



# Validation of semi-analytical, semi-empirical covariance matrices for two-point correlation function for early DESI data

Michael Rashkovetskyi <sup>1</sup>★, Daniel J. Eisenstein <sup>1</sup>, Jessica Nicole Aguilar,<sup>2</sup> David Brooks,<sup>3</sup> Todd Claybaugh,<sup>2</sup> Shaun Cole,<sup>4</sup> Kyle Dawson,<sup>5</sup> Axel de la Macorra,<sup>6</sup> Peter Doel,<sup>3</sup> Kevin Fanning,<sup>7</sup> Andreu Font-Ribera,<sup>8</sup> Jaime E. Forero-Romero,<sup>9</sup> Satya Gontcho A Gontcho,<sup>2</sup> ChangHoon Hahn,<sup>10</sup> Klaus Honscheid,<sup>7</sup> Robert Kehoe,<sup>11</sup> Theodore Kisner,<sup>2</sup> Martin Landriau,<sup>2</sup> Michael Levi,<sup>2</sup> Marc Manera,<sup>8</sup> Ramon Miquel,<sup>8,12</sup> Jeongin Moon,<sup>13,14</sup> Seshadri Nadathur,<sup>15</sup> Jundan Nie,<sup>16</sup> Claire Poppett,<sup>2,17</sup> Ashley J. Ross,<sup>7</sup> Graziano Rossi,<sup>13</sup> Eusebio Sanchez,<sup>18</sup> Christoph Saulder,<sup>19</sup> Michael Schubnell,<sup>20</sup> Hee-Jong Seo,<sup>21</sup> Gregory Tarle,<sup>20</sup> David Valcin,<sup>21</sup> Benjamin Alan Weaver,<sup>22</sup> Cheng Zhao,<sup>23,24</sup> Zhimin Zhou<sup>16</sup> and Hu Zou<sup>16</sup>

*Affiliations are listed at the end of the paper*

Accepted 2023 July 8. in original form 2023 June 13

## ABSTRACT

We present an extended validation of semi-analytical, semi-empirical covariance matrices for the two-point correlation function (2PCF) on simulated catalogs representative of luminous red galaxies (LRGs) data collected during the initial 2 months of operations of the Stage-IV ground-based Dark Energy Spectroscopic Instrument (DESI). We run the pipeline on multiple effective Zel’dovich (EZ) mock galaxy catalogs with the corresponding cuts applied and compare the results with the mock sample covariance to assess the accuracy and its fluctuations. We propose an extension of the previously developed formalism for catalogs processed with standard reconstruction algorithms. We consider methods for comparing covariance matrices in detail, highlighting their interpretation and statistical properties caused by sample variance, in particular, non-trivial expectation values of certain metrics even when the external covariance estimate is perfect. With improved mocks and validation techniques, we confirm a good agreement between our predictions and sample covariance. This allows one to generate covariance matrices for comparable data sets without the need to create numerous mock galaxy catalogs with matching clustering, only requiring 2PCF measurements from the data itself. The code used in this paper is publicly available at <https://github.com/oliverphilcox/RascalC>.

**Key words:** methods: statistical – surveys – software: data analysis – galaxies: statistics – large-scale structure of Universe – cosmology: theory.

## 1 INTRODUCTION

Measurements of the large-scale structure of the Universe are one of the pillars of modern cosmology. The two-point correlation function (2PCF) of galaxies is a particularly important statistical quantity for the large-scale structure, describing the excess probability of finding a galaxy at a given separation from another galaxy, compared to a random distribution. Its measurements have a notable feature at the scale of baryon acoustic oscillations [BAO; first detected by Eisenstein et al. (2005)]. As the corresponding comoving scale is a standard ruler with length set by sound horizon during recombination (more precisely, drag epoch), these measurements particularly constrain the expansion history of the Universe at redshifts between then and now ( $z \sim 1$ ), providing a valuable test for cosmological models. A more detailed overview of the methodology is provided in Weinberg et al. (2013).

We have an exciting opportunity to analyse the data from the Dark Energy Spectroscopic Instrument [DESI; DESI Collaboration (2016, 2022, 2023a, b)], a highly promising 5-yr Stage-IV BAO experiment for large-scale spectroscopic surveys. For example, the sample of luminous red galaxies [LRG; Zhou et al. (2020)] observed during the first 2 months of DESI main survey operations (DESI-M2) yields a BAO scale measurement with 1.7 per cent precision (Moon et al. 2023), which is already comparable to the aggregate precision of 0.77 per cent of preceding leading surveys, BOSS and eBOSS (Alam et al. 2021).

For rigorous interpretation of data, the likelihood is crucial. Fortunately, the distribution of measured clustering statistics is well described by a multivariate Gaussian, which is fully described by mean and covariance matrix. Unfortunately, the latter poses a serious challenge. A fully analytical model for the covariance matrix is desirable because it can provide a fast and stable result. However, it is very hard to construct in the case of galaxy clustering. One of the reasons is that galaxies are high matter overdensities, evolving in a crucially non-linear regime. Another is the complicated effects of

\* E-mail: [mrashkovetskyi@cfa.harvard.edu](mailto:mrashkovetskyi@cfa.harvard.edu)

survey geometry and non-uniform selection. There have been recent promising developments on covariance matrices for power spectra using perturbation theory (Wadekar & Scoccimarro 2020). However, power spectra results are difficult to apply to the correlation functions, because the Fourier transform is not local.

