

Cosmological inference from galaxy-clustering power spectrum: Gaussianization and covariance decomposition

Mike (Shengbo) Wang ^{1,★}, Will J. Percival,^{2,3,1} Santiago Avila ¹, Robert Crittenden ¹ and Davide Bianchi^{1,4}

¹*Institute of Cosmology and Gravitation, University of Portsmouth, Burnaby Road, Portsmouth PO1 3FX, UK*

²*Department of Physics and Astronomy, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada*

³*Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, Ontario N2L 2Y5, Canada*

⁴*Institut de Ciències del Cosmos, University of Barcelona, IEEC-UB, Martí i Franqués, 1, E-08028 Barcelona, Spain*

Accepted 2019 March 18. Received 2019 March 15; in original form 2018 November 19

ABSTRACT

Likelihood fitting to two-point clustering statistics made from galaxy surveys usually assumes a multivariate normal distribution for the measurements, with justification based on the central limit theorem given the large number of overdensity modes. However, this assumption cannot hold on the largest scales where the number of modes is low. Whilst more accurate distributions have previously been developed in idealized cases, we derive a procedure suitable for analysing measured monopole power spectra with window effects, stochastic shot noise, and the dependence of the covariance matrix on the model being fitted all taken into account. A data transformation is proposed to give an approximately Gaussian likelihood, with a variance–correlation decomposition of the covariance matrix to account for its cosmological dependence. By comparing with the modified- t likelihood derived under the usual normality assumption, we find in numerical tests that our new procedure gives more accurate constraints on the local non-Gaussianity parameter f_{NL} , which is sensitive to the large-scale power. A simple data analysis pipeline is provided for straightforward application of this new approach in preparation for forthcoming large galaxy surveys such as DESI and *Euclid*.

Key words: methods: data analysis – methods: statistical – large-scale structure of Universe.

1 INTRODUCTION

The matter power spectrum, which measures the two-point correlation in Fourier space, is an important statistic for describing the large-scale structure of the Universe. It contains all of the information about a Gaussian random field, which describes cosmic density fluctuations on large scales where non-linearities are negligible. Galaxies are linearly biased tracers of the underlying matter distribution on large scales, and thus measurements of the comoving galaxy-clustering power spectrum can provide a wealth of information about fundamental cosmological parameters. Moreover, as the angular positions and redshifts of galaxies are observed, matching the expected comoving clustering offers a geometrical test of the Universe through the distance–redshift relationship, and measurements of redshift space distortions (RSD) provide a powerful probe of structure growth. Upcoming large-volume surveys, including the Dark Energy Spectroscopic Instrument¹ (DESI Collaboration; Aghamousa et al. 2016) and *Euclid*² (*Euclid* Consortium; Laureijs

et al. 2011), will be able to tightly constrain cosmological models with unprecedented precision, but the accuracy of these constraints relies on performing careful statistical analyses.

The multivariate normal distribution is ubiquitous in modelling cosmological observables thanks to the central limit theorem, and this normality assumption is commonly found in likelihood analyses of power spectrum measurements from galaxy surveys (e.g. Alam et al. 2017; Abbott et al. 2018). Given a theoretical model for the data, the key ingredient of a multivariate normal distribution is its covariance matrix, and in the past many efforts have been devoted to the accurate estimation of covariance matrices subject to limited computational resources.

Unbiased covariance matrix estimates are often made from a set of mock galaxy catalogues synthesized using algorithms ranging from fast but approximate perturbation-theory algorithms to slow yet detailed N -body simulations, or different combinations of those (e.g. Manera et al. 2013; Kitaura et al. 2016; Avila et al. 2018). An overview and comparison of those methods is provided in a series of papers by Lippich et al. (2019), Blot et al. (2019), and Colavincenzo et al. (2018). One could further reduce the computational costs and enhance the precision in this estimation step through various statistical techniques, e.g. the shrinkage method for combining empirical estimates and theoretical models (Pope & Szapudi 2008)

* E-mail: mike.wang@port.ac.uk

¹<https://www.desi.lbl.gov/>.

²<https://www.euclid-ec.org/>.

and the covariance tapering method (Kaufman, Schervish & Nychka 2008; Paz & Sánchez 2015).

However, there are multiple caveats to using an estimated covariance matrix. As noted by Hartlap, Simon & Schneider (2007) and known in statistics (see e.g. Anderson 2003), the inverse of an unbiased covariance matrix estimate is biased with respect to the true precision (inverse covariance) matrix that appears in the likelihood function. Therefore, one needs to include a multiplicative correction that is dependent on the data dimension and the number of catalogue samples used for estimation. Further, the error in covariance estimation must be properly propagated to inferred parameter uncertainties, and achieving desired precision requires a large number of catalogue samples (Dodelson & Schneider 2013; Taylor, Joachimi & Kitching 2013; Percival et al. 2014). Rather than correcting the derived errors, one could adopt the Bayesian principle and treat both the unknown underlying covariance matrix and its estimate as random variables, and marginalize over the former given the latter using Bayes' theorem (Sellentin & Heavens 2016). In addition, as mock catalogues are computationally expensive, they are often produced at fixed fiducial cosmological parameters, whereas in reality the covariance matrix may have cosmological dependence and thus has to vary with the model being tested (see e.g. Eifler, Schneider & Hartlap 2009, in the context of cosmic shear). Failure to account for any of these factors could adversely impact parameter estimation, which should be an important concern to future surveys possessing greater statistical power.

Another fundamental issue to be addressed is the normality assumption itself when the premises of the central limit theorem are not fulfilled. In this scenario, an arbitrarily precise covariance matrix estimate is not sufficient for accurate parameter inference due to significant higher moments in the non-Gaussian likelihood, as demonstrated by Sellentin & Heavens (2018) in the context of weak lensing. This also happens to galaxy-clustering measurements on the largest survey scales, where fewer overdensity modes are available due to the finite survey size; if one is to infer parameters sensitive to these large-scale measurements while assuming a Gaussian likelihood, the parameter estimates are likely to be erroneous.

Indeed, in the past there have been efforts to go beyond the Gaussian likelihood approximation in various contexts (e.g. recently Sellentin, Jaffe & Heavens 2017; Seljak et al. 2017). Motivated by constraints imposed by non-negativity of the power spectrum, Schneider & Hartlap (2009), Keitel & Schneider (2011), and Wilking & Schneider (2013) have found a transformation that improves the normality assumption for the bounded configuration-space correlation function which has a non-normal distribution. Sun, Wang & Zhan (2013) have considered the gamma distribution for the power spectrum multipoles and the lognormal plus Gaussian approximation, and assessed their effects on parameter estimation in comparison with the Gaussian approximations. Kalus, Percival & Samushia (2016) have investigated this problem for the three-dimensional power spectrum by deriving the probability distribution of a single-mode estimator for Gaussian random fields, and comparing it with approximations inspired by similar studies for the cosmic microwave background (e.g. Verde et al. 2003; Percival & Brown 2006; Smith, Challinor & Rocha 2006; Hamimeche & Lewis 2008). However, these previous works are either limited to the univariate or bivariate distribution, or have neglected window functions due to survey geometry and selection effects, which could correlate independent overdensity modes.

In this work, we focus on the power spectrum monopole, following the Feldman–Kaiser–Peacock approach (Feldman, Kaiser &

Peacock 1994, hereafter FKP). Our work can also be trivially extended to the second and fourth power-law moments measured using the Yamamoto estimator (Yamamoto et al. 2006). The two approaches are equivalent for the monopole moment. Alternatives to using these estimators on large scales would be to either directly fit the observed overdensity modes or to perform an analysis based on the quadratic maximum likelihood (QML) method (Tegmark 1997; Tegmark et al. 1998), which would be more optimal given the large-scale window effects in the power spectrum, but computationally more demanding due to evaluation of the large data inverse covariance matrix.

For power-spectrum monopole measurements made using the FKP procedure, we derive the underlying non-normal probability distribution for the *windowed* galaxy-clustering power spectrum in the linear regime with *random* shot noise. The multivariate normal distribution is reinstated through a Gaussianizing transformation that improves data normality, and cosmological dependence of the covariance matrix is fully included by the variance–correlation decomposition. Note that our transformation predicts a new variable whose expectation is not the same transformation of the model. Instead it is simply designed to give a likelihood that is multivariate Gaussian in the data vector. The work is presented as follows.

(i) We review the FKP framework for analysing galaxy-clustering measurements in Section 2, and derive the probability distribution of the windowed power spectrum for a Gaussian random field.

(ii) A Gaussianization scheme is presented in Section 3, which gives a new power-spectrum likelihood approximation with both cosmological dependence and random scatter in the estimated covariance matrix taken into consideration.

(iii) We numerically test our procedure and demonstrate its superiority to the traditional likelihood treatments in Section 4 by performing inference on the local non-Gaussianity parameter f_{NL} and comparing the shape of the new likelihood with that of the true likelihood from simulated data sets.

(iv) A simple pipeline is provided in Section 5 for straightforward application of this method. We discuss in Section 6 the applicability of this new approach and motivate future work.

We also provide a summary of notations (Table 1) used in this work, which may be of particular use to the pipeline detailed in Section 5.

2 POWER SPECTRUM ANALYSIS

2.1 Galaxy-clustering measurements in a finite-sized survey

We assume that the galaxy redshifts measured in a given survey have been converted to comoving distances using a fiducial cosmological model at fixed cosmological parameter(s) $\theta = \theta_f$. This is required to measure the comoving local galaxy density, and hence the comoving power spectrum. Rather than recalculating the power spectrum for each model to be tested, we include the cosmological model dependence of this translation in the model to be fitted to the data.

Let $n_g(\mathbf{r})$ be the observed galaxy number density field and $n_s(\mathbf{r})$ be the number density field in an unclustered random catalogue with expectation $\bar{n}(\mathbf{r}) = \langle n_g(\mathbf{r}) \rangle = \alpha \langle n_s(\mathbf{r}) \rangle$, where α matches the observed and catalogue mean densities. Following FKP, one can define the zero-mean field

$$\hat{F}(\mathbf{r}) = \frac{w(\mathbf{r})}{\sqrt{I}} [n_g(\mathbf{r}) - \alpha n_s(\mathbf{r})] \quad (1)$$

Table 1. Notations used in the proposed final likelihood analysis pipeline.

Symbol	Meaning	Remarks
θ	Cosmological parameter(s)	–
$\cdot^{(f)}$	Quantities at the fiducial cosmology	as a superscript in Section 5
$\cdot^{(d)}$	Measurements/data realizations	as a superscript in Section 5
$N_s \equiv m + 1$	Mock catalogue sample size	–
$P(\mathbf{k})$	Power spectrum model	–
$\tilde{P}_0(k)$	Band power spectrum model	–
$a = 1, \dots, p$	Index for k -bins	–
(R_a, η_a)	Gamma distribution shape–scale parameters	varying with the cosmological model
ν_a	Transformation parameter	fixed at $\nu = 1/3$ for simplicity
\mathbf{Z}	Gaussianized data vector	–
μ_a, σ_a^2	Mean and variance for Gaussianized band power	varying with the cosmological model
$\tilde{\Sigma}$	Estimated covariance matrix for Gaussianized data	rescaled with varying cosmology
$\mathcal{L}(\theta)$	Likelihood function	–

for a weighted mask $w(\mathbf{r})$, where the normalization constant is

$$I = \int d^3\mathbf{r} w(\mathbf{r})^2 \bar{n}(\mathbf{r})^2. \quad (2)$$

The underlying galaxy-clustering power spectrum $P_{\text{true}}(\mathbf{k})$ is equivalent to the Fourier transform of the configuration-space two-point correlation $\xi(\mathbf{r})$ by the Wiener–Khinchin theorem (see e.g. Gabrielli et al. 2005). This power spectrum is convolved (denoted with a tilde) with a window due to the mask,

$$\tilde{P}(\mathbf{k}) = \langle |\hat{F}(\mathbf{k})|^2 \rangle = \int \frac{d^3\mathbf{q}}{(2\pi)^3} |\mathcal{G}(\mathbf{k} - \mathbf{q})|^2 P_{\text{true}}(\mathbf{q}) + P_{\text{shot}}, \quad (3)$$

where the function

$$\mathcal{G}(\mathbf{k}) = \frac{1}{\sqrt{I}} \int d^3\mathbf{r} e^{i\mathbf{k}\cdot\mathbf{r}} w(\mathbf{r}) \bar{n}(\mathbf{r}), \quad (4)$$

and there is an additional scale-independent shot noise power (see Appendix A)

$$P_{\text{shot}} = \frac{1 + \alpha}{V} \int d^3\mathbf{r} \bar{n}(\mathbf{r})^{-1}. \quad (5)$$

The convolved power spectrum may be expanded in the basis of Legendre polynomials $L_\ell(\Delta)$,

$$\tilde{P}(\mathbf{k}) = \sum_{\ell=0}^{\infty} \tilde{P}_\ell(k, \Delta) L_\ell(\Delta), \quad (6)$$

where Δ is cosine of the angle between the wavevector \mathbf{k} and the line of sight. With $L_0 \equiv 1$, the monopole of the convolved power spectrum is a spherical average over the shell V_k at radius k ,

$$\tilde{P}_0(k) = \int_{V_k} \frac{d^3\mathbf{k}}{V_k} \tilde{P}(\mathbf{k}) = \int \frac{d^3\mathbf{q}}{(2\pi)^3} \mathcal{W}(k, \mathbf{q}) P_{\text{true}}(\mathbf{q}) + P_{\text{shot}}, \quad (7)$$

where we have introduced the window function

$$\mathcal{W}(k, \mathbf{q}) = \int_{V_k} \frac{d^3\mathbf{k}}{V_k} |\mathcal{G}(\mathbf{k} - \mathbf{q})|^2. \quad (8)$$

Expanding the power spectrum in multipoles allows the computational demand of the convolution to be reduced (Wilson et al. 2017). For a standard linear power spectrum model that is complete with the first three even power-law moments, the convolution could be rewritten requiring three window matrices whose entries are of the form $\mathcal{W}_\ell(k, \mathbf{q})$. However, in order to keep our equations compact, we retain the more general form and focus on the monopole here.

