $\mathbf{p}^{\rm fid}$ is sufficiently close to the true spectrum (such that $\mathsf{C}_{\rm fid}\approx\mathsf{C}_D$), $\delta\mathbf{p}$ can be estimated by minimizing (7)

$$\delta\mathbf{p}=-[\nabla_{\mathbf{p}}\nabla_{\mathbf{p}'}\mathcal{L}]^{-1}\nabla_{\mathbf{p}'}\mathcal{L}, \quad (8)$$

which is just a first-order Newton-Raphson estimate. Inserting the Gaussian likelihood and simplifying gives the maximum-likelihood[5] estimator for a band power p_α,

$$\boxed{\hat{p}_\alpha^{\rm ML}=p_\alpha^{\rm fid}+\frac{1}{2}\sum_\beta F^{-1}_{\alpha\beta}(\mathbf{d}^T\mathsf{C}^{-1}\mathsf{C}_{,\beta}\mathsf{C}^{-1}_{\rm fid}\mathbf{d}-{\rm Tr}[\mathsf{C}^{-1}_{\rm fid}\mathsf{C}_{,\beta}])}, \quad (9)$$

where we have assumed (as is standard practice) that the curvature matrix $\nabla_{\mathbf{p}}\nabla_{\mathbf{p}'}\mathcal{L}$ can be replaced by its (realization-averaged) Fisher matrix defined by

$$F_{\alpha\beta}=\frac{1}{2}{\rm Tr}[\mathsf{C}^{-1}_{\rm fid}\mathsf{C}_{,\alpha}\mathsf{C}^{-1}_{\rm fid}\mathsf{C}_{,\beta}] \quad (10)$$

[9,12,14,23].[6] Writing this in the form of (3)

$$\hat{p}_\alpha^{\rm ML}=p_\alpha^{\rm fid}+\sum_\beta F^{-1}_{\alpha\beta}(\hat{q}_\beta-\bar{q}_\beta), \quad (11)$$

we see that this is just a special case of the general quadratic estimator, with H equal to the covariance matrix in the fiducial cosmology $\mathsf{C}_{\rm fid}$.

As shown in Appendix A, the ML solution is optimal (in the Cramér-Rao sense) if $\mathsf{C}_{\rm fid}$ is equal to the true data covariance C_D (i.e., $\mathbf{p}^{\rm fid}=\mathbf{p}^{\rm true}$) and the underlying density field is Gaussian, such that all connected moments of the density field above the two-point function are vanishing. In this case, the covariance of $\hat{p}_\alpha$ is simply the inverse Fisher matrix $F^{-1}_{\alpha\beta}$.

In reality, the Universe is non-Gaussian except on the largest scales; thus, the quadratic estimator is not strictly optimal. In the weakly non-Gaussian limit, one may construct the leading-order corrections to the maximum-likelihood estimator for $\mathbf{p}^{\rm fid}$ via an Edgeworth expansion (e.g., [30]). These start cubic in the data and are discussed in Appendix B. For most galaxy surveys, this is of limited use, since the power spectrum becomes dominated by shot noise on relatively large scales, though the approach may be of use for dense surveys such as the DESI Bright Galaxy Survey [28].

C. Specialization to spectroscopic surveys

1. Covariance definitions

The quadratic estimator (3) and its optimal variant (9) provide a general framework in which to measure a set of band powers $\mathbf{p}$ from a dataset $\mathbf{d}$ given the signal and noise covariances. For spectroscopic surveys, our dataset is a list of galaxy positions alongside a set of random particles which encode the survey geometry. Here, we define $\mathbf{d}$ as the pixelized field of (data minus random), i.e.,

$$d_i\equiv n_g(\mathbf{r}_i)-\alpha n_r(\mathbf{r}_i), \quad (12)$$

where $\mathbf{r}_i$ is the center of the ith pixel, and n_g (n_r) is the galaxy (random particle) density defined on a grid following some mass assignment scheme, for example, cloud in cell or the more nuanced schemes of Ref. [9] (which can additionally be used to null certain troublesome modes, such as those dominated by fingers-of-God effects). The coefficient α is chosen such that $\langle\mathbf{d}\rangle=0$, i.e., $\alpha=\int d\mathbf{r}n_g(\mathbf{r})/\int d\mathbf{r}n_r(\mathbf{r})$.[7]

The covariance matrix of $\mathbf{d}$ is given by $\mathsf{C}_{ij}=\mathsf{C}(\mathbf{r}_i,\mathbf{r}_j)$,[8] where

$$\begin{aligned}\mathsf{C}(\mathbf{r},\mathbf{r}') &= \langle[n_g-\alpha n_r](\mathbf{r})[n_g-\alpha n_r](\mathbf{r}')\rangle\\ &= n(\mathbf{r})n(\mathbf{r}')\xi(\mathbf{r},\mathbf{r}')+(1+\alpha)n(\mathbf{r})\delta_D(\mathbf{r}-\mathbf{r}')\\ &\equiv \mathsf{S}(\mathbf{r},\mathbf{r}')+\mathsf{N}(\mathbf{r},\mathbf{r}'),\end{aligned} \quad (13)$$

assuming Poisson statistics and writing $\langle n_g(\mathbf{r})\rangle=\alpha\langle n_r(\mathbf{r})\rangle=n(\mathbf{r})$ (suppressing the dependence of C on $\mathbf{p}$ for clarity). We have used that the zero-noise part of the

[5]Strictly, this is only the "maximum likelihood" if $\mathbf{p}^{\rm fid}$ is equal to the true spectrum. This limit can be achieved by iteration: First define some fiducial spectrum $\mathbf{p}^{\rm fid,0}$, which is used in (8) to find an updated estimate $\mathbf{p}^{\rm fid,1}$ and repeat until convergence is reached. While this is possible (and discussed in Ref. [12]), it is time consuming, and the estimator is no longer strictly quadratic. Furthermore, the degree of suboptimality is quadratic in the difference between C_D^{-1} and $\mathsf{C}_{\rm fid}^{-1}$, and thus usually small.

[6]See [13] for a CMB analysis that does not make this assumption.

[7]We may additionally include particle weights in the definition of $\mathbf{d}$, for example, to encapsulate systematics and fiber collisions (but not FKP weights). This gives $d_i\equiv w_g(\mathbf{r}_i)n_g(\mathbf{r}_i)-\alpha w_r(\mathbf{r}_i)n_r(\mathbf{r}_i)$ where $\alpha=\int d\mathbf{r}w_g(\mathbf{r})n_g(\mathbf{r})/\int d\mathbf{r}w_r(\mathbf{r})n_r(\mathbf{r})$. Such weights are included in the analysis of data in Sec. V.

[8]For interpretability, we write most quantities in continuous form, noting that any integrals are strictly sums over the pixel grid, and any Fourier transforms are discrete.

galaxy density field can be written $n_g(\mathbf{r}) = n(\mathbf{r})[1+\delta(\mathbf{r})]$, where $\delta(\mathbf{r})$ has the correlation function $\xi(\mathbf{r},\mathbf{r}') = \langle\delta(\mathbf{r})\delta(\mathbf{r}')\rangle$, and ignored the effects of the pixelization scheme.[9] Writing the correlation function (which, in the general case, will be that of galaxies in redshift space) in Fourier space gives[10]

$$\mathsf{C}(\mathbf{r},\mathbf{r}') = n(\mathbf{r})n(\mathbf{r}')\int_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}\sum_{\ell} P_\ell(k)L_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{r}}') + (1+\alpha)n(\mathbf{r})\delta_D(\mathbf{r}-\mathbf{r}') \tag{14}$$

denoting $\int_{\mathbf{k}} \equiv (2\pi)^{-3}\int d\mathbf{k}$ and expanding in multipoles via the Legendre polynomial L_ℓ, as in the Yamamoto formalism for redshift-space distortions [5].[11] We may similarly define the band-power derivatives assuming $P_\ell(k) = \sum_a \Theta_a(\mathbf{k})P_\ell^a$ for some (thin) k bins with bin weights $\Theta^a(\mathbf{k})$ which are unity if $\mathbf{k}$ is in $|\mathbf{k}|$ bin a and zero else,

$$\mathsf{C}_{,\alpha}(\mathbf{r},\mathbf{r}') = n(\mathbf{r})n(\mathbf{r}')\int_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}\Theta_a(\mathbf{k})L_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{r}}'), \tag{15}$$

where the index α refers to the $\{a,\ell\}$ pair. The above formulas allow the computation of the band powers $\mathbf{p} \equiv \{P_\ell^a\}$ via (3) or (9), in particular, allowing for the implementation of a close-to-optimal estimator by applying the $\mathsf{C}_{\rm fid}^{-1}$ weights to the (data minus random) pixel grid.

2. Relation to standard estimators

Before continuing, we pause briefly to compare the ML estimator of (9) with the windowed-FKP estimator of Ref. [3] conventionally used in the literature (or its more modern variants [6,7]). Although the ML approach requires significantly more computational expense than a simple FKP estimate (mainly due to the necessity to invert the fiducial covariance $\mathsf{C}_{\rm fid}$), it has two key benefits: (1) It produces minimum variance error bars on the power spectrum (and hence, derived parameters), with a particular improvement seen on large scales (the FKP weights are optimal on short scales [9]). (2) By including the noise model and survey geometry in the covariances $\mathsf{C}(\mathbf{p})$, the estimator recovers the *true* unwindowed power spectrum. Furthermore, if one assumes that the shot noise is treated correctly in the fiducial power spectrum, the band-power estimates will also be shot-noise-free; in general, we will leave a free shot noise in the eventual parameter analysis to account for its possible non-Poissonian nature (which may be different in data and simulations).

Our approach requires a fiducial (unwindowed) power spectrum, and, for practical application, a set of simulated realizations, with the same geometry and noise properties as the data (see Sec. III). Practically, we may use a suite of N-body simulations for this purpose. Though the unwindowed galaxy power spectra are not usually known *a priori*, one can fit the mean of the windowed-FKP mock spectra with a window-convolved model such as one-loop effective field theory (hereafter EFT) [32] (given that the cosmology is already known) and use the preconvolved best fit to set the fiducial power spectrum. To avoid biasing the estimated power spectra, we require the theory model to be accurate, such that the model power spectrum is close to the truth; this just requires the χ^2 of the fit to be small. There is no requirement for the fiducial model to match that used to analyze the data however; thus, we do not need to recompute the statistic when comparing different theory prescriptions. An alternative approach would be to *deconvolve* the window from the simulated power spectra to use as a fiducial model, though this is difficult given the finite k range of the observed spectra. We further note that it is possible to remove dependence on the fiducial power spectrum entirely (except in the weights), as discussed below (27), though we retain it to reduce some higher-order effects such as discretization bias.

3. FKP weights

As mentioned above, the ML estimator requires the inverse covariance matrix $\mathsf{C}_{\rm fid}^{-1}$, which, due to its high dimensionality, is difficult and time consuming to compute. For this reason, we consider an alternative quadratic estimator in the form of (3), with weights based on the FKP approach. This uses the weighting matrix

$$\mathsf{H}_{\rm FKP}(\mathbf{r},\mathbf{r}') = n(\mathbf{r})n(\mathbf{r}')P_{\rm FKP}\delta_D(\mathbf{r}-\mathbf{r}') + n(\mathbf{r})\delta_D(\mathbf{r}-\mathbf{r}'), \tag{16}$$

[9]It is possible to self-consistently include this by noting that any pixelized quantity $\tilde{x}$ is equal to $\tilde{x} \equiv \psi * x$, where x is the underlying field, $*$ indicates a convolution, and ψ is the mass assignment window function. Thus, $\mathsf{C}(\mathbf{r},\mathbf{r}') \to [\psi * \mathsf{C} * \psi](\mathbf{r},\mathbf{r}')$. This is more difficult to evaluate, however, since the noise covariance is no longer diagonal, and the field $n(\mathbf{r})$ appearing in C is the *unconvolved* mean field estimated from $\psi^{-1} * \tilde{n}$. This leads to a significant increase in the number of Fourier transforms that need be performed to compute any quantity involving C. Ignoring these effects is a valid assumption in practice, since they are canceled at leading order in the difference estimator of (3).

[10]We neglect the impact of redshift evolution on the galaxy power spectrum, which is valid if the radial extent of the sample is relatively small. One may alternatively include this at linear order, writing $P(k,z) \approx P(k,z_{\rm eff}) \times (D(z)/D(z_{\rm eff}))^2$, leading to two additional (spatially varying) factors of $D(z)/D(z_{\rm eff})$ being inserted into the covariance, giving a slightly modified weight. Technically, this is true only in the absence of redshift space distortions, since $f(z)$ evolves differently in redshift to $D(z)$. Full treatment of this is beyond the scope of this work, though is important for $f_{\rm NL}$-based studies such as Ref. [19].

[11]A more accurate case would be to expand $\xi(\mathbf{r},\mathbf{r}')$ in the angle between $(\mathbf{r}-\mathbf{r}')$ and $(\mathbf{r}+\mathbf{r}')/2$, as in Ref. [31]. The Yamamoto approximation assumes that the line of sight can be approximated as $\mathbf{r}'$, which goes beyond the flat-sky approximation. For the most general treatment, though one beyond the scope of this work, one would instead allow the power spectrum to depend on *two* lines of sight.

which may be seen as the small-scale limit of C with $P(\mathbf{k}) \to P_{\rm FKP} \sim 10^4 h^{-3}\,{\rm Mpc}^3$ [9], additionally setting $\alpha = 0$. Equivalently, this is a simplified form of C, assuming the window function to be compact, such that its characteristic width L satisfies $kL \gg 1$ for all scales of interest k. This is trivially inverted:

$$\mathsf{H}^{-1}_{\rm FKP}(\mathbf{r},\mathbf{r}') = \frac{1}{n(\mathbf{r})[1+n(\mathbf{r})P_{\rm FKP}]}\delta_D(\mathbf{r}-\mathbf{r}'). \quad (17)$$

Inserting this into the general quadratic estimator (3), by construction, gives an unbiased estimator of the unwindowed galaxy power spectrum, which we expect to approach the optimal solution on small scales. Note that this is *not* identical to the conventional FKP scheme of Ref. [3], which assigns weights $w_{\rm FKP}(\mathbf{r}) = 1/(1+\bar{n}(\mathbf{r})P_{\rm FKP})$ to galaxies and random particles, where $\bar{n}(\mathbf{r})$ is some (spatially varying) mean density. Our formalism applies the weight to the grid directly rather than the particles, and normalizes by $n(\mathbf{r})$ rather than some survey-averaged quantity; this is necessary to ensure that we recover the *unwindowed* power-spectrum estimates, unlike in Ref. [3]. Below, we will compare band-power estimates computed using $\mathsf{H}_{\rm FKP}$ to those with the full fiducial covariance $\mathsf{C}_{\rm fid}$. We note that the quadratic scheme with FKP weights is similar to that proposed in Ref. [19] for the measurement of quasar spectra, though their formalism still required window-function convolution.

III. IMPLEMENTATION

To apply the power-spectrum estimators of Sec. II, we perform the following procedure:

1. Given a set of simulations with known cosmology, fit the mean window-convolved power spectrum to the one-loop EFT model of Ref. [32] to determine the best-fit nuisance parameters. The unconvolved theory model (subtracting Poissonian shot noise) gives a smooth model for the fiducial $\mathbf{p}^{\rm fid}$ spectra.
2. Paint the data and randoms to a regular Cartesian grid to define the data vector $\mathbf{d}$.
3. For each ($|\mathbf{k}|$ and multipole) bin α, compute the quadratic estimator terms $\mathbf{d}^T\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}\mathbf{d}$ with $\mathsf{H} = \mathsf{H}_{\rm FKP}$ (16) or $\mathsf{H} = \mathsf{C}_{\rm fid}$ (14). For the latter case, covariances are computed using the (smooth) fiducial power spectra.
4. Compute the bias term and Fisher matrix, as defined in (11) and (4) for the ML and FKP estimators. In both cases, this is done via Monte Carlo methods using a set of simulations which satisfy $\langle \mathbf{d}\mathbf{d}^T\rangle = \mathsf{H}$ (either the simulations used to define the fiducial cosmology for the ML estimators, or rescaled subsets of the random particles for the FKP case).
5. Compute the fiducial power spectrum $\mathbf{p}^{\rm fid}$ in the relevant bins and multipoles from the smooth model.
6. Accumulate the estimators (3) or (9) using the $\mathsf{H}^{-1}_{\rm FKP}$ or $\mathsf{C}^{-1}_{\rm fid}$ pixel weight matrices.

Below, we give practical details concerning the implementation of each of the above steps. All computations are performed using custom PYTHON code making use of the PYFFTW library, with some elements adapted from the PYLIANS and NBODYKIT [33] packages.

A. Computation of $\mathsf{C}_{\rm fid}$ and $\mathsf{C}_{,\alpha}$

To implement the ML quadratic estimators, we must be able to apply the $\mathsf{C}_{\rm fid}$ and $\mathsf{C}_{,\alpha}$ matrices to the pixelized galaxy field.[12] Assuming the density field to have a total of $N_{\rm pix} \sim 10^7$ pixels, the matrices have dimension $N^2_{\rm pix}$; thus, their explicit computation and storage is unfeasible. Instead, we consider their action on a general pixelized field $x(\mathbf{r})$, starting from the definition (14)[13]

$$\begin{aligned}\mathsf{C}[x](\mathbf{r}) &\equiv \int d\mathbf{r}'\mathsf{C}(\mathbf{r},\mathbf{r}')x(\mathbf{r}')\\ &= n(\mathbf{r})\int_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}}\int d\mathbf{r}' e^{-i\mathbf{k}\cdot\mathbf{r}'}n(\mathbf{r}')x(\mathbf{r}')\\ &\quad\times\sum_\ell P_\ell(k)L_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{r}}') + (1+\alpha)n(\mathbf{r})x(\mathbf{r}). \quad (18)\end{aligned}$$

Here, $n(\mathbf{r})$ is the unclustered galaxy density (depending on the survey selection function and mask), which we set equal to the pixelized density field of random particles, and $\{P_\ell(k)\}$ are the fiducial (unwindowed) power-spectrum multipoles. In the flat-sky limit, $L_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{r}}') \to L_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{z}})$ (i.e., we have a uniform line-of-sight $\hat{\mathbf{z}}$ across the survey), this is straightforward to implement; the signal covariance term is $n(\mathbf{r})\mathcal{F}^{-1}[P(\mathbf{k})\mathcal{F}[nx](\mathbf{k})](\mathbf{r})$ for forward (inverse) Fourier operator $\mathcal{F}$ ($\mathcal{F}^{-1}$). In the Yamamoto case, we can expand L_ℓ via the spherical harmonic addition theorem ([34], Eq. 14.30.9), yielding

$$\begin{aligned}\mathsf{C}[x](\mathbf{r}) &= n(\mathbf{r})\sum_\ell\frac{4\pi}{2\ell+1}\sum_{m=-\ell}^{\ell}\int_{\mathbf{k}}P_\ell(k)e^{i\mathbf{k}\cdot\mathbf{r}}Y^*_{\ell m}(\hat{\mathbf{k}})\\ &\quad\times\int d\mathbf{r}'Y_{\ell m}(\mathbf{r}')n(\mathbf{r}')x(\mathbf{r}')e^{-i\mathbf{k}\cdot\mathbf{r}'} + (1+\alpha)n(\mathbf{r})x(\mathbf{r})\\ &= n(\mathbf{r})\sum_\ell\frac{4\pi}{2\ell+1}\sum_{m=-\ell}^{\ell}\mathcal{F}^{-1}[P_\ell(k)Y^*_{\ell m}(\hat{\mathbf{k}})\\ &\quad\times\mathcal{F}[Y_{\ell m}nx](\mathbf{k})](\mathbf{r}) + (1+\alpha)n(\mathbf{r})x(\mathbf{r}). \quad (19)\end{aligned}$$

[12] While $\mathsf{C}_{\rm fid}$ does not appear directly in (9), it is required to efficiently compute the action of $\mathsf{C}^{-1}_{\rm fid}$.