The standard approach is using a sample covariance estimated from mock (simulated) galaxy catalogs. This solution is far from ideal. Such catalogs need to capture key aspects of clustering and be representative of the data or assumed theoretical model. Detailed simulations are computationally expensive, while for precise covariance estimate, a large number of samples is required, increasingly higher as more quantities are measured. This forces a hard compromise between quality and quantity, keeping the total computation time very long. The mock-based covariance production is therefore not characterized by flexibility, as the generation and processing of numerous simulations for an updated data set or alternative model require a huge effort.

Another group of methods is called internal for using only the data itself. The lack of dependence on mocks and model assumptions makes them attractive. These are represented by re-sampling techniques like jackknife and bootstrap, which involve splitting the data into parts. However, it may be hard to ensure that the parts are principally equivalent (so the differences all stem from random fluctuations) and separate the independent contributions to different estimates used for covariance. Mohammad & Percival (2022) attempt to mitigate the latter issue for jackknives by modifying the pair weighting for jackknife covariance estimates. However, Trusov et al. (2023) find that the results are still prone to a density-dependent bias.

Each of the three approaches has its flaws and combining advantages from different ones seems very promising. Thus, we choose to focus on a method that started nearly analytical and employed elements of an internal approach. O’Connell et al. (2016) demonstrated that the covariance matrix for 2PCF in arbitrary survey geometry can be expressed through integrals or sums in configuration space [this result is analogous to Bernstein (1994)], which can be computed efficiently with importance sampling techniques. There are terms containing higher-point correlations (three-point and connected four-point functions), which are highly challenging to model or measure with precision. Thus, instead a procedure of leaving only Gaussian correlation and rescaling of shot noise amplitude was proposed and found to achieve a good agreement with mock-based covariances. This rescaling parameter can be calibrated on a small suite of mocks, or on a jackknife covariance intrinsic to the data analysed (O’Connell & Eisenstein 2019). We find the combination of analytical methods with jackknife more promising than with mocks, because, after a validation, such a procedure does not require the construction of any new mocks to match updated data or an alternative model with different assumptions and does not require any more than jackknife covariance computation, at the same time offering higher smoothness, stability, and invertibility. Philcox et al. (2020) introduced the RASCALC<sup>1</sup> code with a new algorithm, making it easier to account for survey geometry via a catalog of random points, boosting the efficiency several times and extended the formalism to covariances for multiple tracers (galaxy types). Philcox & Eisenstein (2019) developed the estimators for Legendre-binned 2PCF and isotropic three-point correlation function.

<sup>1</sup><https://github.com/oliverphilcox/RascalC> (Oliver Philcox, Daniel Eisenstein, Ross O’Connell, Alexander Wiegand, Misha Rashkovetskiy, Yuting Wang, Ryuichiro Hada, Uendert Andrade)

We have contributed RASCALC covariances for the BAO analysis of DESI-M2 data (Moon et al. 2023). This work accompanies it, focusing on the validation of the approach in realistic circumstances. We limit ourselves to analogs of DESI LRG sample (Zhou et al. 2020) due to the availability of a large suite of mocks with corresponding cuts, providing a good sample covariance matrix for reference. Similarly to Philcox et al. (2020), we process a single mock catalog in essentially the same manner as data and compare the resulting covariance with the sample covariance of clustering measurements in all available mocks, which gives a fair proxy of the pipeline performance on data and is also robust to the mismatch between data and mock clustering. We repeat the procedure multiple times taking a different catalog each time to assess the accuracy of the method, its stability, and fluctuations. In addition, we pay extra attention to the formation of a covariance matrix comparison toolkit. We focus on the meaning of the numbers used and derive reference values for the ideal case when the semi-analytical prediction matches the true underlying covariance. Due to sample variance, these expectation values can be non-trivial, and understanding the noise in the comparison measures is crucially important as well. We choose a smaller number of observables for lower noise and clearer interpretation and further project the covariances into the lower-dimensional and more physically meaningful space of model parameters.

We also note the prospects of standard reconstruction techniques that aim to reverse the large-scale displacements during the times after the drag epoch. Such subsequent evolution leads to broadening and contamination of the BAO peak; thus, undoing it sharpens the feature (Eisenstein et al. 2007). The RASCALC formalism is applicable to reconstructed 2PCF covariance as well with minor adjustments.

This paper is organized in the following manner. We review previous 2PCF estimators and covariances and discuss a modification of random counts computation and a formal extension to reconstructed data in Section 2. In Section 3, we discuss the problem of covariance matrix comparison and present our selection of methods before applying them to RASCALC validation with DESI LRG mocks in Section 4. We conclude in Section 5 by reviewing current accomplishments and future prospects. Appendix A provides more complete details on the covariance matrix estimators. Appendix B provides an overview and derivations of useful properties of covariance matrix comparison metrics.