The windowed power spectrum monopole measured in p bins constitutes the band power data vector

$$\tilde{\mathbf{Y}} \equiv [\hat{P}_0(k_a)]_{a=1}^p, \quad (9)$$

where a hat denotes a realization or an estimator. We can construct another vector

$$\mathbf{Y} \equiv [|\hat{\delta}(\mathbf{q}_i)|^2]_{i=1}^r \quad (10)$$

that estimates the unconvolved power with shot noise at r discrete wavenumbers, where the galaxy overdensity field estimator is

$$\hat{\delta}(\mathbf{r}) = \frac{n_g(\mathbf{r}) - \alpha n_s(\mathbf{r})}{w(\mathbf{r}) \bar{n}(\mathbf{r})} = \frac{\sqrt{I}}{w(\mathbf{r}) \bar{n}(\mathbf{r})} \hat{F}(\mathbf{r}). \quad (11)$$

The discretized analogue to equation (7) for the measured power is then

$$\tilde{\mathbf{Y}} = \mathbf{B}\mathbf{Y}, \quad (12)$$

where the window function is encoded in the mixing matrix

$$B_{ai} = \mathcal{W}(k_a, \mathbf{q}_i), \quad (13)$$

which may be suitably normalized so that

$$\int \frac{d^3\mathbf{q}}{(2\pi)^3} \mathcal{W}(k, \mathbf{q}) = 1 \quad \forall k \quad \Leftrightarrow \quad \sum_i B_{ai} = 1 \quad \forall a. \quad (14)$$

2.2 Distribution of an individual band power measurement

In this subsection we consider the probability distribution of a band power measurement in a single bin at scale k_a . This will be extended to a multivariate distribution including correlations between bins at different scales in Section 3.

2.2.1 Exact hypo-exponential distribution

If $\hat{\delta}(\mathbf{r})$ is directly drawn from a Gaussian random field, then the square amplitude of a single Fourier overdensity mode provides an estimator \hat{P}_i for the unconvolved power $P_i \equiv P(\mathbf{q}_i)$ that is exponentially distributed with the probability density function (PDF)

$$\mathbb{P}[\hat{P}_i] = P_i^{-1} \exp(-\hat{P}_i/P_i), \quad (15)$$

and has expectation $\mathbb{E}[\hat{P}_i] = P_i$ and variance $\text{Var}[\hat{P}_i] = P_i^2$.

However, galaxy formation is a discrete point process, meaning that the overdensity field realization is a *Poisson sample* of the

underlying Gaussian overdensity field (Peebles 1980; Feldman et al. 1994). Consequently, each mode-power estimator Y_i has an additional independent shot noise component ϵ_i ,

$$Y_i = \hat{P}_i + \epsilon_i. \quad (16)$$

In the large galaxy number limit, the shot noise ϵ_i is also exponentially distributed (see Appendix A).

For the bin centred at k_a , the window function mixes mutually independent exponential variables $\{\hat{P}_i\}$ and $\{\epsilon_i\}$ into the band power

$$\tilde{Y}_a = \sum_{i=1}^r B_{ai} Y_i = \sum_{i=1}^r B_{ai} (\hat{P}_i + \epsilon_i). \quad (17)$$

This is an exponential mixture that follows the hypo-exponential distribution (see Appendix B)

$$\mathbb{P}[\tilde{Y}_a; \{\lambda_{i'}\}] = \sum_{i'} \left(\prod_{j' \neq i'} \frac{1}{1 - \lambda_{j'}/\lambda_{i'}} \right) \lambda_{i'}^{-1} \exp(-\lambda_{i'}^{-1} \tilde{Y}_a), \quad (18)$$

with positive scale parameters

$$\lambda_{i'} \in \{B_{ai} P_i, B_{ai} P_{\text{shot}} : i = 1, \dots, r\} \quad (19)$$

being the individual contributions of the unconvolved power and shot noise power to the a th bin. It is understood that a well-defined limit is taken in equation (18) in the case $\lambda_{i'} = \lambda_{j'}$.

We now see that the band power measurement $\tilde{Y}_a = \hat{P}_0(k_a)$ is hypo-exponentially distributed for a Poisson-sampled Gaussian overdensity field. By the central limit theorem, as the number of contributing modes in equation (17) increases, the hypo-exponential variable \tilde{Y}_a converges in distribution to a normal variable. This is the basis for the normality assumption often used in power spectrum analyses: the underlying power as well as shot noise becomes normally distributed, with the latter subtracted as a deterministic quantity to recover the former. However, on the largest scales in a survey where the number of overdensity modes is the fewest, there is clear deviation between the normal distribution and the hypo-exponential distribution.

2.2.2 Gamma distribution approximation

The univariate PDF given by equation (18) is in a difficult form to manipulate as it involves many uncompressed scale parameters $\{\lambda_{i'}\}$. A robust approximation is the exponentially modified gamma distribution with one shape parameter R and two scale parameters (τ, η) (see Golubev 2016), but determining these parameters involves solving a cubic algebraic equation which is cumbersome, making the procedure followed in this paper computationally demanding or even unfeasible. We can adopt a simpler approximation with the gamma distribution

$$\mathbb{P}_\Gamma[\tilde{Y}_a; R, \eta] = \frac{\eta^{-R}}{\Gamma(R)} \tilde{Y}_a^{R-1} e^{-\tilde{Y}_a/\eta}, \quad (20)$$

where by matching the mean and variance of the hypo-exponential distribution we can easily write down the shape and scale parameters

$$R = \mathbb{E}[\tilde{Y}_a]^2 / \text{Var}[\tilde{Y}_a] = \left(\sum_{i'} \lambda_{i'} \right)^2 / \sum_{i'} \lambda_{i'}^2, \quad (21)$$

$$\eta = \text{Var}[\tilde{Y}_a] / \mathbb{E}[\tilde{Y}_a] = \left(\sum_{i'} \lambda_{i'}^2 \right) / \sum_{i'} \lambda_{i'}.$$

In place of the mean and variance parameters of a normal distribution, these two gamma distribution parameters determine the *non-normal* distribution of the band power measurement in each bin, and have a natural interpretation in our context: the shape R is the effective number of independent overdensity modes contributing to the bin under window function mixing, and the scale η is the effective convolved power in that bin. In the limiting case that there is only a single non-vanishing mode (e.g. pure shot noise), the gamma distribution coincides exactly with the hypo-exponential distribution, both of which reduce to the exponential distribution.

In principle, the shape and scale parameters (R, η) could be calculated for each bin using analytical expressions of the band power expectation $\mathbb{E}[\tilde{Y}_a]$ and variance $\text{Var}[\tilde{Y}_a]$, with knowledge of the full window mixing matrix \mathbf{B} and all observed overdensity modes $\{\hat{\delta}(\mathbf{q}_i)\}$. Practically, these quantities are not always available, as fast window convolution is now often performed in configuration space with ever larger surveys and higher resolution requirements, and the windowed power spectrum is computed via a Hankel transform of the convolved two-point correlation multipoles $\tilde{\xi}_\ell(\mathbf{r})$ (Beutler et al. 2017; Wilson et al. 2017). Under these circumstances, the variance of the measured band power $\text{Var}[\tilde{Y}_a]$ may be estimated from mock catalogues, provided the error in this estimation is subdominant compared to other sources of uncertainty. However, it is worth pointing out ongoing efforts in developing accurate analytic covariance matrices that could possibly evade the problem of covariance estimation altogether, e.g. Li et al. (2019).

2.2.3 Normal distribution assumption

For the sake of completeness and for reference, we write down the univariate normal distribution for the band power \tilde{Y}_a in terms of the shape–scale parameters (R, η) for the a th bin,

$$\mathbb{P}_\text{N}[\tilde{Y}_a; R\eta, R\eta^2] = \frac{1}{\sqrt{2\pi R\eta^2}} \exp\left[-\frac{(\tilde{Y}_a - R\eta)^2}{2R\eta^2}\right], \quad (22)$$

so that its expectation and variance match those of the exact hypo-exponential and approximate gamma distributions.

3 GAUSSIANIZATION AND VARIANCE-CORRELATION DECOMPOSITION

A multivariate PDF transformation for $\mathbf{Y} \mapsto \tilde{\mathbf{Y}}$ is captured by the Jacobian factor $\mathcal{J} = \det(\mathbf{B}\mathbf{B}^\top)$, where the distribution of \mathbf{Y} is the product of independent exponential components. The window mixing matrix $\mathbf{B} \in \mathbb{R}_+^{p \times r}$ is non-square ($p < r$), and expressing the transformed PDF explicitly in terms of the band power vector $\tilde{\mathbf{Y}}$, whose components are correlated, requires the inversion of \mathbf{B} . One could introduce $(r - p)$ ‘helper components’ in $\tilde{\mathbf{Y}}$ to pad \mathbf{B} into a square matrix, and eventually marginalize out these additional random variables. However, the linear transformation induced by the padded square matrix will generally map the domain of the random vector \mathbf{Y} from \mathbb{R}_+^r to a different domain in \mathbb{R}^r , making marginalization difficult and susceptible to the ‘curse’ of dimensionality, which is likely as $r \gg p$ for massive data compression.

Instead of trying to determine a full multivariate transformation for the windowed band power, we subscribe to a *component-wise* Gaussianization strategy, and reinstate the multivariate normality assumption in the *Gaussianized* data vector $\mathbf{Z} \leftrightarrow \tilde{\mathbf{Y}}$. The reasoning behind this is two fold. First, each component of the random vector

is now certainly univariate normal, as should be the case for a bona fide multivariate normal distribution. Secondly, if the cross-bin correlation is weak and the covariance matrix has a narrow-band structure, the components become approximately independent and univariate Gaussianization is equivalent to multivariate Gaussianization; this can be achieved with a suitable binning choice where the bin width is chosen to be greater than the known window correlation length. Although sophisticated full Gaussianization schemes exist (e.g. Laparra, Camps-Valls & Malo 2011), they are often iteratively applied based on empirical data and may not be suitable for forward modelling of theoretical models. We thus leave more advanced multivariate Gaussianization methods for future investigations.

We propose a simple Gaussianization scheme using the Box–Cox transformation. This scheme has already been applied in cosmology to deal with non-Gaussian parameter spaces (Joachimi & Taylor 2011; Schuhmann, Joachimi & Peiris 2016), but our context and implementation differ from these studies. An alternative scheme directly transforming a non-normal distribution into the standard normal distribution exists by matching the cumulative distribution function; however, computationally costly numerical integration is necessary for calculating transformed moments, so we relegate this scheme to Appendix C for reference. As we shall see later, the Box–Cox transformation suppresses higher order moments to achieve *approximate* Gaussianization and is sufficiently accurate for our purposes. To perform the transformation, we use the fiducial shape–scale parameters (R_f, η_f) which are determined by the power spectrum model $P_f(k)$ at fixed cosmological parameter(s) $\theta = \theta_f$.