[13] We ignore the dependence of C on $\mathbf{p}$ throughout this section for clarity.

In practice, this is computed with $\frac{1}{2}(1+\ell_{\rm max})(2+\ell_{\rm max})$ real-to-Fourier FFTs using the real form of the spherical harmonics, as in Ref. [7]. This assumes the fiducial spectra are nonzero for all even ℓ up to $\ell_{\rm max}$. A similar expression holds for the application of $\mathsf{C}_{,\alpha}$ to a generic map x starting from (15)

$$\mathsf{C}_{,\alpha}[x](\mathbf{r}) = n(\mathbf{r})\frac{4\pi}{2\ell+1}\sum_{m=-\ell}^{\ell}\mathcal{F}^{-1}[\Theta_a(\mathbf{k})Y^*_{\ell m}(\hat{\mathbf{k}})\mathcal{F}[Y_{\ell m}nx](\mathbf{k})](\mathbf{r}) \tag{20}$$

requiring only $(2\ell+1)$ FFTs. While it is somewhat simpler to consider the action of $\mathsf{C}_{,\alpha}$ on a *pair* of fields (which requires two real-to-complex Fourier transforms and a Fourier-space summation), the action of $\mathsf{C}_{,\alpha}$ on a *single* vector is useful when we wish to apply a number of operations in series, i.e., $\mathsf{C}_{,\alpha}$ then $\mathsf{C}_{\rm fid}^{-1}$.

B. Computation of $\mathsf{C}_{\rm fid}^{-1}$ and $\mathsf{H}_{\rm FKP}^{-1}$

An evaluation of the ML estimator requires the inverse-fiducial-covariance weighted data vector $\mathsf{C}_{\rm fid}^{-1}\mathbf{d}$. Given that the covariance $\mathsf{C}_{\rm fid}$ is too large to be stored and manipulated, we cannot compute $\mathsf{C}_{\rm fid}^{-1}$ directly; thus, we instead opt to use preconditioned conjugate-gradient descent (CGD) methods, as in Ref. [14]. Effectively, this numerically solves the equation $\mathsf{C}_{\rm fid}^{-1}\mathbf{x}=\mathbf{y}$ by finding $\mathbf{y}$ such that $\tilde{\mathsf{C}}^{-1}\mathsf{C}_{\rm fid}\mathbf{y}=\tilde{\mathsf{C}}^{-1}\mathbf{x}$, where $\tilde{\mathsf{C}}^{-1}$ is some preconditioner matrix. This requires repeated application of $\mathsf{C}_{\rm fid}$ to the data, possible via (19). Convergence is expedited by sensible choice of $\tilde{\mathsf{C}}$; we require it to be easily inverted and close to $\mathsf{C}_{\rm fid}$, in the sense that $\tilde{\mathsf{C}}^{-1}\mathsf{C}_{\rm fid}=\mathsf{I}+\mathsf{R}$ where all the eigenvalues of R are less than unity. Here we consider the simple choice,

$$\tilde{\mathsf{C}}(\mathbf{r},\mathbf{r}') = \mathsf{H}_{\rm FKP}(\mathbf{r},\mathbf{r}') = n(\mathbf{r})[1+n(\mathbf{r})P_{\rm FKP}]\delta_D(\mathbf{r}-\mathbf{r}') \tag{21}$$

with $P_{\rm FKP}=10^4h^{-3}\,{\rm Mpc}^3$ (as in Ref. [6]), which is expected to reasonably well approximate $\mathsf{C}_{\rm fid}$. Given this form, we usually find the matrix inversion algorithm to converge (easily tested by computing $\mathbf{x}-\mathsf{C}_{\rm fid}\widetilde{\mathbf{y}}$) in around $N_{\rm it}\approx 50$ iterations. This is a computationally intensive step, requiring $\frac{1}{2}N_{\rm it}(1+\ell_{\rm max})(2+\ell_{\rm max})$ FFTs.

The action of $\mathsf{H}_{\rm FKP}$ on an arbitrary map x is required both for the CGD step of the ML estimator and the unwindowed-FKP approach. Because of the Dirac function, this is straightforward,

$$\mathsf{H}_{\rm FKP}^{-1}[x](\mathbf{r}) = \frac{x(\mathbf{r})}{n(\mathbf{r})[1+n(\mathbf{r})P_{\rm FKP}]}, \tag{22}$$

requiring a simple division in configuration space.[14]

C. Bias term and Fisher matrix

Using the above definitions, we can apply the weighting matrix $\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}$ to the data vector to find $\hat{q}_\alpha$, first computing $\mathbf{e}\equiv\mathsf{H}^{-1}\mathbf{d}$ and $\mathsf{C}_{,\alpha}\mathbf{e}$, then taking the trace $\mathrm{Tr}[\mathbf{e}^T\mathsf{C}_{,\alpha}\mathbf{e}]$—simply a summed product of two density fields. Below, we consider how to compute the bias and Fisher terms, which proceed slightly differently for the two choices of weight matrix.

1. ML estimators

The ML bias and Fisher terms are given by $\bar{q}_\alpha=\frac{1}{2}\mathrm{Tr}[\mathsf{C}_{\rm fid}^{-1}\mathsf{C}_{,\alpha}]$ and $F_{\alpha\beta}=\frac{1}{2}\mathrm{Tr}[\mathsf{C}_{\rm fid}^{-1}\mathsf{C}_{,\alpha}\mathsf{C}_{\rm fid}^{-1}\mathsf{C}_{,\beta}]$. Since the required matrices are dense and we do not have access to their full form, we compute these terms via Monte Carlo methods, as in Refs. [14,26], first writing

$$\begin{aligned} 2\bar{q}_\alpha &= \langle\mathbf{m}^T\mathsf{C}_{\rm fid}^{-1}\mathsf{C}_{,\alpha}\mathsf{C}_{\rm fid}^{-1}\mathbf{m}\rangle,\\ 2F_{\alpha\beta} &= \langle\mathbf{m}^T\mathsf{C}_{\rm fid}^{-1}\mathsf{C}_{,\alpha}\mathsf{C}_{\rm fid}^{-1}\mathsf{C}_{,\beta}\mathsf{C}_{\rm fid}^{-1}\mathbf{m}\rangle, \end{aligned} \tag{23}$$

where the random fields $\mathbf{m}$ satisfy $\langle\mathbf{m}\mathbf{m}^T\rangle=\mathsf{C}_{\rm fid}$.[15] In practice, we can use N-body simulations for these; in particular, those used to define the fiducial band powers $\mathbf{p}^{\rm fid}$, which satisfy the constraint by definition. Computing these traces via Monte Carlo averaging increases the estimator variance by a factor $\sqrt{1+1/N_{\rm mc}}$ when $N_{\rm mc}$ simulations are used; i.e., only ∼0.5% for 100 simulations. For both $\bar{q}_\alpha$ and $F_{\alpha\beta}$, bias will only arise if the equation $\langle\mathbf{m}\mathbf{m}^T\rangle=\mathsf{C}(\mathbf{p}^{\rm fid})$ is not satisfied; i.e., if our fiducial model is not accurate. While we technically require $\mathbf{p}^{\rm fid}=\mathbf{p}^{\rm true}$ for optimality, the increase in output parameter variances resulting from an incorrect N-body cosmology are expected to be small. In practice, we compute the contributions via the following algorithm.

1. For each simulation $\mathbf{m}$:

[14]Note that one requires $n(\mathbf{r})>0$ for the inverse map $\mathsf{H}_{\rm FKP}^{-1}[x]$ to be well defined. For an infinitely large random catalog, violation of this condition occurs outside the survey mask but may be ignored since $x(\mathbf{r})$ will always be zero in these regions. For a finite number of randoms, it is possible for pixels allowed by the survey grid to contain no random particles. To avoid infinities due to the inversion (which will be present also in the full $\mathsf{C}_{\rm fid}^{-1}[x]$ map), we replace such pixels by the average of their 26 nearest neighbors. Reducing the size of the random catalog from 50× to 10× shows this to be of little importance however, even with the finest choice of cell size.

[15]In fact, we can use any invertible matrix A to make the Fisher matrix separable, writing $2F_{\alpha\beta}=\langle\mathbf{a}^T\mathsf{C}_{\rm fid}^{-1}\mathsf{C}_{,\alpha}\mathsf{C}_{\rm fid}^{-1}\mathsf{C}_{,\beta}\mathsf{A}^{-1}\mathbf{a}\rangle$ for $\mathsf{A}=\langle\mathbf{a}\mathbf{a}^T\rangle$. We do not apply this here however, given that we use Monte Carlo simulations already to define the bias term.

(i) Compute $\mathsf{C}_{\rm fid}^{-1}\mathbf{m}$ via preconditioned CGD using (19).
(ii) For each band power α:
  (a) Compute $\mathbf{y}_\alpha = \mathsf{C}_{,\alpha}[\mathsf{C}_{\rm fid}^{-1}\mathbf{m}]$ via (20).
  (b) Accumulate the bias contribution $(\mathsf{C}_{\rm fid}^{-1}\mathbf{m})^T\mathbf{y}_\alpha$.
  (c) Compute $\mathsf{C}_{\rm fid}^{-1}[\mathbf{y}_\alpha]$ via CGD.
  (d) Accumulate the Fisher matrix contribution $\frac{1}{2}\mathbf{y}_\beta^T\mathsf{C}_{\rm fid}^{-1}[\mathbf{y}_\alpha]$ from each choice of β.

Note that we recompute $\mathbf{y}_\beta$ for each choice of α rather than first storing all $N_{\rm bins}$ vectors. This is done to avoid excessive memory requirements but slows down the estimator. In total, we require $N_{\rm mc}(N_{\rm bins}+1)$ matrix inversions to perform the above algorithm and must apply either the $\mathsf{C}_{\rm fid}$ or $\mathsf{C}_{,\alpha}$ operator $N_{\rm mc}[N_{\rm it}+N_{\rm bins}(N_{\rm it}+N_{\rm bins})]$ times, where $N_{\rm it}$ is the number of CGD steps. A faster approach and one advocated by Ref. [14] would be to instead use an approximate form of the Fisher matrix, such as that obtained using the preconditioner matrix $\tilde{\mathsf{C}}^{-1}$ instead of $\mathsf{C}_{\rm fid}^{-1}$. If the estimator is iterated until convergence (updating $\mathbf{p}^{\rm fid}$ at each step), the final estimate does not depend on the exact form of $F_{\alpha\beta}$, justifying this approach. In our context, it is expensive to iterate the algorithm, since one would require a new set of simulations for each $\mathbf{p}^{\rm fid}$ update, and these are expensive to compute, since we require them to reproduce the survey geometry and light cone structure. When we do not recalibrate the simulation cosmology, as here, the full Fisher matrix is required. An additional approach would be set $F_{\alpha\beta}$ equal to the covariance of the $\mathbf{m}^T\mathsf{C}_{\rm fid}^{-1}\mathsf{C}_{,\alpha}\mathsf{C}_{\rm fid}^{-1}\mathbf{m}$ coefficients (which are already used to define the bias vector), noting that this relation is exact for a fully optimal estimator (cf. Appendix A). This is also unfeasible however, since we require a large number of simulations (i.e., $N_{\rm mc} \gg N_{\rm bins}$) to ensure that the Fisher matrix can be easily inverted. Furthermore, this will include additional higher-order contributions from the non-Gaussian four-point function, i.e., $\langle d_i d_j d_k d_l\rangle$, which lead to a nontrivial bias.

2. FKP estimators

To compute the bias and Fisher matrices for the quadratic estimator using the FKP weighting (16), a similar procedure is possible. Analogous to the ML case, we write

$$\bar{q}_\alpha^{\rm QE} = \frac{1}{2}{\rm Tr}[\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}\mathsf{C}_{\rm fid}] = \frac{1}{2}\langle\mathbf{m}^T\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}\mathbf{m}\rangle,$$
$$F_{\alpha\beta}^{\rm QE} = \frac{1}{2}{\rm Tr}[\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}\mathsf{C}_{,\beta}] = \frac{1}{2}\langle\tilde{\mathbf{m}}^T\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}\mathsf{C}_{,\beta}\mathsf{H}^{-1}\tilde{\mathbf{m}}\rangle, \quad (24)$$

where the random fields $\mathbf{m}$ and $\tilde{\mathbf{m}}$ satisfy $\langle\mathbf{m}\mathbf{m}^T\rangle = \mathsf{C}_{\rm fid}$, $\langle\tilde{\mathbf{m}}\tilde{\mathbf{m}}^T\rangle = \mathsf{H}$. As above, we can draw $\mathbf{m}$ from the set of simulations at fiducial cosmology (thus, $\bar{q}_\alpha$ is simply the average of $\hat{q}_\alpha$ over this set of simulations), but the Monte Carlo simulations used to define the FKP Fisher matrix require a different definition, since they must have covariance H rather than $\mathsf{C}_{\rm fid}$. These can be obtained from the set of random particle positions; we split the random catalog into $\approx 1/\alpha$ disjoint pieces (where α is the data-to-random ratio of $\mathbf{d}$) and compute $\tilde{m}_i$ as the difference between each of these and the remainder of the randoms, just as for $\mathbf{d}$. Each density field is then rescaled

$$\tilde{m}(\mathbf{r}) \to \tilde{m}(\mathbf{r}) \times \sqrt{\frac{1+n(\mathbf{r})P_{\rm FKP}}{1+\alpha}}, \quad (25)$$

such that $\langle\tilde{m}\tilde{m}^T\rangle = \mathsf{H}_{\rm FKP}$. Computation of the bias and Fisher matrix then proceeds as above. Because of its diagonal form, $\mathsf{H}_{\rm FKP}^{-1}$ can be applied much faster than $\mathsf{C}_{\rm fid}^{-1}$; however, the process still requires $N_{\rm mc}N_{\rm bins}^2$ applications of $\mathsf{C}_{,\alpha}$ for the Fisher matrix, and $N_{\rm mc}N_{\rm bins}$ for the bias (unless intermediate quantities are stored); thus, the algorithm is not significantly faster unless $N_{\rm bins} \ll N_{\rm it} \sim 50$. Note also that *two* sets of simulations are required in this case, one each for the bias and Fisher terms.

D. Fiducial spectrum

The final component of quadratic estimators (3) and (9) is the set of fiducial band powers $\mathbf{p}^{\rm fid} \equiv \{P_\ell^a\}$, which are related to the continuous power-spectrum model $P(\mathbf{k})$ used to define the fiducial covariance $\mathsf{C}_{\rm fid}$. We consider two possibilities for their computation: (1) Use the smooth fiducial model evaluated at the bin centers, or (2) average the fiducial model over the true k bins, incorporating the full complexities of binning and pixelization. The latter case proceeds by first computing $P(\mathbf{k})$ on the discrete 3D $\mathbf{k}$ grid, then binning it in the same manner as the estimators (20), leading to

$$p_\alpha^{\rm fid,discrete} = \frac{4\pi}{\int_{\mathbf{k}}\Theta_a(\mathbf{k})}\left[\int_{\mathbf{k}}\sum_L\frac{4\pi}{2L+1}P_L(\mathbf{k})Y_{\ell m}^*(\hat{\mathbf{k}})Y_{LM}(\hat{\mathbf{k}}) \times \int\frac{d\mathbf{x}}{V}Y_{\ell m}(\hat{\mathbf{x}})Y_{LM}^*(\hat{\mathbf{x}})\right], \quad (26)$$

where the denominator is just the number of modes in the $|\mathbf{k}|$ bin, and the integrals are sums over all pixels (which have total real-space volume V). This is simply derived from the usual Yamamoto power-spectrum estimator, and, in the limit of infinitely fine pixels and bins, $p_\alpha^{\rm fid,discrete} = P_\ell^a$, as expected.

In this work, we instead adopt the former option, evaluating the theory model at the bin centers, i.e., $P_\ell^a = P_\ell(k_a)$ where k_a lies at the center of bin a. This is of twofold utility: First, it ensures that the fiducial spectra match the theory model used in the later parameter inference (which does not conventionally include bin integration). Second, it removes the leading-order effects of pixelization and gridding in our statistics, thus reducing

bias. This occurs since our estimator is fundamentally the fiducial spectra plus the difference of $\hat{q}_\alpha$ computed in the data and simulations; any effects that are common to both simulations and data will cancel at lowest order. In practice, this allows us to use lower resolution grids to compute the spectra than possible with conventional approaches. Furthermore, using such a difference estimator (with a Monte-Carlo-calibrated bias term) can additionally reduce difficult-to-model nonlinear effects appearing at low k, such as the fingers-of-God effect and non-Poissonian shot noise, provided they are present both in simulations and data.

One further comment on the fiducial spectrum is required. The FKP-based quadratic estimator discussed above relies on both a fiducial model for $P(\mathbf{k})$ both to define $\mathbf{p}^{\rm fid}$ and to compute the bias (via a set of simulated realizations of this cosmology). Given that the standard FKP estimator for the *windowed* power-spectrum multipoles (e.g., [7]) does not require a fiducial model, one may ask whether it is strictly necessary for our unwindowed estimators. As expected, the dependence on $\mathbf{p}^{\rm fid}$ may be completely removed by rewriting (3) in the form

$$\hat{p}_\alpha^{\rm FKP} \to \sum_\beta F_{\alpha\beta}^{-1,{\rm QE}} {\rm Tr}[(\mathbf{d}\mathbf{d}^T - \mathsf{N})\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}], \quad (27)$$

which involves only a redefinition of the bias (and is the form given in Ref. [23]). Here, the bias term depends only on the noise covariance N (assuming that the signal covariance is fully characterized by the set of all linear band powers), which, for FKP weights, could be estimated using subsets of the random catalog, just as for $F_{\alpha\beta}^{\rm QE}$ above, or via simple Fourier transforms of $n(\mathbf{r})$, as detailed in Appendix C. Equation (27) applies also to the ML estimator, though a fiducial model is still required to compute $\mathsf{C}_{\rm fid}^{-1}$ in that case. In this work, we opt to retain the fiducial model, since it has several useful properties; most notably, the fact that the difference estimator [in the form of (3)] is able to cancel the leading-order effects of pixelization, non-Poissonian shot noise, and unsampled bins.[16]

IV. ESTIMATING SUBSPACE COEFFICIENTS

The above sections have presented an extensive discussion concerning the application of quadratic estimators to measuring the unwindowed galaxy power spectrum from spectroscopic surveys. In general, this spectrum is high dimensional (particularly in redshift space), requiring significant computational power to compute the statistic and its sample covariance. For this reason, it is often useful to *compress* the statistic, and a number of methods are available by which to perform this (e.g., [11,27,35–37]). Here, we focus on the subspace compression introduced in Ref. [27] (which gives the optimal linear decomposition for a specific cosmological analysis), but note that the approach is applicable to any compression scheme in which the power spectrum can be written as a linear combination of basis functions.