## 2 METHODS OF COVARIANCE MATRIX ESTIMATION

We start by recapitulating the 2PCF estimators and covariance matrix formalism from O’Connell et al. (2016), O’Connell & Eisenstein (2019), and Philcox et al. (2020), with a revised notation similar to Philcox & Eisenstein (2019). In the following two sections, we discuss a slight modification for optimized disjoint random count computation and an extension for reconstructed data.

### 2.1 Overview of previous work

In a galaxy survey, we may define the 2PCF of tracers  $X$  and  $Y$  through the ratio of pair counts:

$$\hat{\xi}^{XY}(r, \mu) = \frac{N^X N^Y(r, \mu)}{R^X R^Y(r, \mu)} \quad (1)$$

(Landy & Szalay 1993), where  $\mu = \cos \theta$  is used instead of angle  $\theta$  ( $\mu$  can be restricted to  $0 \leq \mu \leq 1$  by symmetry) and  $N^X = D^X - R^X$ .  $D^X$  and  $R^X$  are (weighted) galaxies and random particles (tracing the

expected mean density) of kind  $X$ , respectively. In radial bin  $a$  and angular bin  $c$ , this estimate transforms to

$$(\hat{\xi}^{XY})_a^c = \frac{(N^X N^Y)_a^c}{(R^X R^Y)_a^c} \quad (2)$$

with

$$(N^X N^Y)_a^c = \sum_{i \neq j} n_i^X n_j^Y w_i^X w_j^Y \Theta^a(r_{ij}) \Theta^c(\mu_{ij}) \delta_i^X \delta_j^Y \quad (3)$$

$$(R^X R^Y)_a^c = \sum_{i \neq j} n_i^X n_j^Y w_i^X w_j^Y \Theta^a(r_{ij}) \Theta^c(\mu_{ij}),$$

where we have assigned a cubic grid to the survey such that each cell contains no more than one galaxy,  $n_i^X$  is the expected mean number density of tracer  $X$  in the cell  $i$ ,  $w_i$  is the expected mean weight,  $\delta_i$  is the fractional galaxy overdensity,  $\mu_{ij}$  relates to the angle between the line of sight and the separation vector  $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$  ( $r_{ij}$  being its absolute value), and  $\Theta$  are binning functions (unity if the argument fits into the bin and zero otherwise).

Given the binned 2PCF estimator [equation (2)], the covariance matrix can be computed by definition:

$$\text{cov} \left[ (\hat{\xi}^{XY})_a^c, (\hat{\xi}^{ZW})_b^d \right] = \left\langle (\hat{\xi}^{XY})_a^c (\hat{\xi}^{ZW})_b^d \right\rangle - \left\langle (\hat{\xi}^{XY})_a^c \right\rangle \left\langle (\hat{\xi}^{ZW})_b^d \right\rangle, \quad (4)$$

where  $\langle \rangle$  means ensemble average over realizations of overdensity  $\delta$ . According to equation (3), this affects the  $NN$  counts but not  $RR$ . Thus, the right-hand side of equation (4) has the product of  $RR$  counts as a constant denominator and a sum over sets of four cells (from the product of  $NN$  counts), containing an ensemble average of the product of 4  $\delta$  values, where some of the cells can coincide with each other. These ensemble averages are by definition four-point correlation functions, but same-cell overdensities handled naively give zero separation and divergent values. To overcome this issue, O'Connell et al. (2016) further expanded equation (4) into sums over distinct two-, three-, and four-cell configurations. Additionally, squares of overdensity in one cell (products of overdensities in coinciding cells) were replaced by a shot-noise approximation:

$$(\delta_i^X)^2 \approx \frac{1}{n_i^X} (1 + \delta_i^X). \quad (5)$$

As a result, only products of four, three, or two overdensities in distinct cells remained. After ensemble averaging, these give four-, three-, and two-point correlation functions at non-zero separations, respectively. Lastly, the disconnected (Gaussian) part of the four-point function can be separated from the connected (non-Gaussian) one according to Isserlis' (Wick's) theorem (Isserlis 1918). The full resulting expressions are provided in Appendix A, equation (A2).

Due to high noise in higher-point correlation function measurements and difficulties in their theoretical modelling, an alternative approach of mimicking non-Gaussianity has been established. All the higher-order correlation functions are set to zero (save for disconnected four-point, which reduces to products of two-point functions), but the shot-noise approximation is modified with a factor  $\alpha_{\text{SN}}^X > 1$ , which can have a separate value for each of different samples of galaxies:

$$(\delta_i^X)^2 \approx \frac{\alpha_{\text{SN}}^X}{n_i^X} (1 + \delta_i^X). \quad (6)$$

This increases the correlations on the smallest scales, which is similar to where non-Gaussian effects are the strongest.