3.1 Box–Cox transformation

We define the Box–Cox transformation (Box & Cox 1964) for each component of \mathbf{Y} (index suppressed for brevity) by

$$Z = \tilde{Y}^\nu, \quad \nu > 0, \quad (23)$$

where positive ν is chosen to ensure regularity. The transformed PDF in Z is now

$$\mathbb{P}[Z(\tilde{Y}); R, \eta] = |\mathcal{J}_Z(\tilde{Y})|^{-1} \frac{\eta^{-R}}{\Gamma(R)} \tilde{Y}^{R-1} e^{-\tilde{Y}/\eta}, \quad (24)$$

where $\mathcal{J}_Z(\tilde{Y}) = \nu \tilde{Y}^{\nu-1}$ is the Jacobian for the transformation $\tilde{Y} \mapsto Z$. The transformed K th moment is given by

$$\mathbb{E}[Z^K] = \frac{\Gamma(R + K\nu)}{\Gamma(R)} \eta^{K\nu}, \quad (25)$$

and we can write down the Gaussianized mean and variance

$$\begin{aligned} \mu(R, \eta) &\equiv \mathbb{E}[Z] = \frac{\Gamma(R + \nu)}{\Gamma(R)} \eta^\nu, \\ \sigma^2(R, \eta) &\equiv \text{Var}[Z] = \frac{\Gamma(R + 2\nu)\Gamma(R) - \Gamma(R + \nu)^2}{\Gamma(R)^2} \eta^{2\nu}. \end{aligned} \quad (26)$$

To determine the transformation parameter ν , we demand the third central moment vanish for the *fiducial* model parameters θ_f ,

$$\begin{aligned} 0 &= \mathbb{E}[Z^3] - 3\mathbb{E}[Z^2]\mathbb{E}[Z] + 2\mathbb{E}[Z]^3 \\ &= \left\{ \frac{\Gamma(R_f + 3\nu)}{\Gamma(R_f)} - 3 \frac{\Gamma(R_f + 2\nu)}{\Gamma(R_f)} \frac{\Gamma(R_f + \nu)}{\Gamma(R_f)} \right. \\ &\quad \left. + 2 \left[\frac{\Gamma(R_f + \nu)}{\Gamma(R_f)} \right]^3 \right\} \eta^{3\nu}. \end{aligned} \quad (27)$$

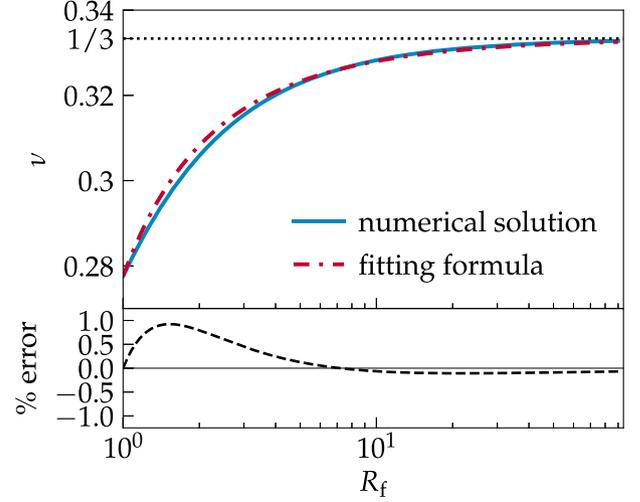


Figure 1. The dependence of the optimal Box–Cox transformation parameter, ν , on the fiducial gamma-distribution shape parameter R_f . *Top panel:* comparison of the numerical solution to equation (27) (solid line) with the empirical fitting formula given by equation (30) (dash-dotted line) for determining the optimal transformation parameter ν as a function of the fiducial shape parameter R_f . *Bottom panel:* the percentage error of the fitting formula from the numerical solution (dashed line) is at the sub per cent level.

The dependence of ν on R_f required to satisfy this constraint is shown in Fig. 1 as a numerical solution (dashed blue line). The observed asymptotic behaviour can be understood by considering the expansion of gamma function ratios (Burić & Elezović 2012)

$$\begin{aligned} \frac{\Gamma(A+b)}{\Gamma(A)} &\sim A^b \left[1 + \frac{(b-1)b}{2} A^{-1} \right. \\ &\quad \left. + \frac{(3b-1)(b-2)(b-1)b}{24} A^{-2} + \dots \right], \end{aligned} \quad (28)$$

so as $R_f \rightarrow \infty$ with $\eta R_f < \infty$ (finite mean), equation (27) becomes

$$0 \simeq R_f^{-2} \nu^3 (3\nu - 1). \quad (29)$$

The non-trivial solution is $\nu = 1/3$. The precise value of ν matters less for increasing R_f owing to the suppression factor R_f^{-2} , a manifestation of asymptotic normality in the limit $R_f \gg 1$.

An empirical fitting formula for the solution to equation (27) is given by

$$\nu \approx \frac{1}{3} + 0.042 \left[1 - \exp(-0.85/R_f^{0.85}) \right]. \quad (30)$$

Comparing this fit to the true solution in Fig. 1 shows that this fitting formula performs well, being accurate to sub per cent levels. As we shall see in Section 4, in fact simply assuming the fixed value $\nu = 1/3$ works very well in realistic situations; our Gaussianization scheme is robust to variation in ν with the fiducial parameter R_f , i.e. the choice of fiducial cosmology.

3.2 Covariance treatment

In this subsection we discuss two general treatments of estimated covariance matrices which could be in the band power data $\tilde{\mathbf{Y}}$ or in the Gaussianized data \mathbf{Z} , although our treatments are applied to the Gaussianized data in Section 5.

3.2.1 Cosmological parameter dependence

Now that we have a Gaussianizing transformation, the model dependence of the covariance matrix still needs to be considered. This sub-subsection shows how the parameter dependence can be included in the covariance matrix estimate analytically. In Section 2, we have ignored the integral constraint in the modelling of the power spectrum, which biases the power spectrum measured due to estimation of the mean galaxy number density from the galaxy catalogue itself; since this offset in the measured power can be subtracted (see e.g. Peacock & Nicholson 1991; Beutler et al. 2014), we do not consider its contribution here in this work.

A generic covariance matrix Σ may be decomposed into the diagonal matrix $\Lambda = (\text{diag}\Sigma)^{1/2}$ and the correlation matrix \mathbf{C} ,

$$\Sigma = \Lambda \mathbf{C} \Lambda, \quad (31)$$

where Λ^2 is the diagonal matrix of the variances. For the band power spectrum on large scales, *whether Gaussianized or not*, the off-diagonal correlation in \mathbf{C} is solely induced by the window function as encoded in the mixing matrix \mathbf{B} . Whilst this mixing matrix does depend on the fiducial cosmological model through the distance–redshift relation, crucially it does not depend on the model being tested through the power spectrum.³ This insight makes the variance–correlation decomposition particularly useful, for the decomposition into Λ and \mathbf{C} is precisely the separation of any cosmology dependence from cosmology independence in the covariance matrix Σ . Therefore one may obtain a covariance matrix estimate $\hat{\Sigma}_f = \Lambda_f \mathbf{C} \Lambda_f$ from mock catalogues produced at a fixed cosmology $\theta = \theta_f$ with the fiducial power spectrum model $P_f(k)$, and calibrate this estimate by rescaling with the diagonal variances to allow for varying cosmology,

$$\hat{\Sigma}(\theta) = \Lambda(\theta) \Lambda_f^{-1} \hat{\Sigma}_f \Lambda_f^{-1} \Lambda(\theta). \quad (32)$$

For instance, for the Gaussianized band power \mathbf{Z} at cosmological parameter(s) θ , the entries in the diagonal variance matrix are

$$\text{diag}\Lambda^2(\theta) = \text{Var}[\mathbf{Z}] = [\sigma_a^2(R_a, \eta_a)]_{a=1}^p \quad (33)$$

as given by equation (26), where the gamma shape–scale parameters (R_a, η_a) for each bin depend on cosmological parameter(s) θ through the power (see equations 19 and 21).

3.2.2 Covariance matrix estimates as random variables

Covariance estimation from mock catalogues gives an inherently random quantity. Instead of directly substituting the estimated covariance in a probability distribution, a more principled approach is to marginalize out the unknown underlying covariance matrix using Bayes’ theorem (Sellentin & Heavens 2016, hereafter SH). Let $\hat{\Sigma}_f$ be an unbiased covariance matrix estimate calculated from $N_s \equiv m + 1$ samples of the data vector from mock catalogues at the fiducial cosmology. Using the uninformative Jeffreys prior on the unknown true covariance Σ_f ,

$$\Pi[\Sigma_f] \propto |\Sigma_f|^{-(p+1)/2}, \quad (34)$$

³A caveat is that, in principle, there could be redshift-space effects that render the overall window function dependent on the underlying cosmology, e.g. a survey boundary defined by a maximum redshift.

and the fact that an empirical covariance matrix estimate $\hat{\Sigma}_f$ has the Wishart distribution conditional on Σ_f ,

$$\mathbb{P}[\hat{\Sigma}_f | \Sigma_f] \propto \frac{|\hat{\Sigma}_f|^{(m-p-1)/2}}{|\Sigma_f|^{m/2}} \exp\left[-\frac{m}{2} \text{tr}\left(\Sigma_f^{-1} \hat{\Sigma}_f\right)\right], \quad (35)$$

one could show that the posterior distribution of $\Sigma_f | \hat{\Sigma}_f$ is inverse Wishart with the PDF

$$\mathbb{P}[\Sigma_f | \hat{\Sigma}_f] \propto \frac{|\hat{\Sigma}_f|^{m/2}}{|\Sigma_f|^{(m+p+1)/2}} \exp\left[-\frac{m}{2} \text{tr}\left(\hat{\Sigma}_f \Sigma_f^{-1}\right)\right]. \quad (36)$$

It is this distribution that one needs to marginalize over to replace the unknown true covariance Σ_f with its estimate $\hat{\Sigma}_f$. The same derivation follows exactly for the rescaled covariance estimate $\hat{\Sigma}(\theta)$, since our decomposition does not affect the SH marginalization procedure (see Appendix D).

In addition to the covariance matrix estimate for the Gaussianized data \mathbf{Z} , if one estimates the gamma distribution parameters $\{R_a\}$ and $\{\eta_a\}$ (see equation 21) from the empirical covariance matrix estimate of the band power data $\tilde{\mathbf{Y}}$, then these parameters should also be considered as random variables; in particular, the shape adopted for the likelihood is itself an estimated quantity, and this could in principle also bias the recovered likelihood. In Section 4, we will see that in practice this is not an issue: for estimated covariance matrices with significant uncertainties, the effect of the covariance estimation (allowed for by the SH marginalization procedure) itself dominates. For estimated covariance matrices with low noise, the two effects can become comparable, but in this situation the size of both effects is small.

3.3 Likelihood form

When the mean vector μ for the Gaussian data vector \mathbf{Z} and its covariance matrix Σ are known exactly, the multivariate normal PDF simply reads

$$\mathbb{P}_N[\mathbf{Z}; \mu, \Sigma] = |\Sigma|^{-1/2} \exp\left[-\frac{1}{2} \chi^2(\mathbf{Z}; \mu, \Sigma)\right], \quad (37)$$

where the quantity

$$\chi^2(\mathbf{Z}; \mu, \Sigma) \equiv (\mathbf{Z} - \mu)^\top \Sigma^{-1} (\mathbf{Z} - \mu). \quad (38)$$

This is the multivariate version of equation (22) but now for the *Gaussianized* band power \mathbf{Z} . The SH procedure replaces the underlying covariance Σ with an estimate $\hat{\Sigma}$, and changes this to a modified *t*-distribution (see Appendix D)

$$\mathbb{P}_t[\mathbf{Z}; \mu, \hat{\Sigma}, m] = c_p |\hat{\Sigma}|^{-1/2} \left[1 + \frac{\chi^2(\mathbf{Z}; \mu, \hat{\Sigma})}{m}\right]^{-(m+1)/2}, \quad (39)$$

where the normalization constant is

$$c_p = (m\pi)^{-p/2} \frac{\Gamma[(m+1)/2]}{\Gamma[(m-p+1)/2]}. \quad (40)$$

The PDF given in equation (39), when regarded as a function of cosmological parameter(s) θ through the dependence of μ and $\hat{\Sigma}$ on the power spectrum model as functions of $\{(R_a, \eta_a)\}$,

$$\mathcal{L}(\theta; \mathbf{Z}, \hat{\Sigma}_f, m) = \mathbb{P}_t[\mathbf{Z}; \mu(\theta), \hat{\Sigma}(\theta), m], \quad (41)$$

is a key result of this work.

In the next section, we shall test different aspects of our procedure used to derive the new likelihood, using Monte Carlo simulations matched to the specifications of future survey data. This leads us to formulating a simple pipeline for likelihood analysis of galaxy surveys, which we present in Section 5.

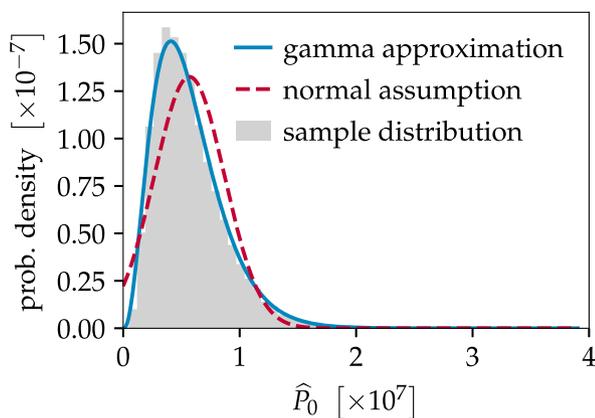


Figure 2. Distribution of the band power spectrum in the lowest- k bin, centred at $k \approx 0.0024 h \text{ Mpc}^{-1}$. The effective number of independent modes in this bin is $R \approx 4$. We compare the exact hypo-exponential distribution (filled region) sampled from 40 000 realizations, the gamma distribution approximation (solid line) and the normal distribution assumption with the same mean and variance (dashed line) for the band power measurement.