A. The compressed subspace

We begin by recapitulating the approach of Ref. [27], in somewhat modified notation. The former work first computed the window-convolved power spectra via standard estimators, then projected these measurements into a subspace of reduced dimension, utilizing basis vectors computed from a singular value decomposition (SVD) in the space of allowed theory models. Here, we work only with unbinned and unwindowed spectra and compute the subspace coefficients directly, i.e., without first computing the band powers. This leads to a significant reduction in computational time.

The continuous-space decomposition is defined by first rotating an arbitrary spectrum $P(\mathbf{k})$ into a noise-weighted function $X(\mathbf{k})$, then decomposing into a set of $N_{\rm SV}$ basis functions $\{V^\alpha(\mathbf{k})\}$ with coefficients $\{c_\alpha\}$:

$$X(\mathbf{k}) = \int_{\mathbf{k}'} C_{\rm fid}^{-1/2}(\mathbf{k},\mathbf{k}')[P(\mathbf{k}') - \bar{P}(\mathbf{k}')],$$
$$X(\mathbf{k}) \approx \sum_{\alpha=1}^{N_{\rm SV}} c_\alpha V^\alpha(\mathbf{k}). \quad (28)$$

In (28), $C_{\rm fid}$ is a fiducial covariance matrix of the power spectrum,[17] the square root indicates a Cholesky decomposition, and $\bar{P}(\mathbf{k})$ is a mean spectrum ensuring that the average over all samples of $X(\mathbf{k})$ has mean zero. This becomes exact in the limit of $N_{\rm SV} \to \infty$. In Ref. [27], the basis vectors $\{V^\alpha\}$ are chosen by drawing a "template bank" of noiseless power spectra from the theory model with some priors on cosmological and nuisance parameters, then performing a SVD on this set to define the basis vectors, with the mean of all templates additionally setting $\bar{P}$.[18] This template bank can also be used to define $N_{\rm SV}$ by enforcing some limit on the χ^2 error incurred by the decomposition averaged across the prior domain. In practice, the basis vectors are computed in a finite number of

[16]To see this, consider the scenario in which our model for N is wrong by an additive factor. If using the form of (27), this will lead to a bias in $\hat{p}_\alpha$, but not from estimator (3) if it is contained within both C_D and $\mathsf{C}_{\rm fid}$; i.e., if it is an effect present in the simulations. Essentially, we will not be biased by modeling inaccuracies in the fiducial covariance (as long as the derivatives $\mathsf{C}_{,\alpha}$ are correct, though these may lead to a slight loss of optimality).

[17]Note that we use C to indicate covariances of the pixelized data, and C to refer to those of the power spectrum.

[18]We note that choosing the basis vectors requires significantly more thought for the bispectrum, since an analogous SVD will not generate separable templates, which are required for fast computation.

bins (and for the first few Legendre multipoles); these are easily converted to continuous form by way of interpolation. To define the decomposition, we must also set the fiducial power-spectrum covariance $C_{\rm fid}$. For an efficient and accurate decomposition, this should be a smooth model and similar to the true covariance $C_{\rm true}(\mathbf{k},\mathbf{k}') = \mathrm{cov}(P(\mathbf{k}), P(\mathbf{k}'))$. In practice, a Gaussian model is usually a good approximation [27].

Inverting (28) allows the (unwindowed) power spectrum to be written as a linear sum of shapes

$$P(\mathbf{k}) \approx \sum_{\alpha=1}^{N_{\rm SV}} c_\alpha \int_{\mathbf{k}'} C_{\rm fid}^{1/2}(\mathbf{k},\mathbf{k}') V^\alpha(\mathbf{k}') + \bar{P}(\mathbf{k}) \equiv \sum_{\alpha=1}^{N_{\rm SV}} c_\alpha \omega^\alpha(\mathbf{k}) + \bar{P}(\mathbf{k}) \tag{29}$$

defining the weighting functions $\omega^\alpha(\mathbf{k})$ in terms of the subspace basis vectors V^α and the fiducial power-spectrum covariance $C_{\rm fid}$. The basis vectors are orthonormal, satisfying $\int_{\mathbf{k}} V^\alpha(\mathbf{k}) V^\beta(\mathbf{k}) = \delta^K_{\alpha\beta}$; this leads to the direct estimator for c_α,

$$c_\alpha = \int_{\mathbf{k}}\int_{\mathbf{k}'} [P(\mathbf{k}) - \bar{P}(\mathbf{k})] C_{\rm fid}^{-1/2}(\mathbf{k},\mathbf{k}') V^\alpha(\mathbf{k}'), = \int_{\mathbf{k}}\int_{\mathbf{k}'} [P(\mathbf{k}) - \bar{P}(\mathbf{k})] C_{\rm fid}^{-1}(\mathbf{k},\mathbf{k}') \omega^\alpha(\mathbf{k}'), \tag{30}$$

which is a convolution of $P(\mathbf{k})$ with the shape functions ω^α. This (in discrete form and applied to the windowed power-spectrum multipoles) is the manner used to estimate c_α from theory and data in the former work. Since we only ever consider the difference between measured and theoretical c_α coefficients, we note that $\bar{P}(\mathbf{k})$ cancels in practical contexts.

B. Quadratic estimators for $\{c_\alpha\}$

In this work, we estimate $\mathbf{c} \equiv \{c_\alpha\}$ directly using the quadratic estimators, rather than via (30). To do so, we first express (29) in terms of multipoles defining the new weightings $\omega^\alpha_\ell(k)$,

$$P(\mathbf{k};\mathbf{r}') \approx \sum_\ell \left[\sum_{\alpha=1}^{N_{\rm SV}} c_\alpha \omega^\alpha_\ell(k) + \bar{P}_\ell(k)\right] L_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{r}}') \tag{31}$$

allowing for dependence on the line-of-sight $\hat{\mathbf{r}}'$, as in the Yamamoto estimator. The $\omega^\alpha_\ell(k)$ functions are, in principle, arbitrary, though we will use them as the interpolated (discrete) SVD basis vectors; i.e., $\omega^\alpha_\ell(k_i) = \omega^\alpha_{\ell,i}$ where i indexes the k bin used to form the template bank. By full analogy with Sec. II, the quadratic estimator for $\mathbf{c}$ is given by (9) or (3), but replacing the $\mathsf{C}_{,\alpha} \equiv \partial\mathsf{C}/\partial P_\alpha$ with $\partial\mathsf{C}/\partial c_\alpha$. This is a straightforward redefinition of (15)

$$\mathsf{C}_{,\alpha}(\mathbf{r},\mathbf{r}') \to \frac{\partial \mathsf{C}}{\partial c_\alpha}(\mathbf{r},\mathbf{r}') = n(\mathbf{r})n(\mathbf{r}') \int_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} \sum_\ell \omega^\alpha_\ell(k) L_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{r}}') \tag{32}$$

inserting ω^α_ℓ in place of the filter functions $\Theta^a(k)$ (which were simply Fourier-space shells). Note that (32) retains the sum over ℓ, since the subspace basis vectors contain contributions from all multipoles used to compute and create the template bank. Additionally, the wave number support of $\omega^\alpha_\ell(k)$ can be set by truncating the template bank spectra at some $k_{\rm min}$ and $k_{\rm max}$. The application of (32) on pixelized fields is analogous to (20), and we note that there is no change to the pixel weighting $\mathsf{H}_{\rm FKP}$ or $\mathsf{C}_{\rm fid}$ (with the latter already depending on the full power spectrum in the fiducial cosmology).

To use the quadratic estimator, we additionally require the fiducial coefficients $c^{\rm fid}_\alpha$. From (30), these are defined by

$$c^{\rm fid}_\alpha = \int \frac{d\mathbf{r}'}{V} \int_{\mathbf{k}} \tilde{\omega}^\alpha(\mathbf{k};\mathbf{r}')[P(\mathbf{k};\mathbf{r}') - \bar{P}(\mathbf{k};\mathbf{r}')] \tag{33}$$

averaging over the line of sight. This uses the inverse weights defined as in (30),

$$\tilde{\omega}^\alpha(\mathbf{k};\mathbf{r}') = \int_{\mathbf{k}'} C^{-1}_{\rm fid}(\mathbf{k},\mathbf{k}') \omega^\alpha(\mathbf{k}';\mathbf{r}') = \sum_\ell (2\ell+1) \tilde{\omega}^\alpha_\ell(k) L_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{r}}') \tag{34}$$

with a multipole decomposition given in the second line. As in Sec. III D, there are two possibilities for computing c_α, either allowing for or removing discreteness effects. In the first case, we estimate the $\tilde{\omega}^\alpha_\ell$ functions from the discretely sampled SVD vectors via $\tilde{\omega}^\alpha_\ell(k_i) = (2\pi^2)/(k_i^2 \delta k_i) \times [C^{-1}_{\rm fid}\omega^\alpha]_{\ell,i}$, with the prefactor necessary to ensure correct normalization. This gives the fiducial coefficients analogous to (26)

$$c^{\rm fid,discrete}_\alpha = 4\pi \left[\int_{\mathbf{k}} \sum_{\ell,L} \frac{4\pi}{2L+1} [P_L(k) - \bar{P}_L(k)] \tilde{\omega}^\alpha_\ell(k) Y^*_{\ell m}(\hat{\mathbf{k}}) Y_{LM}(\hat{\mathbf{k}}) \int \frac{d\mathbf{r}'}{V} Y_{\ell m}(\hat{\mathbf{r}}') Y^*_{LM}(\hat{\mathbf{r}}')\right], \tag{35}$$

where the integral is evaluated as a sum over all points on the real and Fourier-space grids.

The alternative approach, which will be used here, is to perform the angular integrals in (33) analytically, leading to

$$c_\alpha^{\rm fid} \to \int \frac{k^2 dk}{2\pi^2} \tilde{\omega}^\alpha(k)[P_\ell(k) - \bar{P}_\ell(k)]. \tag{36}$$

Given the SVD basis vectors V^α in discrete form, this becomes a matrix multiplication of the fiducial spectrum $P_\ell(k)$ (evaluated at the same wave numbers used to define the SVD) and $\tilde{\omega}_i = \sum_j C_{{\rm fid},ij}^{-1/2} V_j^\alpha$. This approach eliminates leading-order effects of gridding artifacts and the finite-k sampling used to define the basis vectors.

Following the above steps, we obtain a practical estimator for computing the unwindowed subspace coefficients directly from the pixelized data without first computing the band powers. In this context, the subspace decomposition has dual utility: First, the associated dimensionality reduction (i.e., $N_{\rm SV} \ll N_{\rm bins}$) reduces the number of mocks required to compute sample covariance matrices (as in [27]). Second, it significantly expedites the estimator itself (both for the ML and FKP pixel weights). This occurs since the number of FFTs required to compute the statistic scales as $N_{\rm mc} N_{\rm bins}(N_{\rm it} + N_{\rm bins})$ (Sec. III C), with $N_{\rm bins} \to N_{\rm SV}$ in the compressed formalism.

This approach carries an important subtlety. As previously noted, the quadratic estimator is fully unbiased only if we simultaneously estimate all possible $\mathbf{p}$, or equivalently, coefficients $\mathbf{c}$. Using only a subset incurs a small error unless the unsampled coefficients are (a) independent of those sampled or (b) the same in the data and the fiducial model.[19] When estimating band powers this is of little concern, since most unmeasured bins contribute little power in practice, given the lack of a window function. For the coefficients, however, restricting to small $N_{\rm SV}$ can lead to a bias and increased variance in the estimator if there remains significant information outside those coefficients. (Recall that the subspace basis vectors are chosen to maximize information content in the final cosmological analysis, rather than signal-to-noise in the statistic.) In practice, we find this to be the case when using the FKP quadratic estimator for the subspace coefficients (which is less efficient than the ML approach at removing window-function effects) and ameliorate it by measuring twice the number of coefficients, with the latter half discarded when the coefficients are later analyzed via a Markov chain Monte Carlo (MCMC). One can straightforwardly assess whether this effect is present by using only a subset of the coefficients in the estimated $\hat{q}_\alpha$ and $F_{\alpha\beta}$ quantities (3) and checking for variation in $\hat{p}_\alpha$.

[19]Schematically, this can be seen from writing (A3) as $\mathbb{E}[\hat{p}_\alpha] - p_\alpha^{\rm fid} = \sum_{\beta,\gamma} F_{\alpha\beta}^{-1} \mathcal{F}_{\beta\gamma}(p_\gamma^{\rm true} - p_\gamma^{\rm fid})$, where F is a submatrix of the full (infinite-dimensional) Fisher matrix $\mathcal{F}$, and β runs only over the estimated coefficients. If the unsampled bins are uncorrelated, then $\sum_\beta F_{\alpha\beta}^{-1}\mathcal{F}_{\beta\gamma} = \delta_{\alpha\beta}^K$, giving $\hat{p}_\alpha = p_\alpha^{\rm true}$, and similarly if $p_\gamma^{\rm true} = p_\gamma^{\rm fid}$ for those coefficients.

V. DEMONSTRATION ON BOSS DR12

As a practical demonstration of the quadratic estimators discussed herein, we estimate the power-spectrum multipoles and compressed subspace coefficients of a real cosmological dataset. For this, we use the 12th data release (DR12) [38] of the Baryon Oscillation Spectroscopic Survey (BOSS) part of SDSS-III [28,39], specializing to the patch with the largest number density (and volume): the north Galactic cap in the redshift range $0.2 < z < 0.5$, with total volume $1.46h^{-3}$ Gpc^3 and mean redshift $z = 0.38$.

A set of MultiDark-PATCHY (hereafter "PATCHY") simulations[20] [40,41] are used to define the fiducial cosmology, which, by design, have power spectra close to those of BOSS and share its selection function. For each simulation, we compute the windowed power-spectrum multipoles using the estimator of Ref. [7] within NBODYKIT [33] (using the same gridding parameters as for the public BOSS data release), then fit the mean of 1000 simulations to the one-loop EFT model [32] using the CLASS-PT implementation [42]. For the fit, we use $k \in [0.01, 0.25]h$ Mpc^{-1} and $\ell \in \{0, 2\}$, fixing the cosmological parameters to their true values $(\{\Omega_m = 0.307115, \Omega_b = 0.048206, \sigma_8 = 0.8288, n_s = 0.9611, h = 0.6777, A_s = 2.1467 \times 10^{-9}\})$ and varying only the nuisance parameters, encapsulating linear and nonlinear galaxy bias, stochasticity, and the EFT counterterms. Finally, the fiducial power spectrum is set equal to the best-fit theory model computed using the public window-function multipoles and subtracting Poisson shot noise; the latter is necessary to ensure that we do not double count stochasticity in both the signal and noise covariance of the data. We require also a smooth model of the power spectrum for $k > k_{\rm max}$; this is set to linear theory rescaled to avoid discontinuities at $k_{\rm max}$. The exact choice is of only minor importance, since these modes are not used in the later analysis.

A. Estimating unwindowed spectra

Using the above, we may apply the quadratic estimators discussed in Secs. II and III to measure the band powers of both the data and simulations via (3). We will assume the binning $k_{\rm min} = 0h$ Mpc^{-1}, $k_{\rm max} = 0.3h$ Mpc^{-1}, $\Delta k = 0.005h$ Mpc^{-1}, and $\ell \in \{0, 2\}$, somewhat broader than the k range used for the subsequent MCMC analysis to avoid the aforementioned bias from correlated unmeasured bins. For both ML and FKP pixel weights, we require a set of galaxy positions and random particles; for the latter, we use the publicly available catalogs[21] generated for the PATCHY

[20]Publicly available at data.sdss.org/sas/dr12/boss/lss.

[21]Also at data.sdss.org/sas/dr12/boss/lss/.

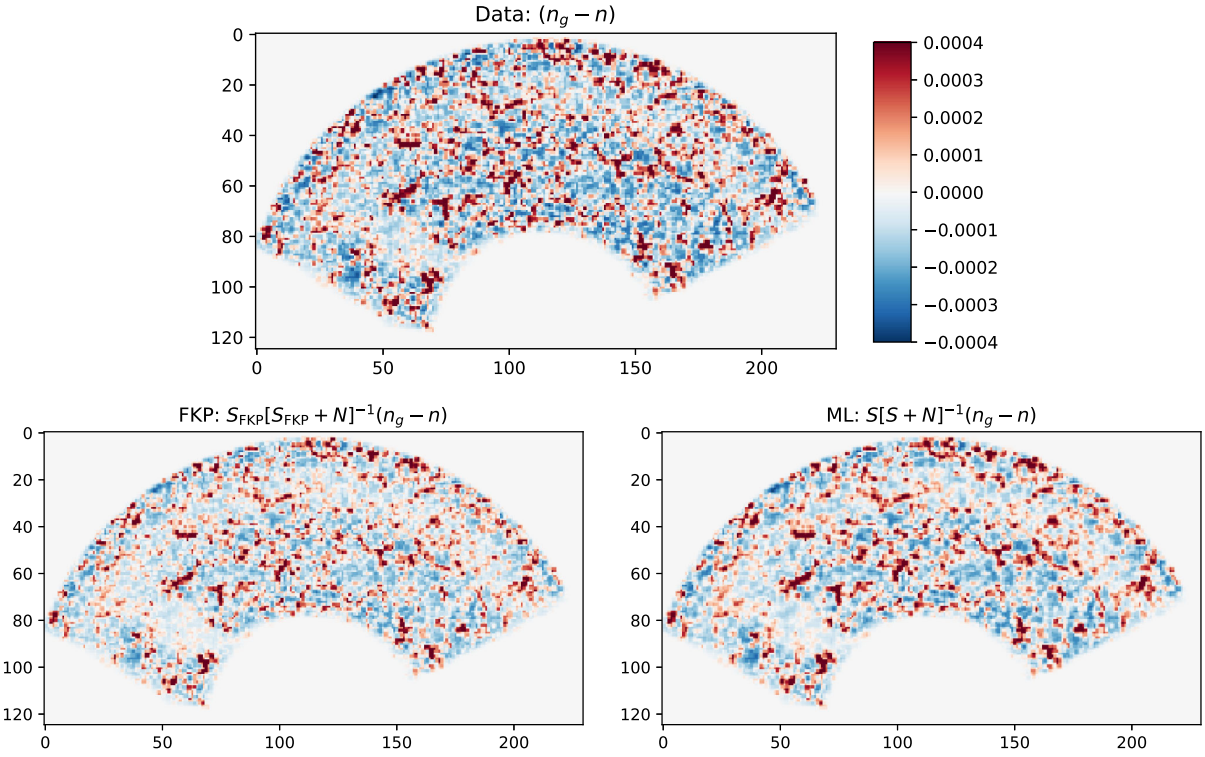

FIG. 1. Visualizations of the overdensity field for a single MultiDark-PATCHY mock, both unweighted and applying Wiener filters using the FKP (16) and ML (9) pixel covariances discussed in this work. Each image depicts a slice of radial width $\Delta z \approx 100h^{-1}$ Mpc, with horizontal and vertical axes labeling pixels (in the transverse direction) with grid size $\approx 11h^{-1}$ Mpc per pixel. We apply the Wiener filters $\mathsf{S}[\mathsf{S}+\mathsf{N}]^{-1}$ for signal and noise covariances S and N rather than the $[\mathsf{S}+\mathsf{N}]^{-1}$ weights used for power-spectrum estimation to facilitate easier interpretation. The ML covariance is inverted using conjugate-gradient descent methods described in Sec. III A.

and BOSS samples which have 50× the average density of the galaxy samples.[22] We include the full particle weights of BOSS (including completeness, systematic, and fiber collision weights, but not FKP particle weights),[23] and translate data and randoms onto Cartesian grids using the fiducial cosmology $\{\Omega_m = 0.31, h = 0.676\}$ (as in BOSS analyses). Gridding is achieved according to triangular-shaped-cloud interpolation utilizing the same cuboidal grids as for the public $P(k)$ spectra, but with the number of pixels in each dimension optionally reduced by a factor $f_{\rm pix}$ of 2 or 3.[24] These have associated Nyquist frequencies $k_{\rm Nyq} \approx (0.6/f_{\rm pix})h$ Mpc^{-1}. We precompute the fiducial power-spectrum multipoles $P_\ell(\mathbf{k})$ and the spherical harmonics $Y_{\ell m}(\hat{\mathbf{k}})$ and $Y_{\ell m}(\hat{\mathbf{r}})$ on the grid to facilitate later computations.