The sums can be transformed to continuous form by changing sums to integrals over positions and replacing cell quantities with continuous functions in 3D space. However, it is convenient to leave them discrete, which allows us to estimate them using importance sampling directly from the random catalog (Philcox et al. 2020), without the need to write functional forms for survey number density, weights, and so on. Two-point function values for pairs of points are interpolated from a grid/table to sampled pair separation  $r_{ij}$ ,  $\mu_{ij}$  (the latter can be computed with respect to the mid-point radius-vector of the pair, or a fixed axis) with the bicubic method, which is in practice based on radially and angularly binned 2PCF estimates. A special iterative correlation function rescaling procedure is used to modify the values of the correlation function on the interpolation grid such that the bin-averaged values of the interpolation result resemble the binned 2PCF estimates more closely.

Originally, O'Connell et al. (2016) fit the shot-noise rescaling to a sample covariance obtained from a smaller set of mocks, with the idea that lower precision was sufficient for obtaining just one parameter as opposed to estimating the whole matrix. O'Connell & Eisenstein (2019) proposed that a jackknife covariance from the data itself can be used instead, eliminating the dependence on mocks completely. Noting that jackknife has issues in the cosmological application, they developed a separate estimator for this covariance taking into account correlations between different estimates.

We follow a modified formalism from Philcox et al. (2020) called *unrestricted jackknife*. According to it, the jackknife correlation function estimate  $\xi_A$  is the cross-correlation function between jackknife region number  $A$  and the whole survey. In other words, an additional weight of particle  $i$  (denoted by  $q_i^A$ ) is equal to one if it belongs to the region  $A$  and zero if not. A pair of particles is additionally weighted by the mean of their weights: one if both belong to the region  $A$ , one half if only one does and zero if both are out of it. Conveniently, the sum of these weights for any selected pair over all jackknife regions is unity. As a consequence, the mean of jackknife 2PCF estimates (weighted by  $RR$  counts) is equal to the full 2PCF estimate. Then an estimator for jackknife covariance is worked out separately [equation (A7)], with a similar shot-noise rescaling procedure.

In principle, shot-noise rescaling [equation (6)] can be different for each tracer, and it can be obtained by fitting the prediction for the jackknife covariance of its auto-correlation function to the data-based jackknife estimate. The resulting  $\alpha_{\text{SN}}^X$  value(s) is used together with the full covariance estimator [equation (A1)] for the final result.

Let us reiterate the key approximation: non-Gaussianity can be mimicked by rescaling shot noise while dropping the terms with higher-order correlation functions. It works because the primary effect of non-Gaussian contributions is an additional correlation at small distances, typically smaller than the bin width of 2PCF used in actual fits and requiring a covariance. Enhancing the shot noise results in increased correlation on infinitely small scales. This should remain a good estimate as long as the correlation functions' contribution to covariances on scales of interest is dominated by their squeezed limits. Whether this is the case is not clear generally, but O'Connell et al. (2016) reported good agreement with large sets of mock catalogs achieved with this method, and Vargas-Magaña et al. (2018) demonstrated the applicability of the approach to BAO analysis.

It is important to specify the means of fitting covariance matrices. Following Philcox et al. (2020), we choose to optimize the Kullback-Leibler (KL) divergence between the RASCALC jackknife precision  $\tilde{\Psi}_J(\alpha_{\text{SN}})$  and the data-based jackknife covariance estimate [given by

equation (A6),  $D_{\text{KL}}[\tilde{\Psi}_J(\alpha_{\text{SN}}), \mathbf{C}_J]$ , where

$$D_{\text{KL}}(\Psi_1, \mathbf{C}_2) = \frac{1}{2} [\text{tr}(\Psi_1 \mathbf{C}_2) - N_{\text{bins}} - \ln \det(\Psi_1 \mathbf{C}_2)]. \quad (7)$$

In this equation,  $N_{\text{bins}}$  is the dimension of covariance matrices (number of correlation function bins). Inversion of the RASCALC covariance has a certain bias. Since it is not expected to obey Wishart statistics like the (mock) sample covariance, the Hartlap factor (Hartlap, Simon & Schneider 2007) is not relevant. A special second-order bias correction has been derived in O’Connell & Eisenstein (2019):

$$\tilde{\Psi} = (\mathbb{I} - \tilde{\mathbf{D}})\tilde{\mathbf{C}}^{-1}$$

$$\tilde{\mathbf{D}} = \frac{N_{\text{subsamples}} - 1}{N_{\text{subsamples}}} \left[ -\mathbb{I} + \frac{1}{N_{\text{subsamples}}} \sum_{i=1}^{N_{\text{subsamples}}} \tilde{\mathbf{C}}_{[i]}^{-1} \tilde{\mathbf{C}}_i \right], \quad (8)$$

which uses the partial covariance estimates  $\tilde{\mathbf{C}}_i$  from  $N_{\text{subsamples}}$  distinct sets of configurations resulting from importance sampling in the estimation of sums [equation (A2) or (A8)] and mean of all the partial estimates but the  $i$ ’th  $\tilde{\mathbf{C}}_{[i]}$ . Thus, the correction is applicable for both jackknife and full covariance.