4 TESTING WITH NUMERICAL SIMULATIONS

In order to test our proposed likelihood form, we create Monte Carlo simulations of data vectors by generating exponentially distributed overdensity mode power and shot noise, and then convolve with a chosen survey window before extracting the measured power spectra. Given the survey volume of DESI and *Euclid*, we consider scales $k = 1.58 \times 10^{-3} - 1.58 \times 10^{-2} h \text{ Mpc}^{-1}$ covering an order of magnitude, and select overdensity wavenumbers \mathbf{q} using an inverse-volume distribution $\mathbb{P}[\mathbf{q}] \propto |\mathbf{q}|^3$. The chosen window function has a Gaussian shape with full width at half-maximum (FWHM) $1.88 \times 10^{-3} h \text{ Mpc}^{-1}$. We divide the scales into $p = 9$ bins so that the cross-bin correlation is weak under this window function. Given an input cosmology, the power spectrum model is specified as follows:

(i) The underlying galaxy power spectrum is calculated using the fitting formula for the matter transfer function $T(k)$ by Eisenstein & Hu (1998), with the large-scale galaxy linear bias fixed at $b_0 = 1.87$ and other cosmological parameters set to *Planck* 2018 values (Aghanim et al. 2018);

(ii) The shot noise power is calculated using equation (5) with a pessimistic number density $\bar{n} = 5 \times 10^{-4} h^3 \text{ Mpc}^{-3}$. The number densities predicted for DESI and *Euclid* are higher (e.g. Duffy 2014), so our analysis is conservative with respect to the effect of shot noise on large scales.

4.1 Testing distribution normality

In Fig. 2, we compare the sampled hypo-exponential distribution (equation 18), the gamma distribution approximation (equation 20) and the normal distribution assumption (equation 22) with the same mean and variance for the band power in the bin centred at $k \approx 0.0024 h \text{ Mpc}^{-1}$, which has an effective number of independent modes $R \approx 4$. The assumed normal distribution has a peak shifted from the underlying hypo-exponential distribution; on the other hand, the gamma distribution is a good approximation that matches both the peak and the tails well.

One may wish to quantify the improvement in *multivariate* normality our component-wise Gaussianization could bring to the

band power data vector. A key defining property of a multivariate normal variable \mathbf{X} is that any projection $\mathbf{X} \mapsto \mathbf{t}^\top \mathbf{X} \in \mathbb{R}$, for some vector \mathbf{t} , gives a univariate normal variable. Hence as a simple multivariate normality test, given a set of samples for \mathbf{X} , one could randomly choose some directions \mathbf{t} and perform univariate normality tests on the projected samples $\{\mathbf{t}^\top \mathbf{x}_i\}$ (see e.g. Shao & Zhou 2010).

We perform the D’Agostino–Pearson normality test (D’Agostino & Pearson 1973) on 10 000 random projections of 40 000 samples of the band power vector $\tilde{\mathbf{Y}}$, which returns the p -value that characterizes how extreme the sample realizations are under the null hypothesis that the underlying distribution were indeed normal. It must be emphasized here that the p -value itself is not a meaningful indicator of normality, as it varies depending on the sample size; rather it is the comparison of the p -values with the same sample that signifies relative departure from normality. We find $p = 0.01$ for $\tilde{\mathbf{Y}}$ without Gaussianization; with Gaussianization $\tilde{\mathbf{Y}} \mapsto \mathbf{Z}$, however, we find improved $p = 0.08$ given these samples.

4.2 Testing covariance treatment

To test the variance–correlation decomposition proposed in Section 3.2, we generate one set of 40 000 band power data realizations with the Hubble parameter set to $H_0 = 67.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$, and an additional set of 40 000 realizations generated with $H_0 = 73.2 \text{ km s}^{-1} \text{ Mpc}^{-1}$. The former set gives a sampled ‘true’ covariance matrix $\hat{\Sigma}$, and the latter gives a ‘fiducial’ covariance estimate $\hat{\Sigma}_f$ which is then rescaled using equation (32) to match the ‘true’ cosmology. For both band power $\tilde{\mathbf{Y}}$ without Gaussianization and Gaussianized band power \mathbf{Z} , Fig. 3 shows that the differences between the directly sampled ‘true’ covariance matrices and the rescaled covariance estimates are small; this validates the decomposition as a means to include cosmological dependence of the covariance matrix.

Since covariance marginalization and decomposition do not mutually affect each other, we do not test the effect of the SH procedure (the modified- t likelihood) separately here but point the reader to Sellentin & Heavens (2016) and Sellentin & Heavens (2018) for reference.

4.3 Testing likelihoods for parameter inference

The ultimate goal of power-spectrum likelihood analysis is to constrain cosmological parameters, so the primary aim of our numerical simulations is to test our new likelihood function after Gaussianization and covariance rescaling.

To summarize, we have proposed the following steps for deriving the new likelihood function (41), which we now tweak to isolate their effects:

(i) *Gaussianization* – the data vector can remain un-Gaussianized $\tilde{\mathbf{Y}}$, Gaussianized $\mathbf{Z}_{1/3}$ with a fixed parameter $\nu = 1/3$, or Gaussianized \mathbf{Z}_ν with a fitted transformation parameter ν given by the formula in equation (30);

(ii) *Covariance rescaling* – the fiducial covariance matrix estimate is either fixed at $\hat{\Sigma}_f$ when calculating the likelihood, or rescaled to $\hat{\Sigma}(\theta)$ using equation (32) to account for parameter dependence;

(iii) *Covariance marginalization* – the Hartlap-debiased precision matrix estimate (Hartlap et al. 2007) can either be substituted directly into the Gaussian likelihood $\mathcal{L}_G(\theta)$ (see equation 37), or the

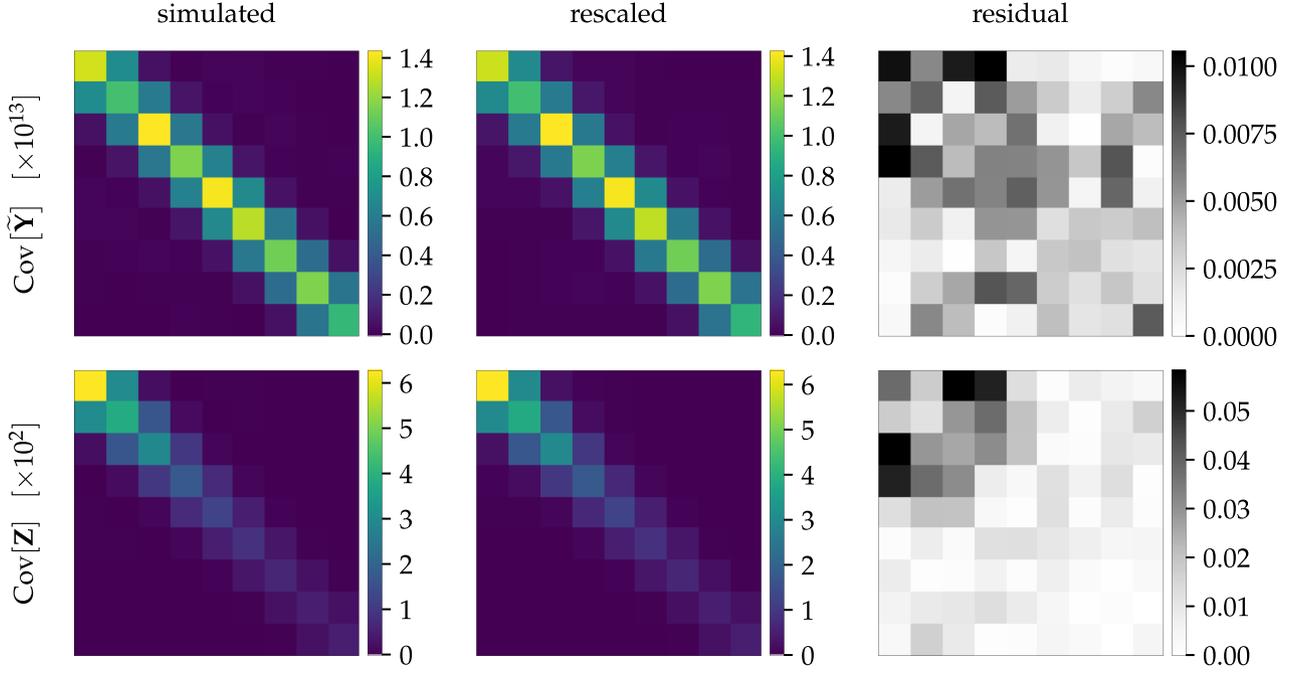


Figure 3. Validation of the variance–correlation decomposition to account for the parameter dependence of the covariance matrix by comparing directly sampled covariance matrices with rescaled covariance matrix estimates. *Left-hand panel:* covariance matrices sampled at a ‘true’ cosmology input ($H_0 = 67.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$); *middle panel:* covariance estimates at the ‘fiducial’ cosmology ($H_0 = 73.2 \text{ km s}^{-1} \text{ Mpc}^{-1}$) rescaled using equation (32) to match the ‘true’ cosmology; *right-hand panel:* residuals (element-wise absolute differences) between the covariance matrices in the left-hand and middle panels. The *top row* and the *bottom row* are produced from 40 000 realizations of the band power $\tilde{\mathbf{Y}}$ and the Gaussianized band power \mathbf{Z} , respectively.

SH marginalization procedure can be used to give the modified- t likelihood $\mathcal{L}_t(\theta)$ (see equation 39).

Different combinations of these choices give the likelihoods tabulated in Table 2. All these likelihoods can be compared with the *true likelihood* constructed from the exponentially distributed modes and shot noise power *prior to convolution*,

$$\mathcal{L}_{\text{true}}(\theta; \mathbf{Y}) = \prod_{i=1}^r \frac{e^{-Y_i/P_i} - e^{-Y_i/P_{\text{shot}}}}{P_i - P_{\text{shot}}}, \quad (42)$$

which is inaccessible in a realistic survey in the presence of the window function.

Since our methods mostly affect measurements on the largest survey scales, the local non-Gaussianity parameter f_{NL} , which is sensitive to the large-scale power, is a well-motivated test parameter (Sun et al. 2013; Kalus et al. 2016). f_{NL} enters the galaxy power spectrum by modifying the constant linear galaxy bias on large scales (Dalal et al. 2008; Matarrese & Verde 2008; Slosar et al. 2008),

$$b_0 \mapsto b_0 + f_{\text{NL}} A(k)(b_0 - 1), \quad (43)$$

which introduces scale dependence via

$$A(k) = \frac{3\Omega_m \delta_c}{k^2 T(k)} \left(\frac{H_0}{c} \right)^2. \quad (44)$$

Here c is the speed of light, Ω_m is the matter density parameter, and $\delta_c \approx 1.686$ is the spherical collapse critical overdensity today. Henceforth we will identify θ with $\theta = f_{\text{NL}}$. We emphasize that the choice of f_{NL} as a test parameter is entirely based on likelihood considerations; our work does not serve as a stringent constraint on primordial non-Gaussianity. To leading order in f_{NL} , which is small

as constrained by *Planck* 2015 results (Ade et al. 2016), we continue to treat galaxy overdensity as a Gaussian random field, albeit with an amplitude modulated by f_{NL} ; this will eventually break down on near the Hubble horizon scale (see e.g. Tellarini et al. 2015).

To properly examine the *ensemble* behaviour of the likelihoods listed in Table 2 with different treatments, we now produce 250 000 data realizations at some ‘true’ input cosmology, and a fixed set of $N_s = 1000$ mock catalogues simulated at the fiducial cosmology $f_{\text{NL}} = 0$. The latter provides a covariance matrix estimate for both the band power and the Gaussianized data. The prior range for f_{NL} is set to be $[-250, 250]$ and scanned through with a resolution of $\Delta f_{\text{NL}} = 0.05$.⁴

To get an initial intuition, we compare in Fig. 4 the true likelihood $\mathcal{L}_{\text{true}}$, the modified- t likelihood $\mathcal{L}_{t,\text{fc}}[\tilde{\mathbf{Y}}]$ without Gaussianization and covariance rescaling, and the new likelihood $\mathcal{L}_{t,\text{rc}}[\mathbf{Z}_\nu]$ derived using our full procedure with fitted transformation parameter ν , all averaged over data realizations produced at $f_{\text{NL}} = 0$. It is clear that our methods produce a superior likelihood approximation to the true likelihood.

4.3.1 Point estimation comparison

For different true f_{NL} parameter inputs and the same fiducial cosmology at $f_{\text{NL}} = 0$, we compare both frequentists’ and Bayes estimators calculated from the likelihoods $\mathcal{L}_{\text{true}}$, $\mathcal{L}_{t,\text{rc}}[\tilde{\mathbf{Y}}]$, $\mathcal{L}_{t,\text{rc}}[\mathbf{Z}_\nu]$,

⁴We have chosen a much wider prior range than the *Planck* 2015 constraint (without polarization) $f_{\text{NL}} = 2.5 \pm 5.7$ (Ade et al. 2016) to demonstrate the robustness of our likelihood treatments to a large range of power spectrum amplitudes.

Table 2. List of likelihoods with different combinations of Gaussianization and covariance rescaling treatments, as well as different functional forms.