With these pixelized fields, we compute the quadratic estimator terms $\mathbf{d}^T\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}\mathbf{d}$ and, for the ML case, $\mathbf{d}^T\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}\mathsf{C}_{,\beta}\mathsf{H}^{-1}\mathbf{d}$ as detailed in Sec. III. For the FKP case, the Fisher matrix is instead computed using subsets of the random catalog as in Sec. III C. This operation is performed both for the data and for 200 PATCHY mocks (and using 50 random partitions to define the FKP Fisher matrix); 50 of these are used to define the bias and Fisher matrix via Monte Carlo averaging,[25] allowing the estimated BOSS band powers $\hat{p}_\alpha$ to be computed via (3) or (9). The remainder are used to form a covariance matrix of the statistic. While this gives a good indication of the statistic correlation structure, it is insufficient for performing parameter inference, as one requires $N_{\rm mocks} \gg N_{\rm bins}$ to obtain a robust inverse covariance matrix estimate. For this reason, we run MCMC only on the compressed subspace vectors (discussed below). Using the ML pixel weights, each simulation requires ∼0.5 (1.5) h to analyze on a four-core machine for $k_{\rm Nyq} = 0.2\,(0.3)\ h\,{\rm Mpc}^{-1}$; for the FKP case, we require $\sim 1(0.5)$ min to compute the bias term in each simulation, though with an additional ∼0.5 (1.5) h for each

[22] See Appendix D for the effects of reducing this to 10×.

[23] When using the FKP pixel weights (i.e., $\mathsf{H} = \mathsf{H}_{\rm FKP}$), we additionally divide the FKP Fisher matrix by a factor $\int d\mathbf{r} n^2(\mathbf{r}) \times \alpha / \sum_i \bar{n}_i$, where $\bar{n}_i$ is the number density associated with the random catalog and i runs over all random particles. This corrects for differences in normalization of the gridded and windowed-FKP estimators when only a finite number of randoms are used.

[24] The dependence of the spectra on $f_{\rm pix}$ is discussed in Appendix D.

[25] As in Sec. III C, using 50 simulations leads to an ≈1% increase in the estimator variance relative to that of infinite mocks.

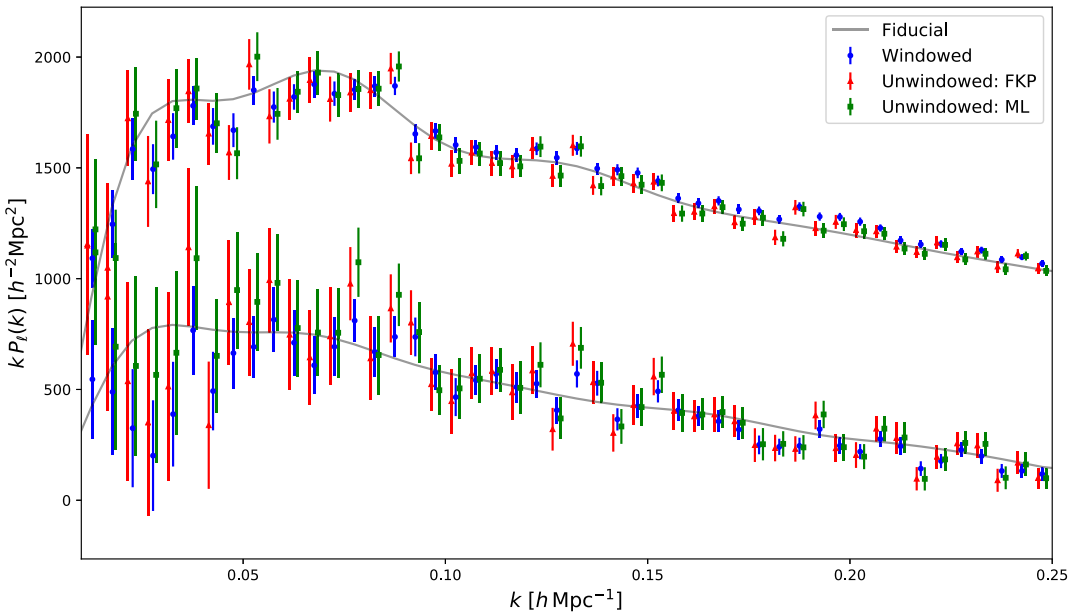

FIG. 2. Galaxy power spectra for the largest-volume chunk of the BOSS DR12 dataset computed both using the conventional algorithm available in NBODYKIT (blue) and the quadratic estimators discussed in this work using FKP (red) and ML (green) pixel weightings. Only the first set of spectra are convolved with the survey window function, though neither include Poissonian shot noise. Data are obtained by applying the formalism of Secs. II and III to the observed BOSS data in the NGC patch with $0.2 < z < 0.5$, using 150 PATCHY simulations to compute the variances. The quadratic estimators grid the data with a Nyquist frequency $k_{\rm Nyq} = 0.3h$ Mpc^{-1} (which is sufficient for this method) and use a fiducial power spectrum shown by the gray lines. The upper (lower) data points show the redshift-space monopole (quadrupole), and we add small lateral displacements for clarity. While the windowed estimates have reduced variance at high k, individual bins are highly correlated, as shown in Fig. 4.

Fisher matrix random catalog partition. As previously discussed, this is similar to the ML case since the number of bins is large.

Before discussing the estimated power-spectrum multipoles, it is instructive to consider the action of the pixel weights in the FKP and optimal formalism. Figure 1 shows a slice of the 3D overdensity field $d(\mathbf{r}) \equiv n_g(\mathbf{r}) - n(\mathbf{r})$, alongside its Wiener-filtered form $\mathsf{S}[\mathsf{S}+\mathsf{N}]^{-1}\mathbf{d}$, using the FKP and ML signal covariances S. Even though the raw data do not include any selection-based weights, the differences between the unfiltered and filtered maps is relatively small, though we note that the former contains more small-scale power. The differences between the FKP and ML weights are again slight, though the ML scheme appears to upweight long-wavelength modes. While this discussion is somewhat rudimentary, it suggests that the differences between the ML and FKP spectra may be small, except on the largest scales.

Figure 2 shows the band-power estimates from the two quadratic estimators alongside the standard (windowed) power-spectrum multipoles computed using NBODYKIT. The quadratic estimators are computed using a grid size corresponding to $k_{\rm Nyq} = 0.3h$ Mpc^{-1}; while this may seem insufficiently fine (indeed, we set $k_{\rm Nyq} = 0.6h$ Mpc^{-1} for the NBODYKIT estimates to avoid aliasing), using broad cells makes little difference to the band-power estimates due to the form of (3) as a difference estimator. This is elaborated upon in Appendix D, displaying different choices of $k_{\rm Nyq}$.

First, we note that, while the spectra from the quadratic estimators are not directly compatible with those from NBODYKIT due to the removal of window-function effects, they are heuristically similar, and notably, both show the same departures from the fiducial power-spectrum model. Especially at high k, it is clear that the windowed estimates have smaller variances than their unwindowed equivalents; this is not an indication that the quadratic estimators are incomplete, but rather due to the presence of significant correlations between bins induced by the window function. Indeed, the total signal-to-noise ratio of the two approaches is very similar. Comparing now the quadratic estimators with FKP and ML pixel weights, we find very similar results, both for the means and variances of each band power, indicating that the FKP form is close to the optimal Gaussian solution for this choice of bins. Further discussion of this may be found in Sec. V D.

Given that individual band powers will have a nontrivial correlation structure (even in the ML case), a more appropriate quantity to plot is the vector $\sum_\beta F_{\alpha\beta}^{-1/2}[\hat{q}_\beta - \bar{q}_\beta]$, which includes all contributions from the data, and, for the ML estimator, has unit covariance in the Gaussian limit [23]. This is plotted in Fig. 3 and allows for better comparison of the two approaches. Notably, the

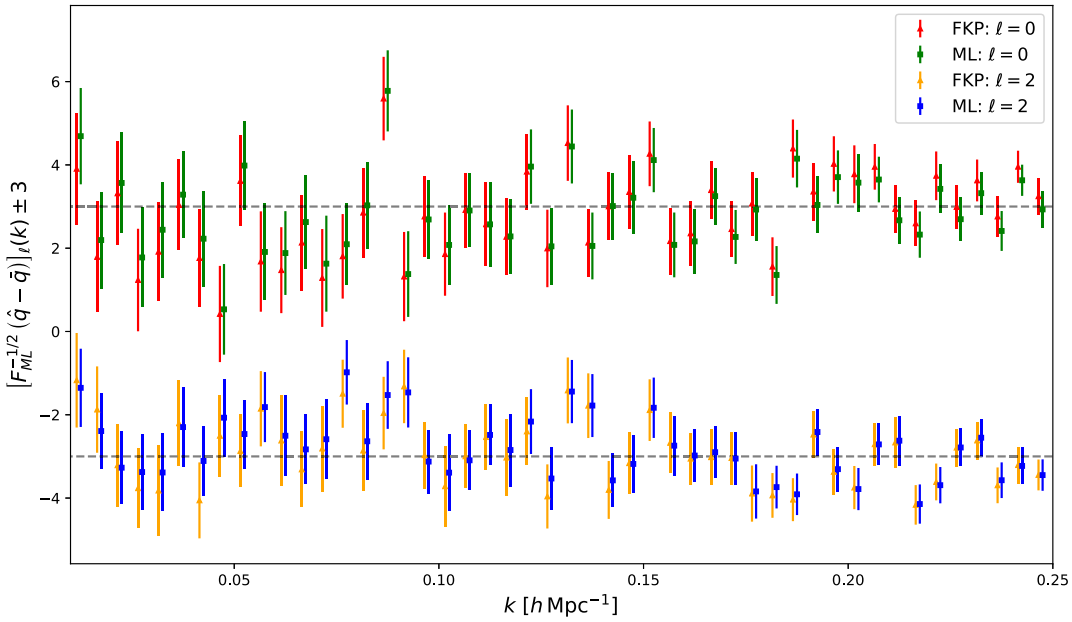

FIG. 3. Unwindowed BOSS power-spectrum estimates from the quadratic estimators rescaled to the convention of Ref. [23]. We plot the quantity $\hat{q}_\alpha - \bar{q}_\alpha$ (where $\hat{q}_\alpha$ is the part of the estimator containing the data, and $\bar{q}$ is its equivalent averaged over mocks) normalized by the Cholesky decomposition of the inverse ML Fisher matrix; this combination has close to unit covariance matrix for the ML sample. The top (bottom) lines show the result for the monopole (quadrupole), and we note consistent estimates and similar error bars in both cases, with only a slight improvement seen from the FKP estimators at low k.

variance of the ML estimates is close to unity for all the bins, implying that the estimator is near optimal. Once again, we find little difference between the FKP and ML approaches, except for a possible slight reduction in the error bars from the ML estimators on the largest scales (which would be useful, for example, for surveys focusing on primordial non-Gaussianity). A more meaningful comparison of the spectra is via their resulting cosmological parameter constraints; these will be discussed in Sec. V C.

Finally, Fig. 4 considers the correlation matrix of the above power spectra computed from the 150 PATCHY mock band-power estimates. The windowed spectra exhibit clear positive correlations for off-diagonal elements, particularly those in neighboring bins, but also pairs at larger separations on small scales. In contrast, the unwindowed quadratic estimator covariances are almost diagonal (and very similar for FKP and ML pixel weights), with just a slight hint of positive correlations at high k and a slight anticorrelation between adjacent bins. This is as expected; the window function smears the true power-spectrum covariance into nearby bins, which also accounts for the reduced variance seen at high k in Fig. 2.

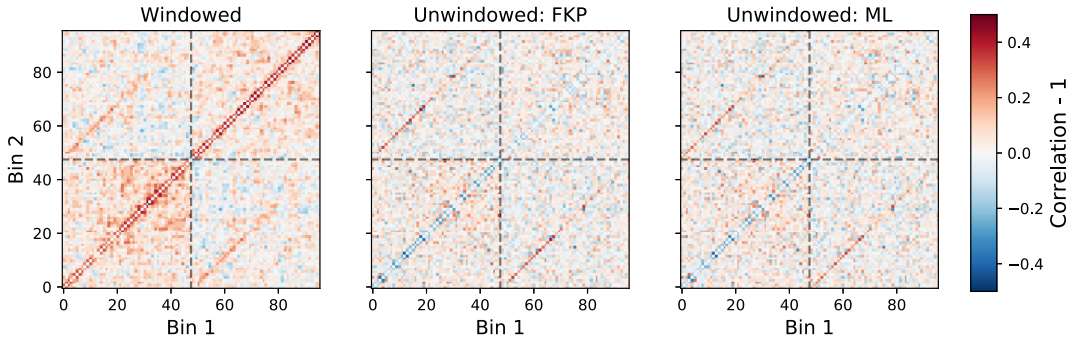

FIG. 4. Correlation matrices for the power spectra plotted in Fig. 2, defined as $\mathrm{corr}_{ab} = \mathrm{cov}_{ab}/\sqrt{\mathrm{cov}_{aa}\mathrm{cov}_{bb}}$ for binned covariance cov_{xy}. These are generated from 150 PATCHY mocks, and bins 0–47 (48–95) refer to the monopole (quadrupole) with increasing k in the range $[0.01, 0.25]h$ Mpc^{-1}. For clarity, we subtract the identity matrix to remove the leading diagonal. We note considerable correlation between neighboring bins and at high k for the windowed spectra which is significantly reduced using the quadratic estimators discussed in this work, though there is a slight anticorrelation of neighboring bins.

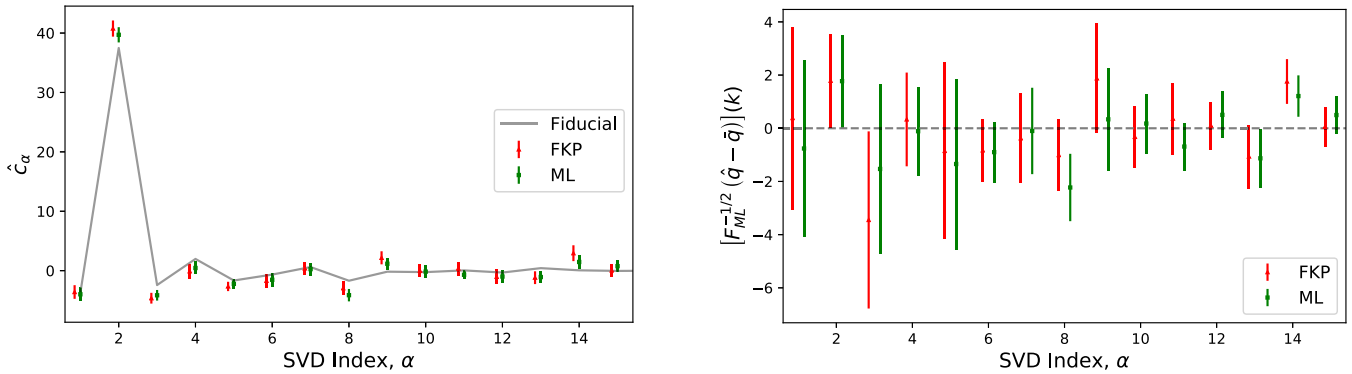

FIG. 5. Compressed coefficients of the BOSS DR12 power spectrum obtained using the quadratic estimators of Sec. IV with FKP (green) and ML (red) pixel weights. The left panel shows the full coefficients alongside the fiducial model in gray (analogous to Fig. 2), while the right employs the rescaling of Fig. 3 to reduce correlations between bins. For the left panel, we note that only the *difference* of fiducial and estimate $\hat{c}_\alpha$ is of interest; the offset cancels in any likelihood. Points show the estimates obtained from the observed BOSS density field, while the error bars give the covariances from 150 PATCHY mocks. A total of 30 coefficients are estimated, but only the first 15 (shown here) are used for the later parameter inference. Each represents the amplitude of a basis vector in the subspace described in Sec. V B. We find similar results from the FKP and ML estimators, as in previous figures.

B. Direct computation of subspace coefficients

Fast and accurate analyses of galaxy power spectra are facilitated by projecting the spectra into information-maximizing subspaces, as in Ref. [27]. This is possible via the algorithms discussed in Sec. IV and allows for robust parameter inferences using only $\mathcal{O}(100)$ mocks. This is far more computationally feasible, given the increase in computing power needed for estimating spectra via configuration-space quadratic estimators.

To compute the compressed spectra, we first require a set of basis vectors generated from a template bank of theory models given prior ranges on cosmological and nuisance parameters. Unlike Ref. [27], we do not need to include window functions and can use dense k-space sampling to ensure information is not lost in binning. We sample 10^4 template bank parameters from the following distributions (varying just $h, \omega_{\rm cdm}$, A_s, and nuisance parameters, for simplicity)

$$\begin{aligned} &h \sim \mathcal{U}(0.6, 0.74), \qquad \omega_{\rm cdm} \sim \mathcal{U}(0.08, 0.16), \\ &(A_s/A_{s,\rm Planck})^{1/2} \sim \mathcal{U}(0.6, 1.4), \\ &b_1 \sim \mathcal{U}(1.7, 2.3), \qquad b_2 \sim \mathcal{N}(0, 1^2) \qquad b_{\mathcal{G}_2} \sim \mathcal{N}(0, 1^2), \\ &c_{s,0} \sim \mathcal{N}(0, 30^2), \qquad c_{s,2} \sim \mathcal{N}(0, 30^2), \\ &P_{\rm shot} \sim \mathcal{N}(0, 5000^3), \qquad b_4 \sim \mathcal{N}(500, 500^3), \end{aligned} \tag{37}$$

where $\mathcal{U}$ and $\mathcal{N}$ represent uniform and normal distributions. The spectra are evaluated for $k \in [0.01, 0.25]h$ Mpc^{-1} with $\Delta k = 0.0005h$ Mpc^{-1} and $\ell \in \{0, 2\}$, choosing the k range to limit the impact of observational systematics and model inaccuracies. Furthermore, we set the fiducial covariance equal to a simple Gaussian model with true survey volume, and include a theoretical error covariance [43] of the form

$$C_{{\rm TE},\ell\ell'}(k, k') = E_\ell(k) E_{\ell'}(k') e^{-(k-k')^2/2\Delta k^2}, \tag{38}$$

where $E_\ell(k) = D^2(z) P_\ell^{\rm tree}(k)(k/0.45)^{3.3}$, similar to that used in Ref. [44]. Here the coherence length $\Delta k = 0.1h$ Mpc^{-1} ensures that we still retain baryon acoustic oscillation information from small scales but are robust to broadband modeling uncertainties. Given the template bank, the subspace is formed via a SVD, and we use only the first 30 basis vectors.[26] Following computation of the basis vectors (which we emphasize is fast, requiring only 10^4 model evaluations, each of which requires one CLASS evaluation and a negligibly fast evaluation of the CLASS-PT module), we compute the subspace coefficients directly, as in Sec. IV, utilizing $N_{\rm mc} = 50$ simulations for the Monte Carlo averages and $N_{\rm sim} = 150$ for the covariances. These numbers are more than sufficient, given that the error on the Fisher matrix scales as $\sqrt{1 + 1/N_{\rm mc}}$ and the error inflation from using too few mocks scales roughly as $(N_{\rm SV} - N_{\rm param})/N_{\rm sim}$ for an analysis measuring $N_{\rm param}$ parameters. The compressed analysis is significantly faster than that for the unwindowed spectra since fewer FFTs are required; it takes ∼7, 20, and 220 min per simulation on a four-core machine using $k_{\rm Nyq} = 0.2$, 0.3 and $0.6h$ Mpc^{-1}, respectively, with ML pixel weights (and a comparatively similar time with FKP weights, but only to compute the Fisher matrix from the random particle subsamples).