## 2.2 Split random-random computation

Keihänen et al. (2019) showed that splitting the random catalog into a number of sub-catalogs of the same size as the data catalog when calculating random–random pairs and excluding pairs across different sub-catalogs provides the optimal error at a fixed computational cost. The splitting can be used in RASCALC. It gives little to no speed-up and impact on results because the importance sampling is too far from complete. However, it can be useful for multinode parallelization. This approach has been used for the data-based RASCALC computation in Moon et al. (2023).

A robust implementation of split random–random pair calculations in RASCALC would require considering only quadruples of random points where members of each pair are from the same sub-catalog, but the pairs can be from different catalogs. However, this has been found to have little to no impact on the results, probably due to the fact that importance sampling covers only a small fraction of all possible configurations. At the same time, such implementation makes the code less efficient and makes it impossible to split the computation of different catalogs between nodes.

## 2.3 Reconstructed two-point function covariance

After standard reconstruction, a common approach is to replace  $N = (D - R)$  by  $N = (D - S)$  ( $S$  being the random point with position shifted in the same manner as data) in the Landy–Szalay estimator [equation (1)], leaving  $RR$  in the denominator, so that it is

$$\hat{\xi}^{XY} = \frac{D^X D^Y - D^X S^Y - S^X D^Y + S^X S^Y}{R^X R^Y}, \quad (9)$$

instead of

$$\hat{\xi}^{XY} = \frac{D^X D^Y - D^X R^Y - R^X D^Y + R^X R^Y}{R^X R^Y}. \quad (10)$$

This means shifted randoms are to be used in sums or integrals representing  $NN$ . These eventually form the sums (or integrals) for  $C$  terms. Thus, strictly speaking, the procedure for reconstructed 2PCF should be:

(i) use shifted randoms for sampling, corresponding to the numerator of (9);

(ii) provide a differently normalized 2PCF as input, namely

$$\hat{\xi}_{\text{in}}^{XY} = \frac{D^X D^Y - D^X S^Y - S^X D^Y + S^X S^Y}{S^X S^Y}, \quad (11)$$

since non-shifted randoms do not appear in the sampling procedure;

(iii) use non-shifted random counts for denominator in equation (A2), or correction function [equation (A15)] in Legendre case, corresponding to the denominator of [equation (9)].

Shifted randoms are individual for each mock catalog. Therefore, they can not be defined clearly for mock-averaged computations. In those cases, we continue to use the non-shifted randoms everywhere for consistency.

## 3 METHODS OF COMPARISON OF COVARIANCE MATRICES

Since a covariance matrix is a high-dimensional object, it can be hard to explore and interpret. Moreover, we run the pipeline multiple times independently and aim to study all the covariance matrix products to assess their stability and fluctuations. Thus, compact and numerical comparison measures are instructive.

### 3.1 Interpretable measures of similarity for covariance matrices

The first characteristic we consider is the KL divergence, a measure of distance between distributions used to fit covariances in RASCALC [Section 2.1, equation (7)]. It is generally defined as an expectation value of the logarithm of the ratio of the two probability distribution functions according to the first distribution:

$$D_{\text{KL}}(P_1 || P_2) = \int \ln \left( \frac{P_1(x)}{P_2(x)} \right) P_1(x) dx \quad (12)$$

By this expression, KL divergence can be seen as an average difference in log-likelihood. For two Gaussian distributions with covariance matrices  $\mathbf{C}_i$  and precision matrices  $\Psi_i = \mathbf{C}_i^{-1}$  describing  $N_{\text{bins}}$  observables (correlation function bins in our setup), it can be found as

$$D_{\text{KL}}(\Psi_1, \mathbf{C}_2) = \frac{1}{2} [\text{tr}(\Psi_1 \mathbf{C}_2) - N_{\text{bins}} - \ln \det(\Psi_1 \mathbf{C}_2)]. \quad (13)$$

This is the expression we will use. O’Connell et al. (2016) show that the KL divergence is related to the log-likelihood if the covariance matrix is estimated from a sample with multivariate normal distribution, which is a very good approximation for correlation function bins, and additionally assuming that the precision matrix is the inverse of the true covariance matrix, characterizing the multivariate normal distribution of measured quantities. This is very appropriate for testing the hypothesis that the RASCALC precision matrix is a precise unbiased estimate.

The next metric assesses how close the first precision matrix is to the inverse of the second covariance matrix, and at the same time, a ‘directional’ root-mean-square relative difference in  $\chi^2$  given by the two covariance matrices (explained in more detail in Appendix B2):

$$R_{\text{inv}}(\Psi_1, \mathbf{C}_2) = \frac{1}{\sqrt{N_{\text{bins}}}} \left\| \mathbf{C}_2^{1/2} \Psi_1 \mathbf{C}_2^{1/2} - \mathbb{I} \right\|_F$$

$$= \sqrt{\frac{\text{tr}[(\Psi_1 \mathbf{C}_2 - \mathbb{I})^2]}{N_{\text{bins}}}}. \quad (14)$$

This measure can also be seen as the average relative difference in the errorbars. Moreover, if the covariance matrix is estimated from



a sample with multivariate normal distribution and the precision matrix is assumed to be true,  $R_{\text{inv}}^2$  is proportional to the  $\chi^2$  computed using covariance of independent covariance matrix elements (Appendix B2). Thus, in this case, it can serve as an approximation of log-likelihood for optimization.