Data variable	No Gaussianization		With Gaussianization		
	Modified- t		Fixed $\nu = 1/3$		
Functional form	Modified- t		Gaussian	Modified- t	Fitted ν
Rescaled covariance estimate (rc)	$\mathcal{L}_{t,rc}[\tilde{\mathbf{Y}}]$		$\mathcal{L}_{G,rc}[\mathbf{Z}_{1/3}]$	$\mathcal{L}_{t,rc}[\mathbf{Z}_{1/3}]$	$\mathcal{L}_{t,rc}[\mathbf{Z}_\nu]$
Fixed covariance estimate (fc)	$\mathcal{L}_{t,fc}[\tilde{\mathbf{Y}}]$		–	$\mathcal{L}_{t,fc}[\mathbf{Z}_{1/3}]$	$\mathcal{L}_{t,fc}[\mathbf{Z}_\nu]$

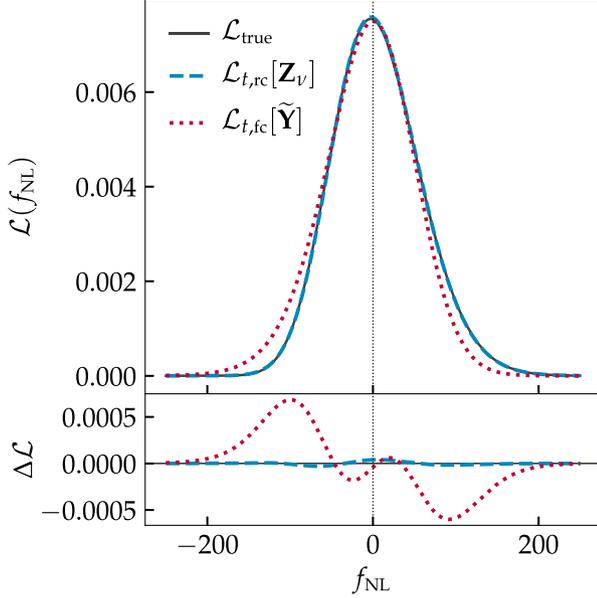

Figure 4. A direct comparison between the true likelihood $\mathcal{L}_{\text{true}}$ (solid line), the modified- t likelihood $\mathcal{L}_{t,rc}[\mathbf{Z}_\nu]$ (dashed line) with Gaussianization and covariance rescaling, and the modified- t likelihood $\mathcal{L}_{t,fc}[\tilde{\mathbf{Y}}]$ (dotted line) without Gaussianization and covariance rescaling, all averaged over data realizations produced at $f_{\text{NL}} = 0$. The bottom panel shows the differences of the likelihoods from $\mathcal{L}_{\text{true}}$.

Table 3. Maximum likelihood estimates of f_{NL} from likelihoods $\mathcal{L}_{\text{true}}$, $\mathcal{L}_{t,rc}[\tilde{\mathbf{Y}}]$, $\mathcal{L}_{t,rc}[\mathbf{Z}_\nu]$, $\mathcal{L}_{t,fc}[\tilde{\mathbf{Y}}]$, and $\mathcal{L}_{t,fc}[\mathbf{Z}_\nu]$ (see Table 2 for definitions), averaged over 250 000 data realizations with different true f_{NL} inputs and the same fiducial cosmology at $f_{\text{NL}} = 0$. The 1σ error bounds are given as the estimated standard deviations.

Input	Maximum likelihood estimates				
	$\mathcal{L}_{\text{true}}$	$\mathcal{L}_{t,rc}[\tilde{\mathbf{Y}}]$	$\mathcal{L}_{t,rc}[\mathbf{Z}_\nu]$	$\mathcal{L}_{t,fc}[\tilde{\mathbf{Y}}]$	$\mathcal{L}_{t,fc}[\mathbf{Z}_\nu]$
50	49 ± 42	42 ± 44	48 ± 42	49 ± 44	51 ± 42
10	9 ± 39	2 ± 41	8 ± 40	9 ± 39	11 ± 39
0	-1 ± 39	-8 ± 41	-2 ± 39	-2 ± 38	1 ± 38
-10	-12 ± 38	-18 ± 40	-12 ± 38	-12 ± 38	-10 ± 37
-50	-52 ± 35	-58 ± 36	-52 ± 35	-52 ± 37	-50 ± 34

$\mathcal{L}_{t,fc}[\tilde{\mathbf{Y}}]$ and $\mathcal{L}_{t,fc}[\mathbf{Z}_\nu]$ (see Table 2 for definitions). The results have been marginalized over our data realizations to assess their ensemble behaviour.

Maximum likelihood estimator – This is a frequentists’ estimator, given by

$$\hat{\theta} = \arg \max \mathcal{L}(\theta). \quad (45)$$

The estimates are compared in Table 3. Their uncertainties are the standard deviations estimated from the ensemble of data realizations.

Table 4. Posterior median estimates of f_{NL} from likelihoods $\mathcal{L}_{\text{true}}$, $\mathcal{L}_{t,rc}[\tilde{\mathbf{Y}}]$, $\mathcal{L}_{t,rc}[\mathbf{Z}_\nu]$, $\mathcal{L}_{t,fc}[\tilde{\mathbf{Y}}]$, and $\mathcal{L}_{t,fc}[\mathbf{Z}_\nu]$ (see Table 2 for definitions), averaged over 250 000 data realizations with different true f_{NL} inputs and the same fiducial cosmology at $f_{\text{NL}} = 0$. The 1σ error bounds are quoted as the equal-tailed 68.3 per cent credible interval.

Input	Posterior median estimates				
	$\mathcal{L}_{\text{true}}$	$\mathcal{L}_{t,rc}[\tilde{\mathbf{Y}}]$	$\mathcal{L}_{t,rc}[\mathbf{Z}_\nu]$	$\mathcal{L}_{t,fc}[\tilde{\mathbf{Y}}]$	$\mathcal{L}_{t,fc}[\mathbf{Z}_\nu]$
50	53 ⁺⁴³ ₋₃₉	48 ⁺⁴⁰ ₋₃₅	53 ⁺⁴³ ₋₃₈	46 ⁺³³ ₋₃₅	51 ⁺³⁹ ₋₃₈
10	13 ⁺⁴¹ ₋₃₆	8 ⁺³⁸ ₋₃₂	13 ⁺⁴⁰ ₋₃₅	6 ⁺³⁶ ₋₃₉	11 ⁺³⁸ ₋₃₈
0	3 ⁺⁴⁰ ₋₃₆	-2 ⁺³⁸ ₋₃₂	3 ⁺⁴⁰ ₋₃₅	-5 ⁺³⁷ ₋₄₀	1 ⁺³⁸ ₋₃₇
-10	-7 ⁺³⁹ ₋₃₅	-12 ⁺³⁷ ₋₃₁	-7 ⁺³⁹ ₋₃₄	-15 ⁺³⁷ ₋₄₁	-9 ⁺³⁸ ₋₃₇
-50	-46 ⁺³⁶ ₋₃₁	-52 ⁺³⁵ ₋₂₈	-46 ⁺³⁶ ₋₃₀	-56 ⁺⁴¹ ₋₄₄	-49 ⁺³⁷ ₋₃₇

Posterior median estimator – With flat priors, the posterior $\mathcal{P}[\theta | \mathbf{X}]$ on the cosmological parameter(s) θ given any observations \mathbf{X} is simply the likelihood $\mathcal{L}(\theta; \mathbf{X})$ suitably normalized. Common Monte Carlo analyses usually return the posterior median or mean as the best estimate (see e.g. Trotta 2008; Hogg & Foreman-Mackey 2018). Here we choose the absolute loss function (Berger 1985)

$$\text{loss}(a, \theta) = |a - \theta| \quad (46)$$

and minimize its expectation to obtain the Bayes estimator that is the posterior median

$$\hat{\theta} = \arg \min_a \mathbb{E}_{\theta|\mathbf{X}}[\ell(a, \theta)]. \quad (47)$$

The associated 1σ uncertainties can be quoted as the equal-tailed 68.3 per cent Bayesian credible interval. The results are displayed in Table 4.

It is evident that overall the new likelihood $\mathcal{L}_{t,rc}[\mathbf{Z}_\nu]$ performs the best in producing closer best estimates as well as error bounds to those from the true likelihood, whether we perform frequentists’ estimation or Bayesian inference. We have found that Gaussianization with a fixed transformation parameter $\nu = 1/3$, i.e. $\mathcal{L}_{t,rc}[\mathbf{Z}_{1/3}]$, gives similar results to Gaussianization with ν fitted by the formula in equation (30); likewise, assuming a wrong fiducial cosmology for Gaussianization, even when it is significantly deviant from the true cosmology, has negligible impact on recovered parameters. This demonstrates that our Gaussianization scheme is robust to variation of the fiducial cosmological model.

4.3.2 Posterior shape comparison

A graphical comparison of likelihood shapes is the quantile–quantile (Q–Q) probability plot (Wilk & Gnanadesikan 1968) of their respective posteriors. We show the f_{NL} percentiles inferred from all likelihoods in Table 2 except $\mathcal{L}_{G,rc}[\mathbf{Z}_{1/3}]$ against f_{NL} percentiles of the true likelihood $\mathcal{L}_{\text{true}}$ in Fig. 5, where we contrast no Gaussianization against Gaussianization, and fixed covariance estimates against rescaled covariance estimates.

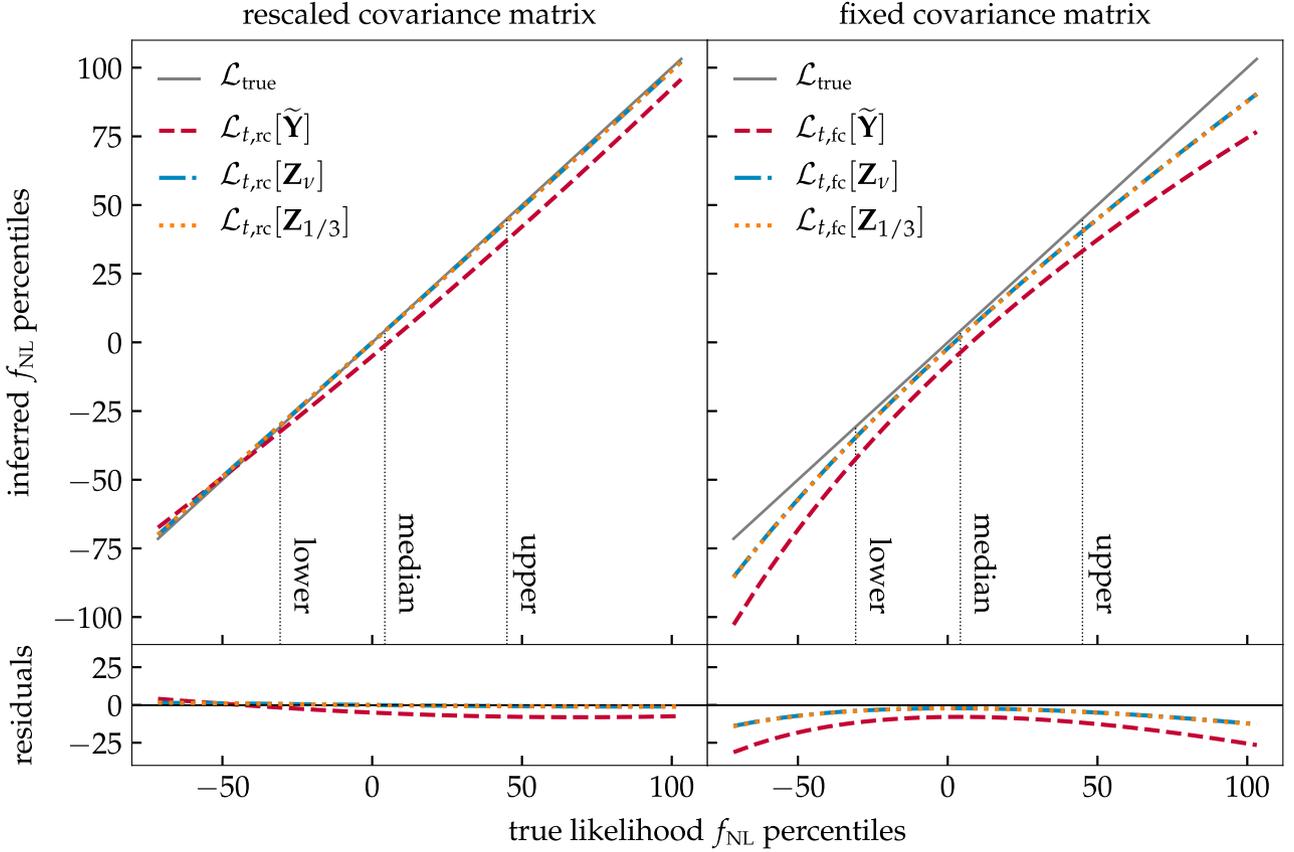


Figure 5. Q–Q plots comparing posterior distributions of the likelihoods without Gaussianization (data variable \tilde{Y} , dashed line), with fitted- ν Gaussianization (data variable Z_ν , dash-dotted lines) and with fixed- ν Gaussianization (data variable $Z_{1/3}$, dotted lines) against the true posterior from $\mathcal{L}_{\text{true}}$ as a reference (solid lines with unit slope). See Table 2 for definitions of the likelihoods. *Top panel:* inferred f_{NL} percentiles from the different likelihoods against the true likelihood $\mathcal{L}_{\text{true}}$ percentiles, averaged over 250 000 data realizations at $f_{\text{NL}} = 0$. *Bottom panel:* the residuals (numerical differences) of corresponding lines from the reference line. *Left-hand panel:* the fiducial covariance estimate is rescaled (‘rc’). *Right-hand panel:* the fiducial covariance estimate is correct but fixed in parameter space (‘fc’). The dotted vertical lines show the f_{NL} posterior median estimate and the 1σ uncertainties from the true likelihood.