[26] As in Ref. [27], we choose $N_{\rm SV}$ by requiring that the basis projection incurs an error in χ^2 (averaged across the prior domain) below 0.1. This gives $N_{\rm SV} \geq 15$, which the former work showed to capture essentially all the cosmological information available. Here, we estimate twice the number of coefficients to minimize the estimator bias from unsampled modes (cf. Sec. IV B). Spectra obtained using $N_{\rm SV} = 15$ are shown in Appendix D.

The first 15 subspace coefficients are shown in Fig. 5, both plotted directly and rescaled in the manner of Fig. 3. The narrative is similar to that of the band powers; the coefficients exhibit some variation around the fiducial coefficients calculated in the PATCHY cosmology but are is consistent between the two choices of weightings, with similar error bars. Notably, the amplitudes for the coefficients are generally close to zero with variances close to unity; this is expected from the SVD decomposition, as discussed in Ref. [27]. The large amplitude of the second coefficient will relate to a significant difference between the mean of the basis vector template bank cosmology and that of the PATCHY (and true) cosmology; in general, each basis vector depends on several different parameters (both cosmological and nuisance) and their priors, making their direct interpretation difficult. We reemphasize that, since the data and theory model are compressed in the same way, any offsets in $\hat{c}_\alpha$ do not enter the likelihood. By changing the strategy by which the template samples are drawn, we can modify the mean $\hat{c}_\alpha$; any such change does not bias the inference however (though the compression requires the fewest basis vectors if the template samples are drawn from the parameter posteriors of the data).

C. Parameter inference without windows

We finally consider the performance of the quadratic estimators in their main application: cosmological parameter inference. For this purpose, we run a simple analysis varying the parameter set $\{h, \omega_{\rm cdm}, A_s\}$ alongside the nuisance parameters, using the likelihoods of [32][27] based on CLASS-PT [42]. These include the analytic marginalization procedures of Ref. [27] for parameters that enter the likelihood linearly (the residual shot noise $P_{\rm shot}$ and the counterterms c_0, c_2, b_4), with only six parameters (the above plus $\{b_1, b_2, b_{\mathcal{G}_2}\}$) being directly sampled. We focus on the analysis of subspace coefficients with both ML and FKP pixel weights (utilizing a grid with $k_{\rm Nyq} = 0.3h$ Mpc^{-1}), noting that the conclusions will apply also to the full unwindowed band powers.

Inference is performed via MCMC using the MONTEPYTHON sampler [45], comparing the observed subspace coefficients $\hat{c}_\alpha$ to those from the finely sampled one-loop power-spectrum model with $k \in [0.01, 0.25]h$ Mpc^{-1}, projected into the subspace via the aforementioned basis vectors. Since the basis vectors are set to zero outside this k range, we do not have contributions from poorly modeled k bins. For the covariance matrix, we use the sample covariance computed from 150 PATCHY mocks supplemented with the theoretical error covariance $C_{\rm TE}$ of (38) and do not apply any Hartlap or Sellentin-Heavens noise corrections [46,47] since $N_{\rm mocks} \gg N_{\rm bins}$. Broad Gaussian priors are applied to the nuisance parameters (barring b_1) with widths given in (37) but centered on the values in the fiducial spectra. We run three MCMC chains, one for each of the FKP- and ML-weighted quadratic estimators discussed in this work, and a third using the full windowed power-spectrum estimates, with a sample covariance drawn from 1000 PATCHY mocks. For the latter, both $C_{\rm TE}$ and the theory model (at each step in the chain) must be convolved with the window function. All chains are run until convergence (assumed when the Gelman-Rubin diagnostic satisfies $|R-1| < 0.05$), which requires ∼50 CPU h per inference.

The output parameter contours are displayed in Fig. 6. The results from the three analyses are statistically consistent for all parameters at $\lesssim 0.8\sigma$, which is as expected since all spectra are computed from the same dataset. Furthermore, we find consistent contour widths when analyzing both power spectra and compressed coefficients; while Ref. [27] demonstrated this for the windowed subspace coefficients, this shows that the same conclusion applies for the directly sampled unwindowed coefficients. The main conclusion from this figure is as follows: Estimating subspace coefficients with quadratic estimators gives statistically consistent parameter posteriors to those obtained using the standard power-spectrum estimators. When the quadratic estimators are adopted, we do *not* need to convolve the theory model in the likelihood; this expedites the sampling by avoiding the need for Hankel transforms.

While the various datasets are consistent overall, we do observe small shifts in some parameters, particularly $\omega_{\rm cdm}$ and the primordial amplitude A_s. Some variation is expected since (a) the ML and FKP estimators weight the data in slightly different manners, and (b) by throwing away noisy modes, the subspace coefficients have different statistical properties to the data. Furthermore, Ref. [27] demonstrated that noise in the covariance matrix for the 1000 mock windowed power-spectrum dataset will cause shifts in the best-fit parameters, while this is reduced for the compressed statistic. If one had analyzed the mean of a set of mocks, these shifts are be expected to vanish (e.g., [27]). Note also that A_s and the linear galaxy bias b_1 are highly correlated (since the real-space linear power spectrum is proportional to $b_1 A_s^2$); thus, the A_s shift in the ML dataset is reflected by a corresponding b_1 shift. Furthermore, the posteriors are not simply reproducing the fiducial cosmology, as evidenced by the tidal galaxy bias $b_{\mathcal{G}_2}$ which is significantly above the input value in all cases. As in previous sections, we find highly consistent results when the subspace coefficients are estimated using FKP and ML pixel weights. In this context, we therefore find little benefit from applying the ML weights instead of the FKP quadratic estimator (though it allows one to compute the Fisher matrix and bias terms on an even footing and is, in practice, not much slower), but we caution that this will depend on the dataset and analysis in question. If one is interested in parameters that impact only large-scale modes (e.g., $f_{\rm NL}$),

[27]These are publicly available at github.com/michalychforever/lss_montepython.

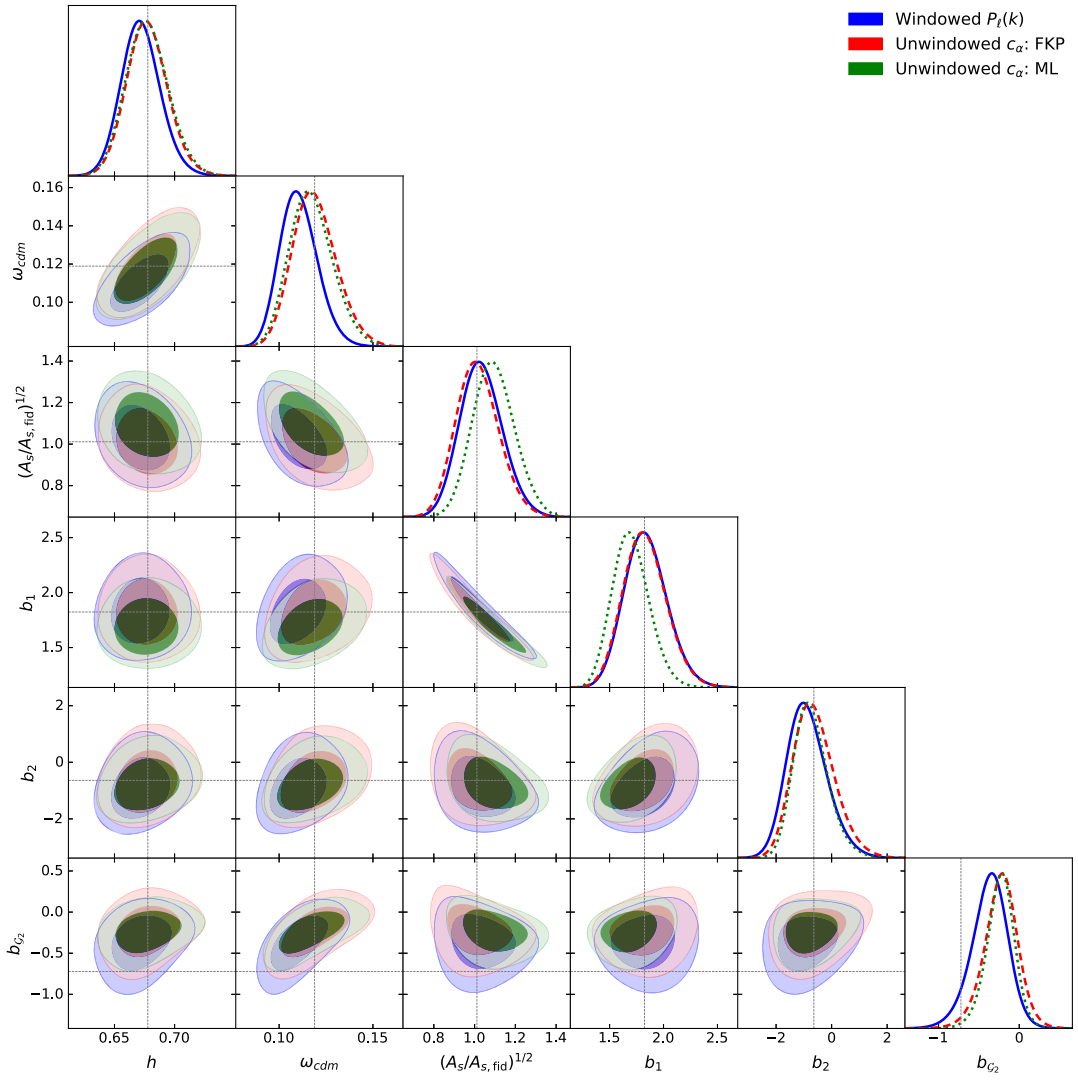

FIG. 6. Posterior contours for a MCMC inference on a segment of the BOSS dataset, either analyzing the windowed power-spectrum multipoles in 98 bins estimated using NBODYKIT (full blue) or 15 compressed subspace coefficients estimated via the quadratic estimators discussed in this work, utilizing FKP (dashed red) or ML (dotted green) pixel weights. The latter are analyzed *without* window convolution of the theory model. One thousand (150) mocks are used to define the sample covariance for the windowed (unwindowed) analyses, supplemented with a correlated theoretical error covariance. We plot only those parameters directly sampled by MONTEPYTHON; another four are marginalized analytically. Thin lines show the parameter values in the fiducial PATCHY cosmology, and we set $\log(10^{10}A_{s,\mathrm{fid}}) = 3.044$. While there are some differences among datasets, all are statistically consistent.

the ML approach is likely to have greater utility as the FKP approach is suboptimal at low k. For this survey, restricting to $k \in [0.01, 0.1]h$ Mpc^{-1} affords similar conclusions, as discussed in Appendix D.

D. Comparison of quadratic estimators: ML versus FKP weights

The above results merit some discussion regarding the optimality of the ML estimator compared to the quadratic estimator with FKP weights.[28] For a BOSS-like survey, the ML estimator clearly does not lead to a higher-precision power spectrum or to a significant increase in constraining power on the underlying parameters. To understand the reasons for this, we first make a number of notes:

(i) If the unclustered number density $n(\mathbf{r})$ is uniform, the FKP weighting scheme is optimal, provided the bins are small.[29]

(ii) On small scales, the FKP estimator is optimal, provided that P_{FKP} is close to the associated

[28] A similar discussion of this in the CMB context is presented in Ref. [24].

[29] For finite-sized bins, the ratio scales as $\int_{\mathbf{k}}[P(k)+1/\bar{n}]^2 \int_{\mathbf{k}}[P(k)+1/\bar{n}]^{-2}/N_{\mathrm{modes}}(k)^2$ in each bin, which is, in practice, very close to unity.

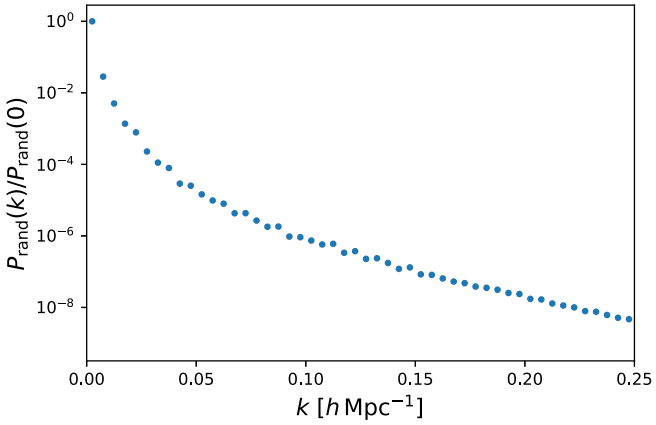

FIG. 7. Power spectrum of the unclustered random particles in the BOSS DR12 data chunk considered in this work, normalized to the $k \to 0$ mode. For simplicity, we plot only the monopole, and adopt the same k binning as in previous plots, but including two additional large-scale modes at $k < 0.01h$ Mpc^{-1}. This is computed as $P_{\rm rand}(k_a) = \int_{\mathbf{k}} \Theta_a(\mathbf{k})|n(\mathbf{k})|^2/N^a_{\rm modes}$, where $N^a_{\rm modes}$ is the number of modes in the bin centered at k_a, and $n(\mathbf{r})$ is the normalized random particle density.

power-spectrum amplitude. This follows directly from the covariance definition (14); if the window function is compact, $n(\mathbf{r})n(\mathbf{r}')\int_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}P(\mathbf{k}) \approx P_{\rm FKP}n(\mathbf{r}r)\delta_D(\mathbf{r}-\mathbf{r}')$, resulting in the FKP form.

(iii) If the density of galaxies is low, such that $\bar{n}P(k) \ll 1$, the covariance is dominated by the Poisson shot-noise component; thus, ML weights are not important (and the ML and FKP quadratic estimators converge).

(iv) Corrections to the FKP weighting are thus expected only to be important on scales where the window function has significant power *and* where $\bar{n}P(k)$ is large.

In the above subsections, we have considered the application of the quadratic power-spectrum estimators to the power-spectrum multipoles of a single BOSS DR12 chunk, which has both large volume ($1.46h^{-3}$ Gpc^3) and relatively low number density ($\bar{n} \sim 2 \times 10^{-4}h^3$ Mpc^{-3}). Furthermore, its window function is remarkably compact, as shown in Fig. 7. The power spectrum of the random particles (and hence, the window function) is less than 1% of its zero lag value even by $k = 0.01h$ Mpc^{-1} and drops significantly beyond this. In practice, we have omitted the first two k bins ($k < 0.01h$ Mpc^{-1}) from all plots in this work due to their contamination by systematic and integral-constraint effects, arising from the unknown mean galaxy density and imperfectly subtracted foreground modes. These results indicate that one may not expect a significant improvement when using the ML weights instead of FKP, matching the conclusions of this work. We argue that it is still useful to implement the quadratic estimator however (even in FKP form) due to the useful property of the resulting spectra being free from the survey window function and leading-order shot-noise and discretization artifacts. Furthermore, the ML approach offers the useful feature that the Fisher matrix and bias terms can be computed directly from the simulation suite used to define the fiducial cosmology and does not lead to much increased computation time. The analysis presented in Appendix D provides further proof of this notion, whereupon the fiducial FKP power $P_{\rm FKP}$ is reduced by a factor of 10 to $P_{\rm FKP} = 10^3h^{-3}$ Mpc^3, considerably below the amplitude of the observed power-spectrum multipole at $k = 0.25h$ Mpc^{-1}. If the estimated power-spectrum multipoles had strong dependence on $P_{\rm FKP}$, this would indicate that the survey had some significant anisotropy, and one may expect stronger constraints to be wrought from the ML estimator. In fact, we find very similar power spectrum estimates in this case, suggesting that the impact of the spatially varying number density $n(\mathbf{r})$ is small here.

We may compare the above conclusions to the results of Ref. [19], which used an optimal quadratic estimator to compute the local-type primordial non-Gaussianity (PNG) parameter $f^{\rm loc}_{\rm NL}$ from the eBOSS quasar sample over a broad redshift range. The study found ∼15%–40% improvements in $\sigma(f^{\rm loc}_{\rm NL})$ (depending on the quasar PNG response), drawing a significantly different conclusion to that of this work. However, this analysis takes a somewhat different approach; while we have opted to construct estimators for the galaxy power spectrum itself, Ref. [19] opted to measure $f^{\rm loc}_{\rm NL}$ directly, and took into account the evolution of the density field across the sample. In this case, the principal difference between the optimal and FKP estimators lies in weighting the galaxies differently as a function of redshift, which, combined with the large redshift range of the sample ($0.8 < z < 2.2$), leads to the increased precision $f_{\rm NL}$ measurements. If one was concerned with measuring just the power spectrum at a single effective redshift and the redshift range was small, we would expect smaller gains, as found here.

One additional subtlety is worth noting: If one compares the unwindowed-FKP-weighted estimators of this work with the standard windowed prescription measuring all power-spectrum modes up to some $k_{\rm max}$, the latter will generically provide a slightly larger signal-to-noise ratio. This occurs since the window function smooths the spectrum with a characteristic scale Δk; thus, the windowed estimator, in fact, includes true power-spectrum modes up to $k_{\rm max} + \Delta k$. Since the variance of a given mode decreases with k, this leads to a slight signal-to-noise ratio increase overall due to the windowed power spectra effectively using a somewhat larger $k_{\rm max}$. This effect is not obvious in the results presented herein (since modes at large k contribute little to the parameter constraints) but would be of greater importance if $k_{\rm max}$ were reduced. If one (at least formally) were to force both spectra to use only preconvolved modes in the same k range, the constraints would be identical however.

It is instructive to consider the scenarios in which one *would* expect the ML approach to outshine that of FKP. Analytic testing of this is difficult, since the benefits of the ML scheme vanish when employing simplifications such as a fixed number density, and any nontrivial scenario requires inversion of the pixel covariance $\mathsf{C}_{\rm fid}$. However, the above discussion implies that the ML approach is expected to be of greatest utility when (a) the survey volume is small, (b) the number density varies significantly across the survey, with characteristic scale similar to the wave numbers of interest, (c) the number density is high, such that we are far from the shot-noise dominated regime, or (d) we are interested in parameters depending primarily on large-scale physics, such as $f_{\rm NL}$. We leave detailed discussion of these regimes to future work, but note that an example satisfying several of these constraints is the DESI Bright Galaxy Survey [28].