The last metric is akin to the mean reduced  $\chi^2$  of samples corresponding to one covariance matrix with respect to the other precision matrix:

$$\chi_{\text{red}}^2(\Psi_1, \mathbf{C}_2) = \frac{1}{N_{\text{bins}}} \text{tr}(\Psi_1 \mathbf{C}_2). \quad (15)$$

It can be seen as the mean ratio of  $\chi^2$  given by the two covariance/precision matrices.

All three metrics are not symmetric, meaning that values for  $(\Psi_1, \mathbf{C}_2)$  and  $(\Psi_2, \mathbf{C}_1)$  may be different, so in principle, it might be informative to consider differences both ways. On the other hand, the sample covariance is less robust than the RASCALC result and its inversion can be less stable. Moreover, computing each metric twice makes the results more numerous and less clear. Finally, the KL divergence [equation (13)] is expected to lose its log-likelihood sense if computed between the sample precision and model covariance, since the latter one does not necessarily follow Wishart distribution. Therefore, we decided to limit ourselves to RASCALC precision matrices and sample covariance matrices.

For a better understanding of the metrics, let us consider the eigenvalues of  $\Psi_1 \mathbf{C}_2$  (alternatively, one can use  $\Psi_1^{1/2} \mathbf{C}_2 \Psi_1^{1/2}$ , which is symmetric) and denote them as  $\lambda_a$ . We would like  $\Psi_1 \rightarrow \mathbf{C}_2^{-1}$  thus all  $\lambda_a \rightarrow 1$ . The metrics then can be expressed as

$$D_{\text{KL}}(\Psi_1, \mathbf{C}_2) = \frac{1}{2} \sum_{a=1}^{N_{\text{bins}}} [\lambda_a - 1 - \ln \lambda_a] \approx \frac{1}{4} \sum_{a=1}^{N_{\text{bins}}} (\lambda_a - 1)^2, \quad (16)$$

$$R_{\text{inv}}(\Psi_1, \mathbf{C}_2) = \sqrt{\frac{1}{N_{\text{bins}}} \sum_{a=1}^{N_{\text{bins}}} (\lambda_a - 1)^2}, \quad (17)$$

$$\chi_{\text{red}}^2(\Psi_1, \mathbf{C}_2) = \frac{1}{N_{\text{bins}}} \sum_{a=1}^{N_{\text{bins}}} \lambda_a. \quad (18)$$

Thus,  $D_{\text{KL}}$  and  $R_{\text{inv}}$  accumulate any deviation of  $\lambda_a$  from 1, although they cannot indicate the direction of such differences. Note that the quadratic expression for  $D_{\text{KL}}$  is approximate so it is not generally degenerate with  $R_{\text{inv}}$ , although as the covariance matrices approach each other, these two measures become more redundant:

$$D_{\text{KL}}(\Psi_1, \mathbf{C}_2) \approx \frac{N_{\text{bins}}}{4} R_{\text{inv}}^2(\Psi_1, \mathbf{C}_2). \quad (19)$$

$\chi_{\text{red}}^2$  can show which covariance matrix is ‘larger’ on average, while deviations in opposite directions may cancel each other.

Next, we would like to understand what to expect from these metrics. For this purpose, we consider the case when the precision matrix is predicted perfectly ( $\Psi_0$ ), ideally matching the true underlying covariance ( $\mathbf{C}_0 = \Psi_0^{-1}$ ) and focus on the noise properties of the sample covariance matrix  $\mathbf{C}_S$  obtained via the standard unbiased estimator for the case when the true mean is not known:

$$C_{S,ab} = \frac{1}{n_S - 1} \sum_{i=1}^{n_S} (\xi_{a,i} - \bar{\xi}_a) (\xi_{b,i} - \bar{\xi}_b) \quad (20)$$

where  $a, b$  denote bin numbers,  $i, j$  index sample numbers, and  $\bar{\xi}_a$  is the estimate of the mean:

$$\bar{\xi}_a \equiv \frac{1}{n_S} \sum_{i=1}^{n_S} \xi_{a,i}. \quad (21)$$

Since the clustering measurements are described well by a multivariate normal distribution, their sample covariance matrix follows the Wishart statistics. This provides a reference of how the metrics behave when the perfect precision matrix is compared to a covariance matrix estimated from  $n_S$  samples with  $N_{\text{bins}}$  bins (or any other Gaussian observables). Full derivations are presented in Appendix B; here, we will only provide the results for mean/expectation values and standard deviations:

$$\begin{aligned} \langle D_{\text{KL}}(\Psi_0, \mathbf{C}_S) \rangle &\approx \frac{N_{\text{bins}}(N_{\text{bins}} + 1)}{4(n_S - 1)}, \\ \sigma[D_{\text{KL}}(\Psi_0, \mathbf{C}_S)] &\approx \frac{1}{2} \sqrt{\frac{N_{\text{bins}}[(N_{\text{bins}} + 1)(n_S + 2N_{\text{bins}} + 2) + 2]}{(n_S - 1)^3}}, \end{aligned} \quad (22)$$

$$\begin{aligned} \langle R_{\text{inv}}(\Psi_0, \mathbf{C}_S) \rangle &\approx \sqrt{\frac{N_{\text{bins}} + 1}{n_S - 1}}, \\ \sigma[R_{\text{inv}}(\Psi_0, \mathbf{C}_S)] &\approx \frac{1}{n_S - 1} \sqrt{\frac{(N_{\text{bins}} + 1)(n_S + 2N_{\text{bins}} + 2) + 2}{N_{\text{bins}}(N_{\text{bins}} + 1)}}. \end{aligned} \quad (23)$$

Naïvely, one could expect  $D_{\text{KL}}$  and  $R_{\text{inv}}$  to become arbitrarily small as  $\Psi_1 \rightarrow \Psi_0$ . However, in reality, they can have large expectation values, especially as the number of bins increases.

$\chi_{\text{red}}^2$ , however, would behave like the reduced  $\chi^2$  with  $N_{\text{bins}} \times (n_S - 1)$  degrees of freedom in this case (see Appendix B3):

$$\begin{aligned} \langle \chi_{\text{red}}^2(\Psi_0, \mathbf{C}_S) \rangle &= 1, \\ \sigma[\chi_{\text{red}}^2(\Psi_0, \mathbf{C}_S)] &= \sqrt{\frac{2}{N_{\text{bins}}(n_S - 1)}}. \end{aligned} \quad (24)$$

It might seem like  $D_{\text{KL}}$  and  $R_{\text{inv}}$  could be unbiased by multiplying one of the matrices by a factor similar to the Hartlap factor (Hartlap et al. 2007), but a lack of bias in  $\chi_{\text{red}}^2$  expectation value suggests that this is not true.

It is notable that a deviation of  $\chi_{\text{red}}^2$  from 1 would contribute to  $R_{\text{inv}}$  – see expressions (18) and (17); their consequence is also that

$$|\chi_{\text{red}}^2(\Psi_1, \mathbf{C}_2) - 1| \leq R_{\text{inv}}(\Psi_1, \mathbf{C}_2). \quad (25)$$

However, the direction of such deviation will not be clear in  $R_{\text{inv}}$ , and non-trivial expectation value can make it harder to interpret. This keeps  $\chi_{\text{red}}^2$  useful in many cases.

### 3.2 Internal convergence assessment

Internal consistency of RASCALC covariance matrices in one run and the convergence of the Monte Carlo integration procedure are also important to assess quantitatively. We propose to employ the above-mentioned methods to accomplish this and provide valuable diagnostics that do not rely on a reference (e.g. sample) covariance and thus can be used in any run, including the pure data-based one. However, we need to note that such a test can only quantify limited sources of uncertainty or error, leaving aside the factors like adequacy of the approximations in the formalism, the precision of the input clustering, and noise in the jackknife covariance estimated from the data.

The RASCALC code provides multiple partial intermediate results corresponding to practically non-overlapping sets of quadruples, triples, and pairs of points. These resulting covariance matrices can be split into two distinct sets of similar size, averaged within them, and compared using the three metrics. In this case, however, the

arguments for the  $\chi_{\text{red}}^2$  weaken – we can expect  $R_{\text{inv}}$  to become arbitrarily low as the number of Monte Carlo samples increases, which would limit the reduced chi-squared via equation (25), and it is not as interesting to understand which of the halves gives a ‘smaller’ matrix. Then  $D_{\text{KL}}$  also becomes more redundant with  $R_{\text{inv}}$  via equation (19). Therefore, it is reasonable to only show  $R_{\text{inv}}$ , which can be seen as an estimate of root-mean-square relative precision (considered over all directions in measurement space).

## 4 APPLICATION TO DESI LRG MOCKS

In this section, we use the described methods on DESI-M2 mocks to assess the performance and stability of the approach on the actual data set. We describe the setup first, then perform intrinsic validation described in Section 4.3, look at the shot-noise rescaling values resulting from jackknife calibration used for the final covariance estimates, validate the RASCALC results by comparison with the mock sample covariance in measurement/observable and parameter space, and finally focus specifically on errorbars on BAO scale.

### 4.1 Mock catalogs and reconstruction method

We use the 999 effective Zel’dovich (EZ) mocks (Chuang et al. 2015; Zhao et al. 2021) with cuts corresponding to the DESI LRG sample [Zhou et al. 2020; described in more detail in Moon et al. (2023)], which will be referred to as DESI-M2 `Firstgen` EZ mocks. Sample covariance based on these does not provide a perfect reference because both the number of mock catalogs and the level of details in each simulation are limited, but the best one can have realistically since increasing one without making the other worse would require even more significant computational resources. Comparing these is also robust to the mismatch between data and mock clustering.