There are two trends that match our expectation: the new likelihoods in the Gaussianized data variable Z matches the shape of the true likelihood $\mathcal{L}_{\text{true}}$ better than the ones in data variable \tilde{Y} without Gaussianization, especially away from the peak and near the tails of the distribution; not rescaling the covariance matrix in parameter space to account for its dependence on cosmology noticeably distorts the error bounds. Again we have also found that Gaussianization with fixed $\nu = 1/3$, i.e. $\mathcal{L}_{t,\text{rc}}[Z_{1/3}]$, produces nearly indistinguishable results from Gaussianization with fitted ν , i.e. $\mathcal{L}_{t,\text{rc}}[Z_\nu]$.

Another quantitative measure of ‘statistical distance’ between a true probability distribution $f(\theta)$ and an approximate probability distribution $g(\theta)$ is the Kullback–Leibler (KL) divergence (Kullback & Leibler 1951)

$$D_{\text{KL}}(f \parallel g) \equiv \int d\theta f(\theta) \ln \frac{f(\theta)}{g(\theta)}. \quad (48)$$

For instance, if we take f to be the posterior of the true likelihood against which we compare the posterior g of the new likelihood, then the expected KL divergence over the entire ensemble of data realizations could quantify the ‘information loss’ due to replacement of the true likelihood with the new.

In Table 5, we list the KL divergence values for all likelihoods in Table 2, except for $\mathcal{L}_{G,\text{rc}}[Z_{1/3}]$, from the true likelihood $\mathcal{L}_{\text{true}}$, averaged over our data realizations generated at different true f_{NL}

inputs. The evidence again suggests that the likelihood $\mathcal{L}_{t,\text{rc}}[Z_\nu]$ in Gaussianized data with the rescaled covariance estimate matches the full shape of true likelihood very well, and this remains the case when we use Gaussianization at fixed $\nu = 1/3$, i.e. $\mathcal{L}_{t,\text{rc}}[Z_{1/3}]$.

4.4 Sources of error in parameter inference

Although the major sources of impact on parameter inference, namely distribution non-normality and parameter dependence of the covariance matrix, have been identified and mitigated by Gaussianization and variance–correlation decomposition, respectively, there are other sources of error which we now consider.

The first potential concern is how the correlations between band powers affect the Gaussianization. We have proposed that the Gaussianization be performed using only the univariate distributions, and hence this will work the best when the off-diagonal correlations in the covariance matrix are relatively weak. Thus for a given window function we want to minimize the number of band powers to be included in the data vector so that the covariance matrix is strongly dominated by the diagonal entries. For future surveys with increasing volume, this will be easier as the window functions will be narrower in Fourier space. However, reducing the number of band powers will also mean that more overdensity modes contribute to each bin, making the statistics more Gaussian. The limit to how few band powers should be included in the data vector is that we

Table 5. Kullback–Leibler (KL) divergence values of f_{NL} posteriors of all likelihoods in Table 2, *except* for $\mathcal{L}_{\text{G,rc}}[\mathbf{Z}_{1/3}]$, from that of the true likelihood $\mathcal{L}_{\text{true}}$, averaged over 250 000 data realizations with different true f_{NL} inputs and the same fiducial cosmology at $f_{\text{NL}} = 0$.

Input	D_{KL} from the true posterior					
	$\mathcal{L}_{t,\text{rc}}[\tilde{\mathbf{Y}}]$	$\mathcal{L}_{t,\text{rc}}[\mathbf{Z}_{1/3}]$	$\mathcal{L}_{t,\text{rc}}[\mathbf{Z}_v]$	$\mathcal{L}_{t,\text{fc}}[\tilde{\mathbf{Y}}]$	$\mathcal{L}_{t,\text{fc}}[\mathbf{Z}_{1/3}]$	$\mathcal{L}_{t,\text{fc}}[\mathbf{Z}_v]$
50	0.12	0.02	0.02	0.35	0.05	0.05
10	0.11	0.02	0.02	0.18	0.04	0.04
0	0.10	0.02	0.02	0.17	0.05	0.05
−10	0.10	0.02	0.02	0.17	0.05	0.05
−50	0.11	0.03	0.03	0.23	0.10	0.09

need to make sure that there are sufficiently many bins to retain the cosmological information in the data.

The second potential concern is that when the band power variance cannot be analytically calculated from the window function mixing matrix and observed overdensity modes, the gamma distribution parameters have to be obtained from mock catalogues, and this comes with additional statistical scatter owing to the estimation of band power variance (see Section 2.2.2). Ideally it needs to be marginalized out together with the unknown full covariance matrix in the SH procedure, but this unfortunately makes the likelihood analytically intractable after Gaussianization.

To this end, we would like to assess the relative impact between covariance estimation of the Gaussianized band power and the estimation of the band power variance for evaluating gamma distribution parameters (R, η). Both estimations are made from the same set of mock catalogues, so we need to consider an ensemble of mock catalogue sets. For 25 000 data realizations, we now generate $N_s = 1000$ mock catalogue samples for *each* of them. In the reproduced Q–Q probability plots (Fig. 6) for likelihoods in the *Gaussianized* data \mathbf{Z} with fixed transformation parameter $\nu = 1/3$ and rescaled covariance estimates, we consider three scenarios of covariance estimation.

(i) The Hartlap-debiased precision matrix estimate is directly substituted into the Gaussian likelihood $\mathcal{L}_{\text{G,rc}}[\mathbf{Z}_{1/3}]$. This scenario corresponds to the dashed lines.

(ii) The unbiased covariance matrix estimate is marginalized with the SH procedure and thus the modified- t likelihood $\mathcal{L}_{t,\text{rc}}[\mathbf{Z}_{1/3}]$ is used. This scenario corresponds to the dash-dotted lines.

(iii) A high-precision covariance estimate is used as a proxy for the exact covariance matrix in the Gaussian likelihood $\mathcal{L}_{\text{G,rc}}[\mathbf{Z}_{1/3}]$. This scenario corresponds to the dotted lines.

The covariance estimate is rescaled for varying cosmology for each of these scenarios, and we compare using analytically calculated shape–scale parameters (R, η) (without ‘+’ markers) with using (R, η) obtained from estimated band power variance (with ‘+’ markers). In addition, we explore the effect of the mock catalogue size on the relative impacts between the two estimations, by adding the same plots (in the right-hand panel) for the case $N_s = 50$. The deviation from the true likelihood in these scenarios are shown as residuals (numerical differences) in the bottom panel of Fig. 6.

The evidence indicates that the SH procedure (the modified- t likelihood) indeed accounts for the statistical scatter of covariance estimation, but this effect is subdominant to the band power variance estimation for evaluating (R, η) if we use a set of $N_s = 1000$ mock catalogue samples – we can see the lines fall into two groups, depending on whether (R, η) are estimated or exact, in the left-hand panel of Fig. 6. If we reduce the sample size of

catalogues to $N_s = 50$, the impact of covariance estimation becomes greater than that of (R, η) estimation – this is evident as the lines corresponding to estimated covariance matrices deviate the most from all other lines. However, both effects are far less significant than the distribution non-normality and cosmological dependence of the covariance matrix, as we have seen in Fig. 5, which are the focal problems addressed in this work. In particular, the errors due to these estimations with $N_s = 1000$ are approaching the errors inherent in our gamma distribution approximation and univariate Gaussianization.

In light of these results, we recommend using the SH marginalization procedure for the covariance matrix estimate, i.e. the modified- t likelihood for parameter inference, even when we cannot do the same for the gamma distribution parameters calculated from the estimated band power variance.

5 APPLICATION PIPELINE

We now present the proposed final pipeline for the straightforward application of our methods. A comprehensive list of the notations used can be found in Table 1 in Section 1; note that in this section we use the superscript (f) to denote quantities evaluated at the fiducial cosmology, and superscript (d) for measurements or data realizations.

Gamma distribution parameters – We model the band power distribution as a gamma distribution in shape–scale parametrization (R, η). Given band power measurements or realizations $\tilde{P}_0^{(d)}(k_a)$ at some cosmology θ , the shape–scale parameters are determined from its mean and variance

$$R_a(\theta) = \mathbb{E}[\tilde{P}_0^{(d)}(k_a)]^2 / \text{Var}[\tilde{P}_0^{(d)}(k_a)],$$

$$\eta_a(\theta) = \text{Var}[\tilde{P}_0^{(d)}(k_a)] / \mathbb{E}[\tilde{P}_0^{(d)}(k_a)]. \quad (49)$$

In the absence of analytic expressions for the band power variance, this should be replaced by a fiducial estimate $\widehat{\text{Var}}[\tilde{P}_0^{(d,f)}(k_a)]$ calculated from mock catalogues and suitably rescaled with the cosmology θ , i.e.

$$\widehat{\text{Var}}[\tilde{P}_0^{(d)}(k_a, \theta)] = \left[\frac{\tilde{P}_0(k_a, \theta)}{\tilde{P}_0^{(f)}(k_a)} \right]^2 \widehat{\text{Var}}[\tilde{P}_0^{(d,f)}(k_a)], \quad (50)$$

which leads to corresponding rescaling for the distribution parameters in equation (49). Note that this rescaling cancels out for the shape parameter R_a , which is in fact independent of θ . This is expected as the effective number of modes is a model-independent quantity.

Data transformation – To make the data distribution approximately multivariate normal, we adopt the Box–Cox transformation whereby the band power measurements are univariately

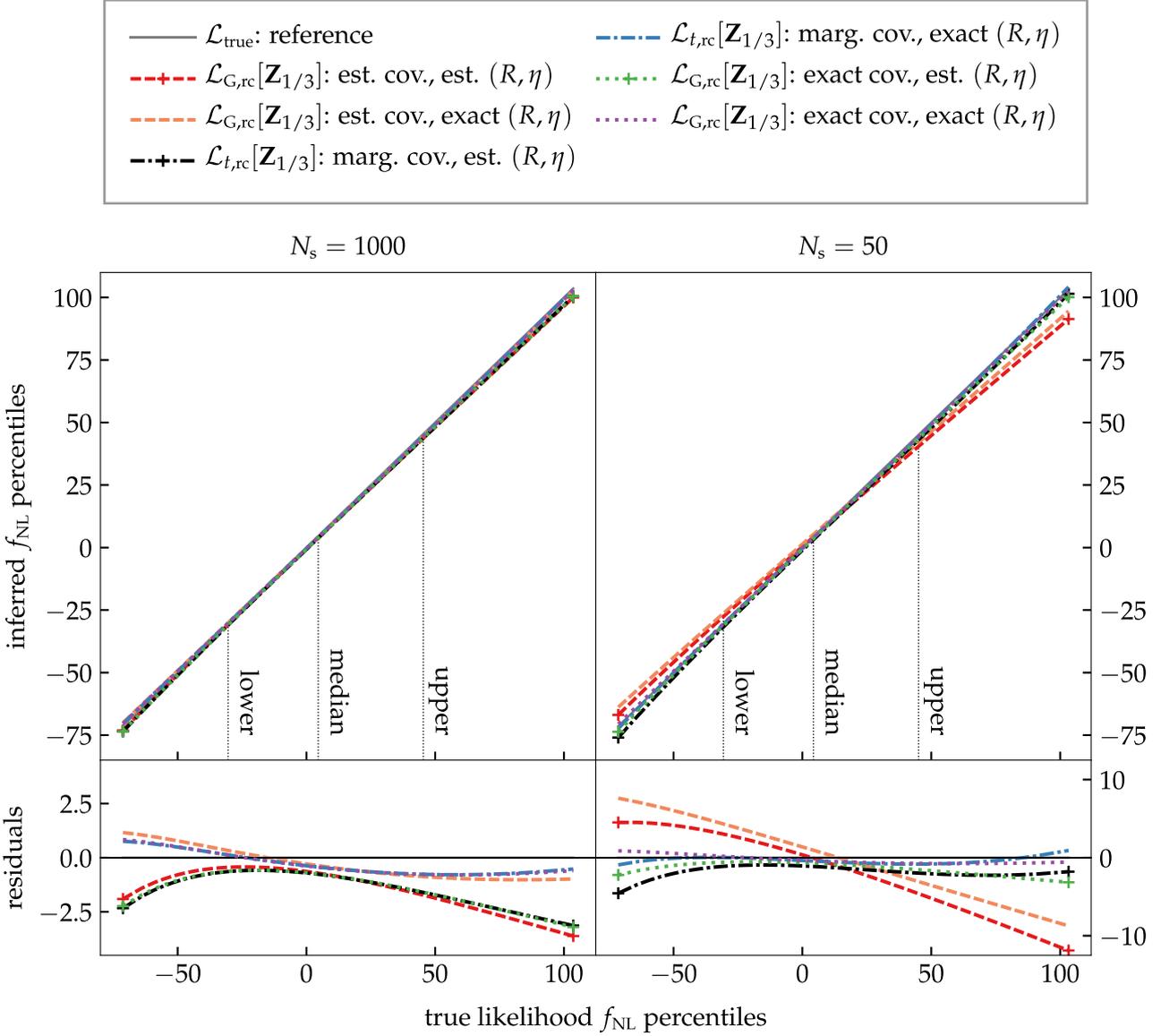


Figure 6. Q–Q plots comparing the relative impact of covariance estimation and band power variance estimation for evaluating gamma distribution parameters (R, η). All likelihoods compared are in the Gaussianized data $\mathbf{Z}_{1/3}$ with the covariance matrix estimate rescaled for varying cosmology. The covariance matrix is either estimated without SH marginalization (Gaussian likelihood $\mathcal{L}_{G,rc}[\mathbf{Z}_{1/3}]$, *dashed lines*), SH marginalized (modified- t likelihood $\mathcal{L}_{t,rc}[\mathbf{Z}_{1/3}]$, *dash-dotted lines*) or exact (Gaussian likelihood $\mathcal{L}_{G,rc}[\mathbf{Z}_{1/3}]$, *dotted lines*); the distribution parameters are either estimated (*with '+' markers*) or exact (*without '+' markers*). *Top panel*: inferred f_{NL} percentiles from the different cases against percentiles of the true likelihood $\mathcal{L}_{\text{true}}$, averaged over 25 000 data realizations at $f_{\text{NL}} = 0$. *Bottom panel*: the residuals (numerical differences) of corresponding lines from the reference line (the true likelihood $\mathcal{L}_{\text{true}}$). *Left-hand panel*: mock catalogues contain $N_s = 1000$ samples. *Right-hand panel*: mock catalogues contain $N_s = 50$ samples. The dotted vertical lines show the f_{NL} posterior median estimate and the 1σ uncertainties from the true likelihood. **Note** – the scales in the bottom panels differ, and also differ from those of the bottom panels in Fig. 5.