VI. SUMMARY AND OUTLOOK

Optimal estimation of the galaxy power spectrum was a subject of fervent interest at the turn of the 21st century (e.g., [3,11,12,23]), but has thereafter been largely ignored. In this work, we discuss the extent to which quadratic estimators can be used to measure the power-spectrum multipoles in a manner that removes the effects of the survey window function. Given a set of simulations with known (or measurable) power spectrum, as well as the galaxy positions and random particle catalog appropriate for the data, quadratic estimators can be constructed to infer the underlying unwindowed power spectra. Two variants are considered here: a ML approach (which is optimal in the limit of Gaussian statistics) and a simpler approximation based on the FKP weighting scheme (e.g., [7]). Both may be computed from the pixelized density field assisted by judicious use of Monte Carlo averaging and, for the ML weights, conjugate-gradient descent techniques.

Our main conclusions are as follows:

(i) Quadratic estimators allow for robust estimation of unwindowed power-spectrum band powers. Output band powers have a close-to-diagonal covariance, unlike the windowed case.

(ii) The estimators are written as the sum of a smooth fiducial model and a difference between a quantity computed from the dataset and simulations. This implies that they are free from the leading-order effects of gridding, binning, fingers of God, and non-Poissonian shot noise, and allows one to use coarse grids (down to a Nyquist frequency of $k_{\rm max}$) in their estimation. Furthermore, unless one includes the first k bin, they are free from integral-constraint effects.

(iii) The approach is simply extended to the direct estimation of compressed (unwindowed) power-spectrum coefficients. This is fast (since the dimensionality is reduced) and does not require the statistic to be binned.

(iv) Applying the ML and FKP quadratic estimators to a data chunk from BOSS shows them to give statistically compatible parameter constraints to those from conventional approaches. While the ML approach requires a little more computational power and, for BOSS, does not lead to an increase in constraining power, some of the terms in its estimation are somewhat simpler to compute than for FKP weights, and the computation time is very similar in practice. We expect this to be of use for surveys that are particularly anisotropic, small volume, dense, or whose analysis principally relies on large-scale modes, such as $f_{\rm NL}$-based studies.

A number of extensions are possible. First, as discussed in Appendix B, one may formulate analogous *cubic* estimators for the power spectrum that are optimal in the presence of mild non-Gaussianity. While this is unlikely to be of great use for BOSS-like samples, it may be of use for deep surveys such as the DESI Bright Galaxy Survey [28], which has considerably lower shot noise and thus contains useful information on short scales. We leave the practical implementation of this algorithm to others, though we note it is only a simple extension of that presented herein.

Of greater interest is the application to unwindowed bispectra. Via a similar route, one may derive a cubic bispectrum estimator, which is optimal in the limit of mild non-Gaussianity. Unlike for the power spectrum, this estimator does not require a fiducial bispectrum model (and hence, knowledge of the bispectrum window), but just a fiducial *power-spectrum* model to compute $\mathsf{C}_{\rm fid}^{-1}$. In this context, compressing the statistic to a (separable) subspace is of great utility, since (a) it dramatically reduces the number of bispectrum bins, and (b) it obviates the need for the theory model to be binned or window convolved (which is difficult to do efficiently). This will be considered in future work.

The techniques discussed in this work thus provide a promising renaissance for the optimal estimators of old, allowing for unwindowed spectral estimates to be efficiently computed both in and beyond the Gaussian limit. We expect them to be of great use in future cosmological analyses, particularly for studies based upon the three-point function.

ACKNOWLEDGMENTS

It is a pleasure to thank Giovanni Cabass, Emanuele Castorina, Mikhail Ivanov, Marcel Schmittfull, Marko Simonović, David Spergel, and Matias Zaldarriaga for insightful conversations that motivated and improved this work. We are additionally grateful to the anonymous referees for extensive comments that helped improve the clarity of this work. O. H. E. P. acknowledges funding from the Nancy Grace Roman Space Telescope (WFIRST) program through Grants No. NNG26PJ30C and

No. NNN12AA01C. The authors are pleased to acknowledge that the work reported in this paper was substantially performed using the Princeton Research Computing resources at Princeton University, which is a consortium of groups led by the Princeton Institute for Computational Science and Engineering and Office of Information Technology's Research Computing.

APPENDIX A: PROPERTIES OF THE QUADRATIC ESTIMATOR

Below, we demonstrate a number of useful properties of the general quadratic estimator of (3), and, by extension the ML estimator central to this work. While we denote the parameters by $\mathbf{p}$, this applies equally well for the subspace coefficients. For clarity, we begin by recapitulating the definition

$$\hat{p}_\alpha = p^{\rm fid}_\alpha + \sum_\beta F^{-1}_{\alpha\beta}(\hat{q}_\beta - \bar{q}_\beta), \qquad \hat{q}_\alpha = \frac{1}{2}\mathbf{d}^T\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}\mathbf{d} \tag{A1}$$

(dropping the "QE" superscript for brevity) for arbitrary positive-definite weighting matrix H.

1. Bias

In expectation, we have

$$\begin{aligned}\mathbb{E}[\hat{p}_\alpha] &= p^{\rm fid}_\alpha + \sum_\beta F^{-1}_{\alpha\beta}(\mathbb{E}[\hat{q}_\beta] - \bar{q}_\beta)\\ &= p^{\rm fid}_\alpha + \frac{1}{2}\sum_\beta F^{-1}_{\alpha\beta}(\mathrm{Tr}[\mathsf{C}_D\mathsf{H}^{-1}\mathsf{C}_{,\beta}\mathsf{H}^{-1}] - \bar{q}_\beta)\end{aligned} \tag{A2}$$

using the definition $\langle\mathbf{d}\mathbf{d}^T\rangle = \mathsf{C}_D$, where C_D is the data covariance. Assuming the covariance to be linear in the parameter p_α, we can write $\mathsf{C}_D - \mathsf{C}_{\rm fid} = \sum_\gamma (p^{\rm true}_\gamma - p^{\rm fid}_\gamma)C_{,\gamma}$ (assuming that the fiducial noise matrix N to be equal to that of the data and that the sum is over all possible band powers), thus,

$$\mathbb{E}[\hat{p}_\alpha] = \sum_\beta F^{-1}_{\alpha\beta}\sum_\gamma p^{\rm true}_\gamma F_{\beta\gamma} = \sum_\gamma \delta^K_{\alpha\gamma}p^{\rm true}_\gamma = p^{\rm true}_\alpha, \tag{A3}$$

where we have inserted the bias and Fisher definitions of (4). We thus find the quadratic estimator to be unbiased for all choices of H.

2. Covariance

The covariance of $\hat{\mathbf{p}}$ can be written in terms of the covariance of $\hat{\mathbf{q}}^{\rm QE}$,

$$\mathrm{cov}(\hat{p}_\alpha, \hat{p}_\beta) = \sum_{\gamma\delta} F^{-1}_{\alpha\gamma}F^{-1}_{\beta\delta}\mathrm{cov}(\hat{q}_\gamma, \hat{q}_\delta) \tag{A4}$$

with

$$\mathrm{cov}(\hat{q}_\alpha, \hat{q}_\beta) = \frac{1}{4}[\langle d_i d_j d_j d_k\rangle - \langle d_i d_j\rangle\langle d_k d_l\rangle]Q^{ij}_\alpha Q^{kl}_\beta \tag{A5}$$

defining $(\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}) = \mathsf{Q}_\alpha \equiv -\partial_\alpha\mathsf{C}^{-1}$ and employing the Einstein summation for repeated pixel indices i, j, k, l. Employing Wick's theorem,

$$\begin{aligned}\mathrm{cov}(\hat{q}_\alpha, \hat{q}_\beta) &= \frac{1}{4}[\mathsf{C}^D_{ik}\mathsf{C}^D_{jl} + \mathsf{C}^D_{jk}\mathsf{C}^D_{il} + \mathsf{T}^D_{ijkl}]\mathsf{Q}^{ij}_\alpha\mathsf{Q}^{kl}_\beta\\ &= \frac{1}{2}\mathrm{Tr}[\mathsf{C}^D\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}\mathsf{C}^D\mathsf{H}^{-1}\mathsf{C}_{,\beta}\mathsf{H}^{-1}]\\ &\quad + \frac{1}{4}\mathsf{T}^D_{ijkl}\mathsf{Q}^{ij}_\alpha\mathsf{Q}^{kl}_\beta,\end{aligned} \tag{A6}$$

where $\mathsf{T}^D_{ijkl} \equiv \langle d_i d_j d_k d_l\rangle_c$ is the connected four-point function of the data. While this is generally complex, for the ML estimator with $p^{\rm fid} = p^{\rm true}$ and neglecting non-Gaussianity, we have $\mathsf{C}_{\rm fid} = \mathsf{C}_D$ and $\mathsf{T}_D = 0$, and thus,

$$\begin{aligned}\mathrm{cov}(\hat{q}^{\rm ML}_\alpha, \hat{q}^{\rm ML}_\beta) &\to \frac{1}{2}\mathrm{Tr}[\mathsf{C}^{-1}_{\rm fid}\mathsf{C}_{,\alpha}\mathsf{C}^{-1}\mathsf{C}_{,\beta}] = F_{\alpha\beta}\\ &\Rightarrow \mathrm{cov}(\hat{p}^{\rm ML}_\alpha, \hat{p}^{\rm ML}_\beta) \to F^{-1}_{\alpha\beta};\end{aligned} \tag{A7}$$

thus, the Fisher matrix is just the inverse covariance matrix. From this, we also expect the quantities $\sum_\beta F^{-1/2}_{\alpha\beta}\hat{q}^{\rm ML}_\beta$ to have a covariance close to the identity matrix.[30]

3. Optimality

The Cramér-Rao theorem states that, if an estimator $\hat{p}_\alpha$ is optimal, it must satisfy

$$\mathrm{cov}(\hat{p}_\alpha, \hat{p}_\beta)|_{\rm CR\ bound} = \left\langle\frac{1}{2}\frac{\partial^2\mathcal{L}[\mathbf{d}](\mathbf{p})}{\partial p_\alpha\partial p_\beta}\right\rangle^{-1} \equiv \mathcal{I}^{-1}_{\alpha\beta}, \tag{A8}$$

where the right-hand side is the inverse Fisher information for negative log-likelihood $\mathcal{L}[\mathbf{d}](\mathbf{p}) = -2\log L[\mathbf{d}](\mathbf{p})$ depending on parameters $\mathbf{p}$ and data $\mathbf{d}$ (e.g., [48]). Assuming the Gaussian likelihood of (6) evaluated at the true covariance C_D, the Fisher information is straightforwardly derived,

$$\begin{aligned}\mathcal{I}_{\alpha\beta} &= \left\langle\frac{1}{2}\mathrm{Tr}[\mathsf{C}^{-1}\mathsf{C}_{,\alpha}\mathsf{C}^{-1}\mathsf{C}_{,\beta}\mathsf{C}^{-1}(2\mathbf{d}\mathbf{d}^T - \mathsf{C})]\right\rangle_{\mathsf{C}=\mathsf{C}^D}\\ &= \frac{1}{2}\mathrm{Tr}[\mathsf{C}^{-1}\mathsf{C}_{,\alpha}\mathsf{C}^{-1}\mathsf{C}_{,\beta}]_{\mathsf{C}^D=\mathsf{C}} \equiv F_{\alpha\beta}|_{\mathsf{C}=\mathsf{C}^D}.\end{aligned} \tag{A9}$$

Comparison with (A7) shows that the ML estimator of (9) saturates its Cramér-Rao bound and is thus optimal if

[30]Reference [23] advocates plotting this quantity for this reason.

$\mathbf{p}^{\rm fid} = \mathbf{p}^{\rm true}$, neglecting non-Gaussianity. This is, in fact, guaranteed for any ML estimator.

APPENDIX B: OPTIMAL ESTIMATORS WITH WEAK NON-GAUSSIANITY

If the likelihood of $\mathbf{d}$ is non-Gaussian, the quadratic power-spectrum estimator becomes nonoptimal, even in the limit of $\mathsf{C}_{\rm fid} = \mathsf{C}_D$. Here, we discuss the leading-order corrections to the maximum-likelihood estimator in the presence of weak non-Gaussianity, either from higher-order Poissonian noise corrections or density field non-Gaussianities. We begin with the Edgeworth expansion for the likelihood around the Gaussian form of (6) (henceforth denoted L_G)

$$L[\mathbf{d}](\mathbf{p}) = L_G[\mathbf{d}](\mathbf{p})\left[1+\frac{1}{3!}\mathsf{B}^{ijk}\mathcal{H}_{ijk}+\frac{1}{4!}\mathsf{T}^{ijkl}\mathcal{H}_{ijkl} + \frac{1}{6!}(\mathsf{B}^{ijk}\mathsf{B}^{lmn}+9\,\text{perms})\mathcal{H}_{ijklmn}+\cdots\right] \tag{B1}$$

(e.g., [30,49]), where the Hermite tensors $\mathcal{H}_{i_1\ldots i_n}$ are defined in Ref. [30], depending on products of $\mathsf{C}^{-1}\mathbf{d}$ and C^{-1}.[31] Of importance here is the first defined by

$$\mathcal{H}_{ijk} = h_i h_j h_k - (\mathsf{C}_{ij}h_k + 2\,\text{perms}), \tag{B2}$$

where $h_i \equiv [\mathsf{C}^{-1}\mathbf{d}]_i$. In (B1), we have included only terms involving the three-point correlator $\mathsf{B}^{ijk} \equiv \langle d^i d^j d^k\rangle$ and four-point correlator $\mathsf{T}^{ijkl} \equiv \langle d^i d^j d^k d^l\rangle - [\langle d^i d^j\rangle\langle d^k d^l\rangle + 2\,\text{perms}]$. Note that all terms are assumed to be evaluated at the fiducial cosmology. More useful is the negative log-likelihood

$$\ell[\mathbf{d}](\mathbf{p}) = \ell_G[\mathbf{d}](\mathbf{p}) - \frac{1}{3!}\mathsf{B}^{ijk}\mathcal{H}_{ijk} - \frac{1}{4!}\mathsf{T}^{ijkl}\mathcal{H}_{ijkl} + \frac{1}{72}\mathsf{B}^{ijk}\mathsf{B}^{lmn}[\mathcal{H}_{ijk}\mathcal{H}_{lmn} - \mathcal{H}_{ijklmn}] + \cdots \tag{B3}$$

expanding the logarithm to second order and noting that the Hermite tensors are fully symmetric.

To proceed, we require the gradient of $\ell[\mathbf{d}](\mathbf{p})$ with respect to the band powers, as in (7). Note that the band powers appear *both* in the inverse covariance matrices C^{-1} (present in the Hermite tensors) and the three-point averages B^{ijk} through Poisson noise corrections. Following some algebra, this gives the leading-order contribution

$$\begin{aligned}-\partial_\alpha\ell[\mathbf{d}](\mathbf{p}) =& \frac{1}{2}\mathsf{Q}^\alpha_{ij}[d^i d^j - \mathsf{C}^{ij}] \\ &+\frac{1}{6}\mathsf{B}^{ijk}_{,\alpha}\mathsf{C}^{-1}_{il}\mathsf{C}^{-1}_{jm}\mathsf{C}^{-1}_{kn}[d^l d^m d^n - 3\mathsf{C}^{lm}d^n] \\ &-\frac{1}{2}\mathsf{B}^{ijk}\mathsf{C}^{-1}_{il}\mathsf{C}^{-1}_{jm}\mathsf{Q}^\alpha_{kn}[d^l d^m d^n - d^l\mathsf{C}^{mn} - \mathsf{C}^{lm}d^n] \\ &+\cdots\end{aligned} \tag{B4}$$

using $\mathsf{Q}_\alpha \equiv \mathsf{C}^{-1}\mathsf{C}_{,\alpha}\mathsf{C}^{-1}$, as before. Here, the first term is the Gaussian piece, with the second and third being cubic corrections from noise and gravitational non-Gaussianities. This motivates the following power-spectrum estimator containing quadratic and cubic contributions [by analogy to (9) for the Gaussian ML case]:

$$\begin{aligned}\hat{p}^{\rm NG}_\alpha =& p^{\rm fid}_\alpha + \sum_\beta F^{-1,{\rm NG}}_{\alpha\beta}(\hat{q}^{\rm NG}_\beta - \bar{q}^{\rm NG}_\beta), \\ \hat{q}^{\rm NG}_\alpha =& \frac{1}{2}\mathsf{Q}^\alpha_{ij}d^i d^j + \frac{1}{6}\mathsf{B}^{ijk}_{,\alpha}\mathsf{C}^{-1}_{il}\mathsf{C}^{-1}_{jm}\mathsf{C}^{-1}_{kn}[d^l d^m d^n - 3\mathsf{C}^{lm}d^n] \\ &-\frac{1}{2}\mathsf{B}^{ijk}\mathsf{C}^{-1}_{il}\mathsf{C}^{-1}_{jm}\mathsf{Q}^\alpha_{kn}[d^l d^m d^n - d^l\mathsf{C}^{mn} - \mathsf{C}^{lm}d^n].\end{aligned} \tag{B5}$$

While the bias and Fisher term could be derived from the realization-averaged second derivative $\partial_\alpha\partial_\beta\ell[\mathbf{d}](\mathbf{p})$, this is arduous to compute, so we here take an alternative route, first noting that $\hat{\mathbf{q}}^{\rm NG}$ has expectation

$$\begin{aligned}\mathbb{E}[\hat{q}^{\rm NG}_\alpha] =& \frac{1}{2}\mathsf{Q}^\alpha_{ij}\mathsf{C}^{ij}_D + \frac{1}{6}\mathsf{B}^{ijk}_{,\alpha}\mathsf{C}^{-1}_{il}\mathsf{C}^{-1}_{jm}\mathsf{C}^{-1}_{kn}\mathsf{B}^{lmn}_D \\ &-\frac{1}{2}\mathsf{B}^{ijk}\mathsf{C}^{-1}_{il}\mathsf{C}^{-1}_{jm}\mathsf{Q}^\alpha_{kn}\mathsf{B}^{lmn}_D,\end{aligned} \tag{B6}$$

where the subscripts D denote the data. Both the two- and three-point expectations C_D and B_D can be written in the form $\mathsf{X}_D = \mathsf{X} + \sum_\gamma (p^{\rm true}_\gamma - p^{\rm fid}_\gamma)X_{,\gamma}$, assuming the fiducial model to have the correct noise properties.[32] This gives

$$\begin{aligned}\mathbb{E}[\hat{q}^{\rm NG}_\alpha] =& \sum_\gamma (p^{\rm true}_\gamma - p^{\rm fid}_\gamma)\left[\frac{1}{2}\mathsf{Q}^\alpha_{ij}\mathsf{C}^{ij}_{,\gamma} + \frac{1}{6}\mathsf{B}^{ijk}_{,\alpha}\mathsf{C}^{-1}_{il}\mathsf{C}^{-1}_{jm}\mathsf{C}^{-1}_{kn}\mathsf{B}^{lmn}_{,\gamma} - \frac{1}{2}\mathsf{B}^{ijk}\mathsf{C}^{-1}_{il}\mathsf{C}^{-1}_{jm}\mathsf{Q}^\alpha_{kn}\mathsf{B}^{lmn}_{,\gamma}\right] \\ &+\left[\frac{1}{2}\mathsf{Q}^\alpha_{ij}\mathsf{C}^{ij} + \frac{1}{6}\mathsf{B}^{ijk}_{,\alpha}\mathsf{C}^{-1}_{il}\mathsf{C}^{-1}_{jm}\mathsf{C}^{-1}_{kn}\mathsf{B}^{lmn} - \frac{1}{2}\mathsf{B}^{ijk}\mathsf{C}^{-1}_{il}\mathsf{C}^{-1}_{jm}\mathsf{Q}^\alpha_{kn}\mathsf{B}^{lmn}\right] \\ \equiv& \sum_\gamma (p^{\rm true}_\gamma - p^{\rm fid}_\gamma)F^{\rm NG}_{\alpha\gamma} + \bar{q}^{\rm NG}_\alpha\end{aligned} \tag{B7}$$

[31] For clarity, we omit the explicit dependence of $L[\mathbf{d}]$ on the correlators beyond $\mathbf{p}$. For the purposes of this section, we also omit the subscripts on C, assuming all covariances to be defined at the true parameters.

[32] This also assumes the fiducial model to have the correct gravitational three-point function since it is contained within B; small deviations are unlikely to cause significant bias to the estimator however.

defining the bias and Fisher matrix in the final line, such that the resulting $\hat{p}_\alpha^{\rm NG}$ estimators satisfy $\mathbb{E}[\hat{\mathbf{p}}^{\rm NG}] = \mathbf{p}^{\rm fid}$. For $\mathbf{p}^{\rm fid} \to \mathbf{p}^{\rm true}$, this gives the optimal cubic correction to the power-spectrum estimator, to the extent in which we can neglect non-Gaussianity of the signal and noise beyond the three-point function.

Computation of the relevant terms in $\hat{p}_\alpha$ is possible given the vectors $\mathsf{C}^{-1}\mathbf{d}$ and $\mathsf{Q}^\alpha\mathbf{d}$, just as for the Gaussian case. Defining $h_i = [\mathsf{C}^{-1}\mathbf{d}]_i$ and $\phi_i^\alpha = [\mathsf{Q}^\alpha\mathbf{d}]_i$, we can write the estimator more succinctly as

$$\begin{aligned}\hat{q}_\alpha^{\rm NG} &= \frac{1}{2}d^i\phi_i^\alpha + \frac{1}{6}\mathsf{B}^{ijk}_{,\alpha}[h_ih_jh_k - 3\mathsf{C}_{ij}h_k] \\ &\quad - \frac{1}{2}\mathsf{B}^{ijk}[h_ih_j\phi_k^\alpha - h_i\mathsf{Q}^\alpha_{jk} - \mathsf{C}^{-1}_{ij}\phi_k^\alpha] \\ &= \frac{1}{2}d^i\phi_i^\alpha + \frac{1}{6}\mathsf{B}^{ijk}_{,\alpha}[h_ih_jh_k - 3\langle\tilde{h}_i\tilde{h}_j\rangle h_k] \\ &\quad - \frac{1}{2}\mathsf{B}^{ijk}[h_ih_j\phi_k^\alpha - h_i\langle\tilde{\phi}_j^\alpha\tilde{h}_k\rangle - \langle\tilde{h}_i\tilde{h}_j\rangle\phi_k^\alpha], \qquad \text{(B8)}\end{aligned}$$

where in the second line, we have replaced the C_{ij} and Q^α_{ij} matrices with averages over Monte Carlo realizations $\tilde{\mathbf{d}}$ generated at the fiducial cosmology (denoting $\tilde{h} \equiv \mathsf{C}^{-1}\tilde{\mathbf{d}}$, $\tilde{\phi}^\alpha \equiv \mathsf{Q}^\alpha\tilde{\mathbf{d}}$), allowing the operator to be computed as a product of two or three fields in configuration space, given the B and $\mathsf{B}_{,\alpha}$ weighting tensors, just as in Sec. III. Note the appearance of terms involving *both* Monte Carlo simulations $\mathbf{m}$ and the data $\mathbf{d}$. These did not appear in the Gaussian estimator and require two pixel grids being in memory simultaneously. While on its own, this does not pose a difficulty, we caution that, unless carefully treated, it may require extensive computation time to compute a sample covariance matrix, since the term is different for each simulation entering the covariance.

Following a similar approach, we can compute $\bar{q}_\alpha$ via Monte Carlo averaging as in (23):

$$\bar{q}_\alpha^{\rm NG} = \frac{1}{2}\langle\tilde{d}^i\tilde{\phi}_i^\alpha\rangle + \frac{1}{6}\langle\mathsf{B}^{ijk}_{,\alpha}\tilde{h}_i\tilde{h}_j\tilde{h}_k\rangle - \frac{1}{2}\langle\mathsf{B}^{ijk}\tilde{h}_i\tilde{h}_j\tilde{\phi}_k^\alpha\rangle. \qquad \text{(B9)}$$

While a similar procedure can be performed for $F^{\rm NG}_{\alpha\beta}$, it is more difficult to obtain a simply applicable form in this case, though Ref. [26] presents a discussion of this in the context of CMB bispectrum estimators. An alternative method is to note that, assuming the estimator to be close to optimal,

$$\mathrm{cov}(\hat{q}_\alpha^{\rm NG} - \hat{q}_\alpha^{\rm G}, \hat{q}_\beta^{\rm NG} - \hat{q}_\beta^{\rm G}) \approx F^{\rm NG}_{\alpha\beta} - F^{\rm G}_{\alpha\beta}, \qquad \text{(B10)}$$

where the superscript G indicates the Gaussian results of Sec. II. Since the Gaussian matrix can be straightforwardly computed as in Sec. III C, computing the non-Gaussian residual in this manner is expected to yield an accurate (and invertible) overall result, assuming it to be a small perturbation.

It remains to specify the form of B^{ijk}. For the spectroscopic surveys considered in this work, we have the three-point average

$$\begin{aligned}\mathsf{B}^{ijk} &= \langle[n_g - n](\mathbf{r}_i)[n_g - n](\mathbf{r}_j)[n_g - n](\mathbf{r}_k)\rangle \\ &= n(\mathbf{r}_i)\delta_D(\mathbf{r}_i - \mathbf{r}_j)\delta_D(\mathbf{r}_i - \mathbf{r}_k) \\ &\quad + [n(\mathbf{r}_i)n(\mathbf{r}_j)\xi(\mathbf{r}_i - \mathbf{r}_j)\delta_D(\mathbf{r}_i - \mathbf{r}_k) + 2\,\text{perms}] \\ &\quad + n(\mathbf{r}_i)n(\mathbf{r}_j)n(\mathbf{r}_k)\zeta(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k), \qquad \text{(B11)}\end{aligned}$$

where ζ is the three-point correlation function. Expressing the correlators in Fourier space, this yields

$$\begin{aligned}\mathsf{B}^{ijk} &= n(\mathbf{r}_i)\delta_D(\mathbf{r}_i - \mathbf{r}_j)\delta_D(\mathbf{r}_i - \mathbf{r}_k) + \left[n(\mathbf{r}_i)n(\mathbf{r}_j)\delta_D(\mathbf{r}_i - \mathbf{r}_k)\int_{\mathbf{k}} P(\mathbf{k};\mathbf{r}_j)e^{i\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_j)} + 2\,\text{perms}\right] \\ &\quad + n(\mathbf{r}_i)n(\mathbf{r}_j)n(\mathbf{r}_k)\int_{\mathbf{k}_1\mathbf{k}_2\mathbf{k}_3} B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)e^{i\mathbf{k}_1\cdot\mathbf{r}_{ij} + i\mathbf{k}_2\cdot\mathbf{r}_{jk} + i\mathbf{k}_3\cdot\mathbf{r}_{ki}}(2\pi)^3\delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \qquad \text{(B12)}\end{aligned}$$

denoting $\mathbf{r}_{ab} \equiv \mathbf{r}_a - \mathbf{r}_b$ and allowing for the line-of-sight dependence of $P(\mathbf{k})$ but assuming the bispectrum B to be isotropic for simplicity.[33] This now depends on the fiducial power-spectrum $P(\mathbf{k})$ and bispectrum $B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)$. The band-power derivative is given by

$$\mathsf{B}^{ijk}_{,\alpha} = n(\mathbf{r}_i)n(\mathbf{r}_j)\delta_D(\mathbf{r}_i - \mathbf{r}_k)\int_{\mathbf{k}}\Theta_a(\mathbf{k})L_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{r}}_j)e^{i\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_j)} + 2\,\text{perms}. \qquad \text{(B13)}$$

arising only from Poisson noise contractions.

[33]This procedure may be analogously generalized for the redshift-space bispectrum multipoles, just involving additional spherical harmonic factors in the estimators.

In the case of vanishing gravitational non-Gaussianity, i.e., $B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\equiv 0$, the power-spectrum estimators simplify, since they depend only on fields at two physical locations rather than three. In particular, the action of B and $\mathsf{B}_{,\alpha}$ on three fields $\{\mathbf{x},\mathbf{y},\mathbf{z}\}$ is given by

$$\begin{aligned}\mathsf{B}^{ijk}x_i y_j z_k &\to \int d\mathbf{r}[nxyz](\mathbf{r}) + \left[\int d\mathbf{r}_1 d\mathbf{r}_2 [nx](\mathbf{r}_1)[nyz](\mathbf{r}_2)\int_{\mathbf{k}} P(\mathbf{k})e^{i\mathbf{k}\cdot(\mathbf{r}_1-\mathbf{r}_2)} + 2\,\text{perms.}\right]\\ &= \int d\mathbf{r}[nxyz](\mathbf{r}) + \left[\int_{\mathbf{k}} P(\mathbf{k})[nx](\mathbf{k})[nyz](-\mathbf{k}) + 2\,\text{perms.}\right],\\ \mathsf{B}^{ijk}_{,\alpha}x_i y_j z_k &\to \int_{\mathbf{k}} \Theta_\alpha(\mathbf{k})[nx](\mathbf{k})[nyz](-\mathbf{k}) + 2\,\text{perms.}\end{aligned} \tag{B14}$$

assuming isotropic $P(\mathbf{k})$ for simplicity. Both may be straightforwardly computed via Fourier transforms and summations of the pixel grid, allowing for the $\hat{q}_\alpha$ coefficients to be swiftly computed. Indeed, the computational expense of computing the estimator at next-to-leading order in non-Gaussian noise is not parametrically larger than that of the Gaussian case, since it requires only quantities (h and ϕ) already computed for the Gaussian ML estimator. When allowing for *gravitational* non-Gaussianity, the operators become more difficult to apply, since the action of B cannot be simply reduced to a single summation in real or Fourier space. The calculation is greatly simplified if the fiducial gravitational bispectrum is separable however.

APPENDIX C: ESTIMATOR SIMPLIFICATIONS FOR FKP PIXEL WEIGHTS

Using the FKP weights of (16), the quadratic estimator (3) can be implemented either as for the ML estimator (Sec. III) or more directly, without the use of a fiducial cosmology. In this section, we briefly discuss the latter option.

Writing $\mathsf{H}_{\rm FKP} = w_{\rm FKP}(\mathbf{r})/n(\mathbf{r})$ where $w_{\rm FKP}(\mathbf{r}) \equiv [1+n(\mathbf{r})P_{\rm FKP}]^{-1}$ is the familiar FKP weight, the action of $\mathsf{H}^{-1}_{\rm FKP}$ on a map x is simply

$$\mathsf{H}^{-1}_{\rm FKP}[\mathbf{x}](\mathbf{r}) = \frac{w_{\rm FKP}(\mathbf{r})x(\mathbf{r})}{n(\mathbf{r})}, \tag{C1}$$

noting that we apply the weights to the grid rather than to the particles directly, in contrast to the windowed-FKP approach. $\hat{\mathbf{q}}^{\rm FKP}$ is given by

$$\begin{aligned}\hat{q}^{\rm FKP}_\alpha &= \frac{1}{2}\mathbf{d}^T\mathsf{H}^{-1}_{\rm FKP}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}_{\rm FKP}\mathbf{d} = \frac{1}{2}\int d\mathbf{r}d\mathbf{r}'\frac{w_{\rm FKP}(\mathbf{r})d(\mathbf{r})}{n(\mathbf{r})}\mathsf{C}_{,\alpha}(\mathbf{r},\mathbf{r}')\frac{w_{\rm FKP}(\mathbf{r}')d_{\rm FKP}(\mathbf{r}')}{n(\mathbf{r}')}\\ &= \frac{1}{2}\frac{4\pi}{2\ell+1}\sum_{m=-\ell}^{\ell}\int_{\mathbf{k}}\Theta_a(\mathbf{k})Y^*_{\ell m}(\mathbf{k})[w_{\rm FKP}d](-\mathbf{k})[Y_{\ell m}w_{\rm FKP}d](\mathbf{k}),\end{aligned} \tag{C2}$$

inserting (15), and applying the Yamamoto approximation, as in (20). This can be easily evaluated with a Fourier transform. Following the form of (27), the bias and Fisher components are given similarly as

$$\begin{aligned}\bar{q}^{\rm FKP}_\alpha &= \frac{1}{2}\text{Tr}[\mathsf{H}^{-1}_{\rm FKP}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}_{\rm FKP}\mathsf{N}] = \frac{(1+\alpha)}{2}\int d\mathbf{r}n(\mathbf{r})w^2_{\rm FKP}(\mathbf{r})\int_{\mathbf{k}}\Theta_a(\mathbf{k})L_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{r}})\\ &= \frac{(1+\alpha)}{2}\frac{4\pi}{2\ell+1}\sum_{m=-\ell}^{\ell}\int d\mathbf{r}n(\mathbf{r})w^2_{\rm FKP}(\mathbf{r})Y_{\ell m}(\mathbf{r})\int_{\mathbf{k}}Y^*_{\ell m}(\mathbf{k})\Theta_a(\mathbf{k}),\\ F^{\rm FKP}_{\alpha\beta} &= \frac{1}{2}\text{Tr}[\mathsf{H}^{-1}_{\rm FKP}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}_{\rm FKP}\mathsf{C}_{,\beta}]\\ &= \frac{1}{2}\int d\mathbf{r}d\mathbf{r}'n(\mathbf{r})w_{\rm FKP}(\mathbf{r})n(\mathbf{r}')w_{\rm FKP}(\mathbf{r}')\int_{\mathbf{k}\mathbf{k}'}\Theta_a(\mathbf{k})\Theta_b(\mathbf{k})L_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{r}}')L_{\ell'}(\hat{\mathbf{k}}'\cdot\hat{\mathbf{r}})e^{i(\mathbf{k}-\mathbf{k}')\cdot(\mathbf{r}-\mathbf{r}')}\\ &= \frac{1}{2}\frac{(4\pi)^2}{(2\ell+1)(2\ell'+1)}\sum_{m=-\ell}^{\ell}\sum_{m'=-\ell'}^{\ell'}\int_{\mathbf{k}\mathbf{k}'}\mathcal{F}[w_{\rm FKP}nY_{\ell'm'}](\mathbf{k}-\mathbf{k}')\mathcal{F}[w_{\rm FKP}nY_{\ell m}](\mathbf{k}'-\mathbf{k})Y^*_{\ell m}(\mathbf{k})\Theta_a(\mathbf{k})Y^*_{\ell'm'}(\mathbf{k}')\Theta_b(\mathbf{k}').\end{aligned} \tag{C3}$$

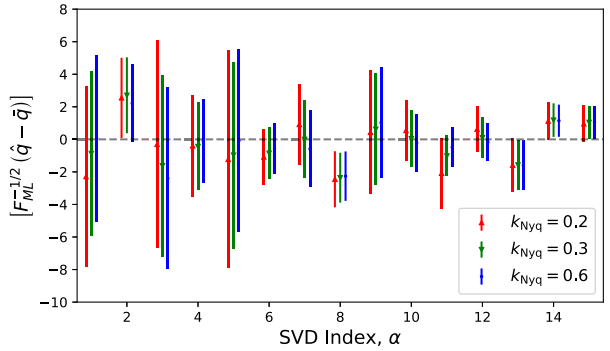

FIG. 8. Comparison of the subspace coefficients estimated from BOSS DR12 using three different choices of pixelization grid, with associated Nyquist frequencies given in $h\,\mathrm{Mpc}^{-1}$ units in the caption. The quadratic estimator is rescaled as in Fig. 5 (right panel) for clarity, with the dashed line indicating the values in the fiducial cosmology. All points are computed using the ML quadratic estimator of (9), using 150 PATCHY mocks to compute the covariances. We note very little difference from reducing the pixel density.

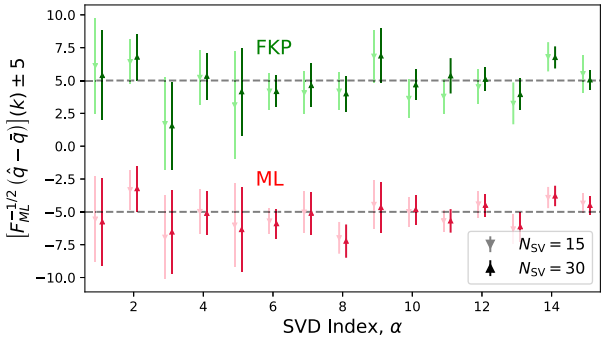

FIG. 9. Estimates of the subspace coefficients as a function of the total number of coefficients sampled ($N_{\rm SV}$). Results with 15 (30) coefficients are shown in light (dark) with the top (bottom) datasets computed using the FKP (ML) estimator. All cases assume $k_{\rm Nyq}=0.3h\ \mathrm{Mpc}^{-1}$, with the variances computed from 150 PATCHY mocks, and data are rescaled as in Fig. 5. Since the coefficients are somewhat correlated, there is a slight decrease in the error bars as $N_{\rm SV}$ is increased when using FKP weights.

The Fisher matrix term can be written as a convolution and a Fourier-space summation allowing for straightforward computation. Ignoring the effects of pixelization, $\bar{q}_\alpha = \frac{1}{2}(1+\alpha)N^a_{\rm modes}\delta^K_{\ell 0}\int d\mathbf{r} n(\mathbf{r}) w^2_{\rm FKP}(\mathbf{r})$ where $N^a_{\rm modes} = \int d\mathbf{k}\Theta_a(\mathbf{k})$, and $\hat{q}_\alpha$ and $\bar{q}_\alpha$ are just the usual windowed-FKP spectrum and shot noise multiplied by $N^a_{\rm modes}$ and a constant factor. In our case, multiplication by the Fisher matrix inverse $F^{-1,\rm FKP}_{\alpha\beta}$ acts to remove the effects of the window function and the bias term removes Poissonian shot noise. For a uniform survey and isotropic $P(k)$, $\mathcal{F}[w_{\rm FKP}n](\mathbf{p}) = (2\pi)^3\delta_D(\mathbf{p})w_{\rm FKP}n$; thus, $F^{\rm FKP}_{\alpha\beta} = \frac{1}{2}\delta^K_{\alpha\beta}N^\alpha_{\rm modes}w^2_{\rm FKP}n^2$, such that all bins are uncorrelated and the density field enters only via the combination $|d(\mathbf{k})|^2$, as expected.

APPENDIX D: QUADRATIC ESTIMATOR HYPERPARAMETERS

Below, we provide brief discussion of the effects of varying a number of hyperparameters controlling the implementation and application of the quadratic estimators, including the effects of binning, weighting, and pixelization.