Gaussianized,

$$\tilde{P}_0^{(d)}(k_a) \mapsto Z_a \equiv \left[\tilde{P}_0^{(d)}(k_a) \right]^{\nu_a}. \quad (51)$$

Whilst the transformation parameters ν_a for each bin can be determined using the fitting formula given by equation (30) as a function of the fiducial shape parameter $R_a^{(f)}$, we have found little gain over keeping this fixed at $\nu_a = 1/3$, which we favour for reasons of simplicity. After the transformation, the mean $\mu_a(\theta)$ and variance $\sigma_a^2(\theta)$ of the Gaussianized band power Z_a at cosmology θ are given by equation (26) for each bin.

Likelihood evaluation – The remaining quantity needed for likelihood evaluation is the covariance matrix estimate $\hat{\Sigma}(\theta)$ for

the *Gaussianized* data \mathbf{Z} , which is allowed to vary with cosmology by rescaling the fiducial estimate $\hat{\Sigma}^{(f)}$ from $N_s \equiv m + 1$ mock catalogue samples,

$$\hat{\Sigma}(\theta) = \mathbf{D} \hat{\Sigma}^{(f)} \mathbf{D}, \quad (52)$$

where the diagonal matrix \mathbf{D} consists of entries

$$D_{aa} = \sigma_a(\theta) / \sigma_a(\theta^{(f)}). \quad (53)$$

We recommend using the modified- t distribution obtained with SH marginalization as the final likelihood (see equations 39 and 41),

$$\mathcal{L}(\theta; \mathbf{Z}, \hat{\Sigma}^{(f)}, m) = \mathbb{P}_t[\mathbf{Z}; \boldsymbol{\mu}(\theta), \hat{\Sigma}(\theta), m]. \quad (54)$$

Our simulations in Section 4 have shown that the impact from the errors in the poorly determined covariance matrix entries dominates over problems caused by the estimated gamma distribution parameters in our likelihood. Consequently, it is worth marginalizing out scatter of the estimated covariance matrix using the SH procedure even if we cannot simultaneously perform an equivalent procedure for the gamma distribution parameters.

Standard Bayesian inference can be readily performed now to extract cosmological parameter estimates and associated uncertainties, or to sample the posterior distribution in a multidimensional parameter space using Monte Carlo techniques.

6 CONCLUSION

In preparation for next-generation galaxy surveys such as DESI and *Euclid*, we have revisited the Gaussian likelihood assumption commonly found in galaxy-clustering likelihood analyses, which may adversely impact cosmological parameter inference from measurements limited by sample size on the largest survey scales. Extending previous work by Schneider & Hartlap (2009), Keitel & Schneider (2011), Wilking & Schneider (2013), Sun et al. (2013), and Kalus et al. (2016), we have carefully derived the distribution of the band power spectrum (windowed power spectrum monopole) in the linear regime while taking window effects and random shot noise into account; in particular, we have

- (i) devised a Gaussianization scheme using the Box–Cox transformation to improve data normality;
- (ii) proposed a variance–correlation decomposition of the covariance matrix to allow for varying cosmology;
- (iii) presented a simple pipeline for straightforward application of this new methodology (Section 5).

We always recommend rescaling the covariance matrix estimate using our decomposition as its parameter dependence has a significant impact on parameter estimation. Although below the largest survey scales the normal distribution may be a good approximation for the band power measurements, we still recommend the use of our Gaussianization scheme for its simplicity.

With numerical simulations, we have tested the likelihood derived from the new procedure for both point estimation and shape comparison with the true likelihood inaccessible in real surveys. By focusing on the local non-Gaussianity f_{NL} , which is a sensitive parameter for the large-scale power spectrum, we have demonstrated noticeable improvement in parameter inference brought by Gaussianization and covariance rescaling. Whilst Gaussianizing transformations are not new, our set-up, motivation and implementation differ from previous works by, for instance, Wilking & Schneider (2013) and Schuhmann et al. (2016).

However, an all-encompassing formalism for galaxy-clustering power spectrum analysis is still out of reach. Towards the non-linear regime where overdensity modes are no longer independent but coupled due to gravitational evolution, the power-spectrum covariance structure is fundamentally more complex, and non-negligible shot noise can also deviate from the Poisson sampling prescription (Bernardeau et al. 2002). The analysis covered in this paper focuses on the windowed power spectrum monopole in the FKP framework, but this could also be applied to power-law moment estimators with even exponents in the local plane-parallel approximation (Yamamoto et al. 2006). We leave further extensions to the current analysis to future work.

ACKNOWLEDGEMENTS

MSW, WJP, and DB acknowledge support from the European Research Council (ERC) through the Darksurvey grant 614030. SA acknowledges support from the UK Space Agency through grant ST/K00283X/1. RC is supported by the Science and Technology Facilities Council (STFC) grant ST/N000668/1.

Numerical computations are performed on the Sciama High Performance Computing (HPC) cluster which is supported by the Institute of Cosmology and Gravitation (ICG), the South East Physics Network (SEPnet) and the University of Portsmouth.

REFERENCES

- Abbott T. M. C. et al., 2018, *Phys. Rev. D*, 98, 043526
 Ade P. A. R. et al., 2016, *A&A*, 594, A17
 Aghamousa A. et al., 2016, preprint (arXiv:1611.00036)
 Aghanim N. et al., 2018, preprint (arXiv:1807.06209)
 Alam S. et al., 2017, *MNRAS*, 470, 2617
 Anderson T. W., 2003, *An Introduction to Multivariate Statistical Analysis*, 3rd edn. Wiley, New York
 Avila S. et al., 2018, *MNRAS*, 479, 94
 Berger J. O., 1985, *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, Berlin
 Bernardeau F., Colombi S., Gazta naga E., Scoccimarro R., 2002, *Phys. Rep.*, 367, 1
 Beutler F. et al., 2014, *MNRAS*, 443, 1065
 Beutler F. et al., 2017, *MNRAS*, 466, 2242
 Blot L. et al., 2019, *MNRAS*, 485, 2806
 Box G. E. P., Cox D. R., 1964, *J. Royal Stat. Soc.*, 26, 211
 Burić T., Elezović N., 2012, *Integral Transforms Spec. Funct.*, 23, 355
 Colavincenzo M. et al., 2018, *MNRAS*, 482, 4883
 D’Agostino R., Pearson E. S., 1973, *Biometrika*, 60, 613
 Dalal N., Dore O., Huterer D., Shirokov A., 2008, *Phys. Rev. D*, 77, 123514
 Dodelson S., Schneider M. D., 2013, *Phys. Rev. D*, 88, 063537
 Duffy A. R., 2014, *Ann. Phys.*, 526, 283
 Eifer T., Schneider P., Hartlap J., 2009, *A&A*, 502, 721
 Eisenstein D. J., Hu W., 1998, *ApJ*, 496, 605
 Feldman H. A., Kaiser N., Peacock J. A., 1994, *ApJ*, 426, 23
 Gabrielli A., Sylos Labini F., Joyce M., Pietronero L., 2005, *Statistical Physics for Cosmic Structures*. Springer-Verlag, Berlin Heidelberg
 Golubev A., 2016, *J. Theor. Biol.*, 393, 203
 Gupta A., Nagar D., 2000, *Matrix Variate Distributions*. Chapman and Hall/CRC, New York
 Hamimeche S., Lewis A., 2008, *Phys. Rev. D*, 77, 103013
 Hartlap J., Simon P., Schneider P., 2007, *A&A*, 464, 399
 Hogg D. W., Foreman-Mackey D., 2018, *ApJS*, 236, 11
 Joachimi B., Taylor A. N., 2011, *MNRAS*, 416, 1010
 Kalus B., Percival W. J., Samushia L., 2016, *MNRAS*, 455, 2573
 Kaufman C. G., Schervish M. J., Nychka D. W., 2008, *J. Am. Stat. Assoc.*, 103, 1545
 Keitel D., Schneider P., 2011, *A&A*, 534, A76
 Kitaura F.-S. et al., 2016, *MNRAS*, 456, 4156
 Kullback S., Leibler R. A., 1951, *Ann. Math. Stat.*, 22, 79
 Laparra V., Camps-Valls G., Malo J., 2011, *IEEE Trans. Neural Netw.*, 22, 537
 Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
 Li Y., Singh S., Yu B., Feng Y., Seljak U., 2019, *J. Cosmol. Astropart. Phys.*, 01, 016
 Lippich M. et al., 2019, *MNRAS*, 482, 1786
 Manera M. et al., 2013, *MNRAS*, 428, 1036
 Matarrese S., Verde L., 2008, *ApJ*, 677, L77
 Neuts M. F., 1981, *Matrix-Geometric Solutions in Stochastic Models: an Algorithmic Approach*. Dover Publications Inc., New York
 Paz D. J., Sánchez A. G., 2015, *MNRAS*, 454, 4326
 Peacock J. A., Nicholson D., 1991, *MNRAS*, 253, 307

Peebles P. J. E., 1980, *The Large-Scale Structure of the Universe*. Princeton Univ. Press, Princeton

Percival W. J., Brown M. L., 2006, *MNRAS*, 372, 1104

Percival W. J. et al., 2014, *MNRAS*, 439, 2531

Pope A. C., Szapudi I., 2008, *MNRAS*, 389, 766

Schneider P., Hartlap J., 2009, *A&A*, 504, 705

Schuhmann R. L., Joachimi B., Peiris H. V., 2016, *MNRAS*, 459, 1916

Seljak U., Aslanyan G., Feng Y., Modi C., 2017, *J. Cosmol. Astropart. Phys.*, 12, 009

Sellentin E., Heavens A. F., 2016, *MNRAS*, 456, L132

Sellentin E., Heavens A. F., 2018, *MNRAS*, 473, 2355

Sellentin E., Jaffe A. H., Heavens A. F., 2017, preprint (arXiv:1709.03452)

Shao Y., Zhou M., 2010, *J. Multivar. Anal.*, 101, 2637

Slosar A., Hirata C., Seljak U., Ho S., Padmanabhan N., 2008, *J. Cosmol. Astropart. Phys.*, 08, 031

Smith S., Challinor A., Rocha G., 2006, *Phys. Rev. D*, 73, 023517

Sun L., Wang Q., Zhan H., 2013, *ApJ*, 777, 75

Taylor A., Joachimi B., Kitching T., 2013, *MNRAS*, 432, 1928

Tegmark M., 1997, *Phys. Rev. D*, D55, 5895

Tegmark M., Hamilton A. J. S., Strauss M. A., Vogeley M. S., Szalay A. S., 1998, *ApJ*, 499, 555

Tellarini M., Ross A. J., Tasinato G., Wands D., 2015, *J. Cosmol. Astropart. Phys.*, 07, 004

Trotta R., 2008, *Contemp. Phys.*, 49, 71

Verde L. et al., 2003, *ApJS*, 148, 195

Wilking P., Schneider P., 2013, *A&A*, 556, A70

Wilk M. B., Gnanadesikan R., 1968, *Biometrika*, 55, 1

Wilson M. J., Peacock J. A., Taylor A. N., de la Torre S., 2017, *MNRAS*, 464, 3121

Wishart J., 1928, *Biometrika*, 20A, 32

Yamamoto K., Nakamichi M., Kamino A., Bassett B. A., Nishioka H., 2006, *PASJ*, 58, 93

APPENDIX A: SHOT NOISE POWER AND ITS DISTRIBUTION

Here we derive the amplitude of the shot noise power and consider its distribution, which affects the power spectrum likelihood. Following the calculations in Peebles (1980) and Feldman et al. (1994) for the two-point correlation of the Poisson-sampled overdensity field $\hat{\delta}$ (see equation 11), we have

$$\begin{aligned} \langle \hat{\delta}(\mathbf{q}) \hat{\delta}^*(\mathbf{q}') \rangle &= \int d^3\mathbf{r} d^3\mathbf{r}' e^{i(\mathbf{q}-\mathbf{q}')\cdot\mathbf{r}} e^{i\mathbf{q}'\cdot(\mathbf{r}-\mathbf{r}')} \\ &\times \left[\xi(\mathbf{r}-\mathbf{r}') + \frac{1+\alpha}{\bar{n}(\mathbf{r})} \delta^D(\mathbf{r}-\mathbf{r}') \right] \\ &= P_{\text{true}}(\mathbf{q}) \delta_{\mathbf{q},\mathbf{q}'}^K + \frac{1+\alpha}{V} \int d^3\mathbf{r} e^{i(\mathbf{q}-\mathbf{q}')\cdot\mathbf{r}} \bar{n}(\mathbf{r})^{-1}, \end{aligned} \quad (\text{A1})$$

where $\delta^D(\mathbf{r})$ is the Dirac delta function and $\delta_{\mathbf{q},\mathbf{q}'}^K$ is the Kronecker delta function. This expectation value contains both the underlying power spectrum and the scale-invariant shot noise power

$$P_{\text{shot}} = \frac{1+\alpha}{V} \int d^3\mathbf{r} \bar{n}(\mathbf{r})^{-1}. \quad (\text{A2})$$

To determine the distribution of the stochastic shot noise, we consider the scenario where N galaxies are randomly located at $\{\mathbf{x}_i\}_{i=1}^N$ in a finite volume. In this set-up, the overdensity field and its Fourier transform are

$$\delta(\mathbf{r}) = \frac{V}{N} \sum_{i=1}^N \delta^D(\mathbf{r}-\mathbf{x}_i) - 1, \quad \delta(\mathbf{k}) = \frac{1}{N} \sum_{i=1}^N e^{i\mathbf{k}\cdot\mathbf{x}_i}, \quad (\text{A3})$$

where in $\delta(\mathbf{k})$ we have dropped a Dirac delta term that vanishes for $\mathbf{k} \neq \mathbf{0}$. In the large galaxy number limit $N \rightarrow \infty$, regardless of the

detailed distribution of the summands $\exp(i\mathbf{k}\cdot\mathbf{x}_i)$ where $\{\mathbf{x}_i\}_{i=1}^N$ are independently uniformly distributed, $\delta(\mathbf{k})$ becomes a Gaussian random field by the central limit theorem (Peacock & Nicholson 1991). Hence the shot noise power is also exponentially distributed (cf. Section 2.2.1), and will overlay the exponential distribution of any underlying power if there is any intrinsic structure in galaxy clustering.

APPENDIX B: HYPO-EXPONENTIAL DISTRIBUTION

Here we derive the form of the hypo-exponential PDF (equation 18) introduced in Section 2. We will also show that the sum of independently *identically* distributed exponential random variables follows the gamma distribution. This motivates the gamma distribution approximation of the hypo-exponential distribution in Section 2.2.2.

Let X be the sum of independent exponential variables $\{X_i\}_{i=1}^r$ with PDF

$$\mathbb{P}_i[X_i = x_i] = \beta_i \exp(-\beta_i x_i), \quad (\text{B1})$$

where $\{\beta_i^{-1}\}_{i=1}^r$ are their respective scale parameters. Then the PDF of X is the convolution of the individual PDFs \mathbb{P}_i ,

$$\begin{aligned} \mathbb{P}[X = x] &= \left(\begin{matrix} r \\ * \\ i=1 \end{matrix} \mathbb{P}_i \right) [x] \equiv \int \prod_{i=1}^r dt_i \mathbb{P}_i[x - t_i] \\ &\times \prod_{j=2}^r \mathbb{P}_j[t_{j-1} - t_j] = \sum_i \mathbb{P}_i[x] \prod_{\substack{j \neq i \\ j=1}}^r \frac{\beta_j}{\beta_j - \beta_i}. \end{aligned} \quad (\text{B2})$$

We prove the last equality by induction on r : the initial statement for $r = 2$ is easy to check, so we only need to establish the inductive step

$$\begin{aligned} \mathbb{P}[X = x] &= \int_0^x dt \sum_{i=1}^{r-1} \left(\prod_{\substack{j \neq i \\ j=1}}^{r-1} \frac{\beta_j}{\beta_j - \beta_i} \right) \mathbb{P}_i[t] \mathbb{P}_r[x - t] \\ &= e^{-\beta_r x} \sum_{i=1}^{r-1} \left(\prod_{\substack{j \neq i \\ j=1}}^{r-1} \frac{\beta_j}{\beta_j - \beta_i} \right) \beta_i \beta_r \int_0^x dt e^{-(\beta_i - \beta_r)t} \\ &= \sum_{i=1}^{r-1} \left(\prod_{\substack{j \neq i \\ j=1}}^{r-1} \frac{\beta_j}{\beta_j - \beta_i} \right) \frac{\beta_i \mathbb{P}_r[x] - \beta_r \mathbb{P}_i[x]}{\beta_i - \beta_r} \\ &= \sum_{i=1}^r \left(\prod_{\substack{j \neq i \\ j=1}}^r \frac{\beta_j}{\beta_j - \beta_i} \right) \mathbb{P}_i[x]. \end{aligned} \quad (\text{B3})$$

This is an example of the hypo-exponential family of distributions, sometimes also referred to as the generalized Erlang distribution (Neuts 1981). We note here the particular case $\beta_i = \beta_j$ for some $i \neq j$, when two variables are also identically distributed. Using the formula above by taking the limit $\Delta\beta \equiv \beta_i - \beta_j \rightarrow 0$, the PDF of $(X_i + X_j)$ is

$$\begin{aligned} \mathbb{P}[X_i + X_j = x] &= \lim_{\epsilon\beta \rightarrow 0} \frac{\beta_i (\beta_i - \Delta\beta)}{\Delta\beta} e^{-\beta_i x} (1 - e^{x\Delta\beta}) \\ &= \beta_i^2 x e^{-\beta_i x}. \end{aligned} \quad (\text{B4})$$

We recognize this as a gamma distribution $\Gamma(2, \beta_i^{-1})$ in the shape-scale parametrization.

This recovers the usual result that the sum of independently *identically* distributed exponential random variables follows the gamma distribution; it also motivates our gamma distribution approximation of the hypo-exponential distribution.

APPENDIX C: ERROR-FUNCTION TRANSFORMATION

We consider an alternative to the Box–Cox transformation adopted as our default Gaussianization scheme. This is derived by matching the cumulative distribution functions (CDF) and involves the (complementary) error function. Whilst this scheme is exact in principle, it requires computationally costly numerical integrations for calculating transformed moments.

To this end, we seek an invertible transformation $\tilde{Y} \mapsto Z$ of the gamma random variable with fiducial shape–scale parameters (R_f, η_f) , where $Z \sim \mathcal{N}(0, 1)$ is a standard normal variable with zero mean and unit variance, by matching the CDFs

$$\int_0^{\tilde{Y}} dt \mathbb{P}_\Gamma[t; R_f, \eta_f] = \int_{-\infty}^Z dt \mathbb{P}_\mathcal{N}[t; 0, 1], \quad (\text{C1})$$

where the gamma PDF is given by equation (20) and the normal PDF with zero mean and unit variance is

$$\mathbb{P}_\mathcal{N}[t; 0, 1] = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}. \quad (\text{C2})$$

The solution with $dZ/d\tilde{Y} > 0$ gives the transformation

$$Z = -\sqrt{2} \operatorname{erfc}^{-1} \left[2 \frac{\gamma(R_f, \tilde{Y}/\eta_f)}{\Gamma(R_f)} \right], \quad (\text{C3})$$

where $\operatorname{erfc}^{-1}(x)$ is the inverse of the complementary error function

$$\operatorname{erfc}(x) \equiv \frac{2}{\sqrt{\pi}} \int_x^\infty dt e^{-t^2}. \quad (\text{C4})$$

For a gamma variable \tilde{Y} with different shape–scale parameters (R, η) transformed using equation (C3), the PDF in Z is now

$$\mathbb{P}[z; R, \eta] = \mathbb{P}_\mathcal{N}[t; 0, 1] \frac{\mathbb{P}_\Gamma[\tilde{y}(z); R, \eta]}{\mathbb{P}_\Gamma[\tilde{y}(z); R_f, \eta_f]}, \quad (\text{C5})$$

with mean and variance given by

$$\begin{aligned} \mu(R, \eta) &= \int_{-\infty}^{\infty} dz z \mathbb{P}[z; R, \eta], \\ \sigma^2(R, \eta) &= \int_{-\infty}^{\infty} dz z^2 \mathbb{P}[z; R, \eta] - \mu(R, \eta)^2. \end{aligned} \quad (\text{C6})$$

These results are analogous to equation (26) for the Box–Cox Gaussianization scheme; however, in this case accurate evaluation of transformed moments requires computationally expensive numerical integration for each parameter pair (R, η) . For this reason we do not implement this scheme in our pipeline.

APPENDIX D: COVARIANCE MARGINALIZATION

We now show that the covariance matrix decomposition, which is used in rescaling the fiducial covariance estimate to allow for cosmological dependence, does not affect the SH procedure for

marginalizing out the scatter due to covariance estimation using simulated data samples.

Let us consider the unbiased estimator of the true covariance matrix $\Sigma \in \mathbb{R}^{p \times p}$ for $(m+1)$ samples $\{\mathbf{X}_i\}_{i=1}^{m+1}$ of a random vector,

$$\hat{\Sigma} = \frac{1}{m+1} \sum_i (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^\top,$$

$$\text{where } \bar{\mathbf{X}} = \frac{1}{m+1} \sum_i \mathbf{X}_i. \quad (\text{D1})$$

The distribution of $\hat{\Sigma}$ conditional on Σ is Wishart $W_p(\Sigma/m, m)$ with the PDF (Wishart 1928)

$$\begin{aligned} \mathbb{P}_W[\hat{\Sigma} | \Sigma] &= \frac{|\Sigma/m|^{-m/2}}{2^{mp/2} \Gamma_p(m/2)} |\hat{\Sigma}|^{(m-p-1)/2} \\ &\times \exp\left[-\frac{m}{2} \operatorname{tr}(\Sigma^{-1} \hat{\Sigma})\right], \end{aligned} \quad (\text{D2})$$

where the multivariate gamma function Γ_p is defined by

$$\Gamma_p(x) \equiv \pi^{(p-1)p/4} \prod_{j=1}^p \Gamma\left(x + \frac{1-j}{2}\right). \quad (\text{D3})$$

We also introduce the inverse Wishart distribution $W_p^{-1}(m\hat{\Sigma}, m)$ with the PDF

$$\begin{aligned} \mathbb{P}_{W^{-1}}[\Sigma | \hat{\Sigma}] &= \frac{|m\hat{\Sigma}|^{m/2}}{2^{mp/2} \Gamma_p(m/2)} |\Sigma|^{-(m+p+1)/2} \\ &\times \exp\left[-\frac{m}{2} \operatorname{tr}(\hat{\Sigma} \Sigma^{-1})\right]. \end{aligned} \quad (\text{D4})$$

Both the Wishart and inverse Wishart distribution possess the following invariance property (Gupta & Nagar 2000): given a non-singular matrix $\mathbf{D} \in \mathbb{R}^{p \times p}$,

- (i) if $\Psi \sim W_p(\mathbf{A}, m)$, then $\mathbf{D}^\top \Psi \mathbf{D} \sim W_p(\mathbf{D}^\top \mathbf{A} \mathbf{D}, m)$;
- (ii) if $\Psi \sim W_p^{-1}(\mathbf{A}^{-1}, m)$, then $\mathbf{D}^\top \Psi \mathbf{D} \sim W_p^{-1}(\mathbf{D}^\top \mathbf{A}^{-1} \mathbf{D}, m)$.

Since in our covariance matrix decomposition the rescaling matrix is the diagonal matrix of standard deviations (see equations 31 and 32), the cosmology-varying covariance matrix $\Sigma(\theta)$ has the inverse Wishart posterior distribution $W_p^{-1}(m\hat{\Sigma}(\theta), m)$,

$$\begin{aligned} \mathcal{P}(\Sigma(\theta) | \hat{\Sigma}(\theta)) &= \frac{|m\hat{\Sigma}|^{m/2}}{2^{mp/2} \Gamma_p(m/2)} |\Sigma|^{-(m+p+1)/2} \\ &\times \exp\left[-\frac{m}{2} \operatorname{tr}(\hat{\Sigma} \Sigma^{-1})\right]. \end{aligned} \quad (\text{D5})$$

This shows that the variance–correlation decomposition does not change the SH marginalization step in Section 3.2.

Finally, by marginalizing the normal distribution over the posterior distribution of $\Sigma | \hat{\Sigma}$ derived above, we can replace the unknown covariance matrix tribution over the posterior distribution of $\Sigma(\theta)$ with an unbiased estimate $\hat{\Sigma}(\theta)$ from $(m+1)$ data samples. This leads to the modified- t distribution (equation 39) introduced in Sellentin & Heavens (2016), which we recommend using as the likelihood form in our analysis pipeline.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.