1. Size of pixelization grid

A necessary step for computation of the quadratic estimators is assigning the data to a cuboidal grid, allowing for Fourier transforms to be easily computed via the FFT algorithm. For the simple windowed power-spectrum estimates, the pixel width L is usually chosen to ensure that the associated Nyquist frequency ($k_{\rm Nyq} = 2\pi/L$) is at least twice $k_{\rm max}$; this ensures that the spectra are free from discretization artifacts, though, due to the $\mathcal{O}(N_g \log N_g)$ scaling of FFTs, a denser grid requires greater computational time. In the quadratic estimator formalism, the leading-order effects of pixelization are canceled since the band powers are computed from a difference of quantities computed in the data and simulations; it may thus be feasible to use a smaller grid size.

Subspace coefficients computed with the ML estimator (Sec. IV) with three choices of pixel width are shown in Fig. 8. We observe consistent results with all three, albeit with a very slight increase in the error bars for $k_{\rm Nyq} = 0.2h\ \mathrm{Mpc}^{-1}$. This is particularly notable for the coarsest grid, since $k_{\rm Nyq}$ is below the $k_{\rm max} = 0.25h\ \mathrm{Mpc}^{-1}$ used to construct the basis vectors. This conclusion depends somewhat on the choice of basis. The subspace vectors preferentially weight wave numbers most sensitive to cosmological and nuisance parameters; thus, the high-k regions most affected by gridding artifacts are down-weighted. Repeating the analysis for the band powers shows that the results are independent of $k_{\rm Nyq}$ providing $k_{\rm Nyq} > k_{\rm max}$. The latter conclusion applies also to the FKP-based quadratic estimator. This weak restriction allows us to use relatively coarse grids and hence expedite the computation.

2. Number of basis vectors

As discussed in Sec. IV B, the measured subspace coefficients carry some dependence on the total number estimated, $N_{\rm SV}$. This is shown in Fig. 9 plotting the first 15 subspace coefficients from the ML and FKP estimators using two choices of $N_{\rm SV}$. For the ML case, the results are consistent, but the FKP approach suffers from inflated parameter covariances with $N_{\rm SV} = 15$, particularly for the later coefficients. This arises due to weighting-specific correlations between measured and unmeasured coefficients that bias the estimator (and increase its variance). We conclude that, when using the FKP estimator, it is

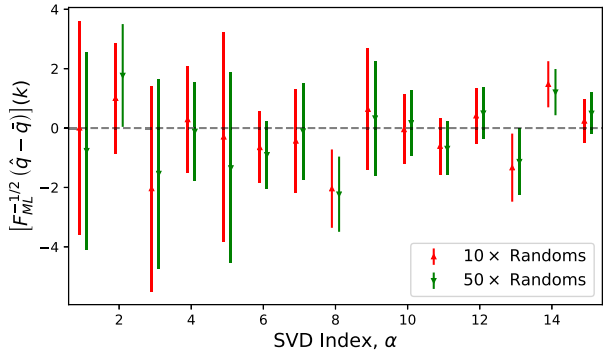

FIG. 10. As Fig. 8, but comparing the estimated subspace coefficients when the number of random particles is reduced. Results from random catalogs with 10× (50×) the galaxy density are shown in red (green) and all datasets use $k_{\rm Nyq} = 0.3h$ Mpc^{-1}, the ML estimator, and $N_{\rm SV} = 30$. While there are small shifts in the coefficients, these are not systematic, and are well within the error bars.

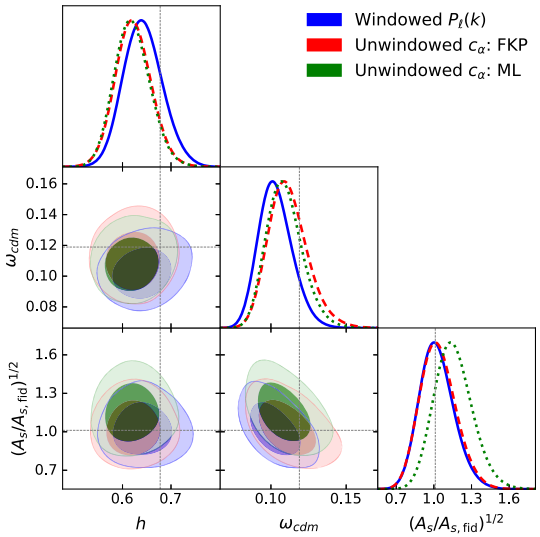

FIG. 11. As Fig. 6, but reducing $k_{\rm max}$ to $0.1h$ Mpc^{-1}. For clarity, we show only the cosmological parameters. Our conclusions regarding the efficacy of the various estimators are unchanged.

important to measure slightly more coefficients than will be used in the final analysis.

3. Number of random particles

The random particle density $n_r(\mathbf{r})$ enters the algorithm both to define the pixelized data $d(\mathbf{r}) = n_g(\mathbf{r}) - n_r(\mathbf{r})$ and as a proxy of the background number density $n(\mathbf{r})$ appearing in the pixel covariance matrix (14). Figure 10 considers the impact of this for the ML case [which is expected to be the most affected, since several of the $n(\mathbf{r})$ fields cancel when applying FKP weighting]. Reducing the density of randoms from 50 to 10 times that of the data causes slight stochastic shifts in the output coefficients, which is expected (and a similar effect is seen for the windowed power-spectrum estimators). Noting that the shifts are well within the error bars even for 10× randoms, we expect our main results (which use 50× randoms) to be robust.

4. Choice of $k_{\rm max}$

When performing parameter inference with the quadratic estimators, an important hyperparameter is the maximum wave number used, here set to $k_{\rm max} = 0.25h$ Mpc^{-1} in this work. While the exact choice is not of particular importance due to the correlated theoretical error that reduces sensitivity to short poorly modeled scales by substantially reducing $k_{\rm max}$, we can check how the quadratic estimator affects the information content of the large-scale modes. Since much of the information on cosmological parameters comes from intermediate scales (which allow one to constrain nuisance parameters [32]), this is a valid test of the approach. To this end, we have reduced $k_{\rm max}$ to $0.1h$ Mpc^{-1}, and recomputed the subspace coefficients of Sec. IV (since a change in $k_{\rm max}$ necessarily leads to a change in the basis vectors), before running the parameter inference as in Sec. V. The resulting constraints on cosmological parameters are shown in Fig. 11. Notably, our conclusions are very similar to those with $k_{\rm max} = 0.25h$ Mpc^{-1}: The quadratic estimators lead to slight (but not significant) shifts in the cosmological parameters, but the FKP and ML weights are similar, though with a slight shift to higher A_s in the latter case.

5. FKP weighting

Finally, we comment on the effects of changing the FKP coefficient $P_{\rm FKP}$ appearing in the FKP pixel weight matrix of (16). By reducing this from $P_{\rm FKP} = 10^4 h^{-3}$ Mpc^3 to $P_{\rm FKP} = 10^3 h^{-3}$ Mpc^3 (significantly below the value of the power-spectrum monopole on all tested scales), we would expect to increase the power-spectrum error bars if the effects of nonuniform survey number density were significant. The resulting power-spectrum multipoles are shown in Fig. 12, alongside those with the fiducial FKP weight used in Fig. 2. While there are minor perturbations to the spectra due to the different choice of weighting, we find statistically consistent results, and, importantly, do not observe inflated error bars, even at low k. While this may appear somewhat surprising, given that the FKP weight is not optimal in this case, it indicates that, for a broad survey such as BOSS, the exact pixel weighting scheme does not have a significant impact, likely because the window function is relatively compact, and the number density relatively consistent across the survey. That this does not have a significant impact further suggests that the ML approach will not greatly improve the BOSS measurement of $P_\ell(k)$, just as found in this work.

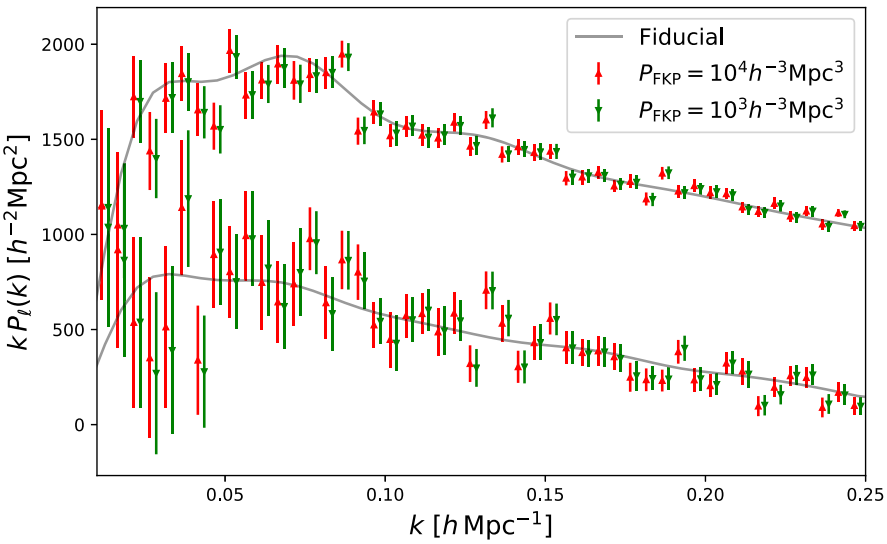

FIG. 12. As Fig. 2, but comparing two different FKP weighting schemes with $P_{\rm FKP} = 10^4 h^{-3}$ Mpc^3 (red) and $10^3 h^{-3}$ Mpc^3 (green). The second choice of weighting is suboptimal, since it significantly underestimates the true power. In both cases, spectra are computed using the quadratic estimators of Sec. II, and are thus unbiased by the survey window function.

[1] F. Beutler, H.-J. Seo, S. Saito, C.-H. Chuang, A. J. Cuesta, D. J. Eisenstein *et al.*, The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Anisotropic galaxy clustering in Fourier space, Mon. Not. R. Astron. Soc. **466**, 2242 (2017).

[2] H. Gil-Marín, J. E. Bautista, R. Paviot, M. Vargas-Magaña, S. de la Torre, S. Fromenteau *et al.*, The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Measurement of the BAO and growth rate of structure of the luminous red galaxy sample from the anisotropic power spectrum between redshifts 0.6 and 1.0, Mon. Not. R. Astron. Soc. **498**, 2492 (2020).

[3] H. A. Feldman, N. Kaiser, and J. A. Peacock, Power-spectrum analysis of three-dimensional redshift surveys, Astrophys. J. **426**, 23 (1994).

[4] K. Yamamoto, Optimal weighting scheme in redshift-space power spectrum analysis and a prospect for measuring the cosmic equation of state, Astrophys. J. **595**, 577 (2003).

[5] K. Yamamoto, M. Nakamichi, A. Kamino, B. A. Bassett, and H. Nishioka, A measurement of the quadrupole power spectrum in the clustering of the 2dF QSO survey, Publ. Astron. Soc. Jpn. **58**, 93 (2006).

[6] D. Bianchi, H. Gil-Marín, R. Ruggeri, and W. J. Percival, Measuring line-of-sight-dependent Fourier-space clustering using FFTs, Mon. Not. R. Astron. Soc. **453**, L11 (2015).

[7] N. Hand, Y. Li, Z. Slepian, and U. Seljak, An optimal FFT-based anisotropic power spectrum estimator, J. Cosmol. Astropart. Phys. 07 (2017) 002.

[8] D. Sorini, An optimally weighted estimator of the linear power spectrum disentangling the growth of density perturbations across galaxy surveys, J. Cosmol. Astropart. Phys. 04 (2017) 029.

[9] M. Tegmark, A. J. S. Hamilton, M. A. Strauss, M. S. Vogeley, and A. S. Szalay, Measuring the galaxy power spectrum with future redshift surveys, Astrophys. J. **499**, 555 (1998).

[10] M. Tegmark, How to measure CMB power spectra without losing information, Phys. Rev. D **55**, 5895 (1997).

[11] M. Tegmark, A. N. Taylor, and A. F. Heavens, Karhunen-Loève eigenvalue problems in cosmology: How should we tackle large data sets?, Astrophys. J. **480**, 22 (1997).

[12] J. R. Bond, A. H. Jaffe, and L. Knox, Estimating the power spectrum of the cosmic microwave background, Phys. Rev. D **57**, 2117 (1998).

[13] J. Borrill, Power spectrum estimators for large CMB datasets, Phys. Rev. D **59**, 027302 (1998).

[14] S. P. Oh, D. N. Spergel, and G. Hinshaw, An efficient technique to determine the power spectrum from cosmic microwave background sky maps, Astrophys. J. **510**, 551 (1999).

[15] A. J. S. Hamilton, Power spectrum estimation I. Basics, Lect. Notes Phys. **665**, 415 (2008).

[16] A. J. S. Hamilton, Power spectrum estimation II. Linear maximum likelihood, Lect. Notes Phys. **665**, 433 (2008).

[17] R. E. Smith and L. Marian, Towards optimal estimation of the galaxy power spectrum, Mon. Not. R. Astron. Soc. **454**, 1266 (2015).

[18] R. E. Smith and L. Marian, What is the optimal way to measure the galaxy power spectrum?, Mon. Not. R. Astron. Soc. **457**, 4285 (2016).

[19] E. Castorina, N. Hand, U. Seljak, F. Beutler, C.-H. Chuang, C. Zhao *et al.*, Redshift-weighted constraints on primordial non-Gaussianity from the clustering of the eBOSS DR14 quasars in Fourier space, J. Cosmol. Astropart. Phys. 09 (2019) 010.

[20] A. J. S. Hamilton, M. Tegmark, and N. Padmanabhan, Linear redshift distortions and power in the IRAS Point Source Catalog Redshift Survey, Mon. Not. R. Astron. Soc. **317**, L23 (2000).

[21] M. Tegmark, A. J. S. Hamilton, and Y. Xu, The power spectrum of galaxies in the 2dF 100k redshift survey, Mon. Not. R. Astron. Soc. **335**, 887 (2002).

[22] M. Tegmark, S. Dodelson, D. J. Eisenstein, V. Narayanan, R. Scoccimarro, R. Scranton *et al.*, The angular power spectrum of galaxies from early Sloan Digital Sky Survey Data, Astrophys. J. **571**, 191 (2002).

[23] M. Tegmark, M. R. Blanton, M. A. Strauss, F. Hoyle, D. Schlegel, R. Scoccimarro *et al.*, The three-dimensional power spectrum of galaxies from the Sloan Digital Sky Survey, Astrophys. J. **606**, 702 (2004).

[24] G. Efstathiou, Myths and truths concerning estimation of power spectra: The case for a hybrid estimator, Mon. Not. R. Astron. Soc. **349**, 603 (2004).

[25] U. Seljak, G. Aslanyan, Y. Feng, and C. Modi, Towards optimal extraction of cosmological information from nonlinear data, J. Cosmol. Astropart. Phys. 12 (2017) 009.

[26] K. M. Smith and M. Zaldarriaga, Algorithms for bispectra: Forecasting, optimal analysis and simulation, Mon. Not. R. Astron. Soc. **417**, 2 (2011).

[27] O. H. E. Philcox, M. M. Ivanov, M. Zaldarriaga, M. Simonović, and M. Schmittfull, Fewer mocks and less noise: Reducing the dimensionality of cosmological observables with subspace projections, Phys. Rev. D **103**, 043508 (2021).

[28] A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, L. E. Allen *et al.* (DESI Collaboration), The DESI Experiment Part I: Science, targeting, and survey design, arXiv:1611.00036.

[29] H. Gil-Marín, W. J. Percival, L. Verde, J. R. Brownstein, C.-H. Chuang, F.-S. Kitaura, S. A. Rodríguez-Torres, and M. D. Olmstead, The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: RSD measurement from the power spectrum and bispectrum of the DR12 BOSS galaxies, Mon. Not. R. Astron. Soc. **465**, 1757 (2017).

[30] E. Sellentin, A. H. Jaffe, and A. F. Heavens, On the use of the Edgeworth expansion in cosmology I: How to foresee and evade its pitfalls, arXiv:1709.03452.

[31] O. H. E. Philcox and Z. Slepian, Beyond Yamamoto: Anisotropic power spectra and correlation functions with pairwise lines-of-sight, arXiv:2102.08384.

[32] M. M. Ivanov, M. Simonović, and M. Zaldarriaga, Cosmological parameters from the BOSS galaxy power spectrum, J. Cosmol. Astropart. Phys. 05 (2020) 042.

[33] N. Hand, Y. Feng, F. Beutler, Y. Li, C. Modi, U. Seljak, and Z. Slepian, NBODYKIT: An open-source, massively parallel toolkit for large-scale structure, Astron. J. **156**, 160 (2018).

[34] NIST, NIST Digital Library of Mathematical Functions, DLMF, https://dlmf.nist.gov/.

[35] A. F. Heavens, R. Jimenez, and O. Lahav, Massive lossless data compression and multiple parameter estimation from galaxy spectra, Mon. Not. R. Astron. Soc. **317**, 965 (2000).

[36] R. Scoccimarro, The bispectrum: From theory to observations, Astrophys. J. **544**, 597 (2000).

[37] J. Alsing and B. Wandelt, Generalized massive optimal data compression, Mon. Not. R. Astron. Soc. **476**, L60 (2018).

[38] S. Alam, M. Ata, S. Bailey, F. Beutler, D. Bizyaev, J. A. Blazek *et al.*, The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Cosmological analysis of the DR12 galaxy sample, Mon. Not. R. Astron. Soc. **470**, 2617 (2017).

[39] D. J. Eisenstein, D. H. Weinberg, E. Agol, H. Aihara, C. A. Prieto, S. F. Anderson *et al.*, SDSS-III: Massive spectroscopic surveys of the distant Universe, the Milky Way, and extra-solar planetary systems, Astron. J. **142**, 72 (2011).

[40] S. A. Rodríguez-Torres, C.-H. Chuang, F. Prada, H. Guo, A. Klypin, P. Behroozi *et al.*, The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Modelling the clustering and halo occupation distribution of BOSS CMASS galaxies in the Final Data Release, Mon. Not. R. Astron. Soc. **460**, 1173 (2016).

[41] F.-S. Kitaura, S. Rodríguez-Torres, C.-H. Chuang, C. Zhao, F. Prada, H. Gil-Marín *et al.*, The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Mock galaxy catalogues for the BOSS Final Data Release, Mon. Not. R. Astron. Soc. **456**, 4156 (2016).

[42] A. Chudaykin, M. M. Ivanov, O. H. E. Philcox, and M. Simonović, Nonlinear perturbation theory extension of the Boltzmann code CLASS, Phys. Rev. D **102**, 063533 (2020).

[43] T. Baldauf, M. Mirbabayi, M. Simonović, and M. Zaldarriaga, LSS constraints with controlled theoretical uncertainties, arXiv:1602.00674.

[44] A. Chudaykin and M. M. Ivanov, Measuring neutrino masses with large-scale structure: Euclid forecast with controlled theoretical error, J. Cosmol. Astropart. Phys. 11 (2019) 034.

[45] T. Brinckmann and J. Lesgourgues, MONTEPYTHON3: Boosted MCMC sampler and other features, Phys. Dark Universe **24**, 100260 (2019).

[46] 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).

[47] E. Sellentin and A. F. Heavens, Quantifying lost information due to covariance matrix estimation in parameter inference, Mon. Not. R. Astron. Soc. **464**, 4658 (2017).

[48] A. Heavens, Statistical techniques in cosmology, arXiv:0906.0664.

[49] D. Babich, Optimal estimation of non-Gaussianity, Phys. Rev. D **72**, 043003 (2005).