The reconstruction method is also the same as in Moon et al. (2023): the iterative procedure (Burden, Percival & Howlett 2015) implemented in the `IterativeFFTR` reconstruction algorithm of the `PYRECON` package<sup>2</sup> with the `RecIso` convention. Three iterations are used with a Gaussian smoothing kernel of width  $15 h^{-1}\text{Mpc}$ . An approximate growth rate and the expected bias are assumed.

### 4.2 Setup

For this study, we have performed separate runs using 2PCF measured from single LRG DESI-M2 `Firstgen` EZ mocks catalogs. This has been repeated 10 times for pre- and post-recon. In the latter case, individual shifted random catalogs have been used for each mock, following the procedure we described in Section 2.3.

Pre-reconstruction galaxies and randoms were assigned unity weights, for post-reconstruction FKP weights (Feldman, Kaiser & Peacock 1994) were used, given by

$$w_{\text{FKP}} = \frac{1}{1 + n(z)CP_0} \quad (26)$$

where  $n(z)$  is the weighted number density (per volume),  $C$  is the mean completeness for the sample, and  $P_0$  is a fiducial power-spectrum amplitude. For LRG,  $C = 0.579$  and  $P_0 = 10^4 (h^{-1}\text{Mpc})^3$

(Moon et al. 2023). We note that the weighting schemes are not exactly the same as for real data, but since weights are included explicitly in the covariance estimators, we expect RASCALC to work with any fixed choice applied consistently for 2PCF measurements and Monte Carlo integration.

For the importance sampling input, 10 random catalogs were used in pre-recon computations and 20 in post-recon, like for the `RR (SS)` pair counts computation for the 2PCF estimates. These randoms have been concatenated before being provided to RASCALC executable. We assign 60 jackknife regions assigned by a  $K$ -means subsampler based on data positions (but not weights) as in DESI-M2 data, compute the jackknife covariance matrix, and use it to calibrate the shot-noise rescaling.

We note that some validation has been performed in Moon et al. (2023): RASCALC covariance matrix based on 2PCF averaged over all LRG DESI-M2 `Firstgen` EZ mocks with no shot-noise rescaling applied (and using unshifted randoms in the post-reconstruction case) has been compared to the sample covariance matrix in terms of  $\chi^2$  of BAO fits, best-fit values and standard deviations of BAO isotropic scale parameter  $\alpha$ , yielding a good agreement. However, there are significant limitations to this approach:

- (i) effect of noise in the input clustering is significantly smaller in 2PCF averaged over  $\approx 1000$  mocks than in the real data, which is close to a single mock catalog;
- (ii) possible differences in other parameters of the BAO model or more generic aspects of correlation function have not been assessed.

Single-mock runs address the first issue since each of them is a fair proxy of the data. The use of covariance matrix comparison metrics from Section 3.1 expands on the second one. We keep the result with mock-average 2PCF, labeled ‘Average G’ (Gaussian), to assess the importance of precision of input clustering.

We also consider a shot-noise-rescaled version of the run with mock-averaged clustering, labelled ‘Average NG’ (non-Gaussian). We note that calibration of an all-mocks run on a jackknife estimate can be ambiguous or require a repeated computation of all the pair counts with jackknives, which was not done before because jackknives are not necessary for the sample covariance. Thus we choose to fit the full covariance matrix to the mock sample covariance by minimizing the KL divergence between them [analogously to the jackknife procedure described in Section 2.1 and equation (7)]. Since only one parameter is varied in the fit, a perfect agreement is still not guaranteed. On the other hand, this setup is clearly idealized and would be closer to the closest possible match to the mock covariance the RASCALC method can provide. It comprises another useful reference to compare to the data-like performance on single mocks.

We only consider 45 radial bins, spanning  $4 h^{-1}\text{Mpc}$  each from 20 to  $200 h^{-1}\text{Mpc}$ . The RASCALC covariances are produced with single angular bins, which is a simplifying assumption since treating monopole more precisely in Legendre mode with shot-noise rescaling would require 2 runs per data set, as explained in Appendix A3. The 2PCF measurements for the mock sample covariance use a more precise monopole estimate provided by `PYCORR`.<sup>3</sup> We also project them into a BAO model parameter space using the derivatives near the best fit (Fisher forecast). The model uses only a separation range from 48 to  $148 h^{-1}\text{Mpc}$ .

<sup>2</sup><https://github.com/cosmodesi/pyrecon> (Arnaud de Mattia, Martin J. White, Julian E. Bautista, Pedro Rangel Caetano, Sesh Nadathur, Enrique Pailles, Grant Merz, Davide Bianchi)

<sup>3</sup><https://github.com/cosmodesi/pycorr> (Arnaud de Mattia, Lehman Garrison, Manodeep Sinha, Davide Bianchi, Svyatoslav Trusov, Enrique Pailles, Seshadri Nadathur, Craig Warner, James Lasker)

**Table 1.** Intrinsic pre- and post-reconstruction convergence test results in measurement space.  $R_{\text{inv}}$  estimate root-mean-square relative precision, averaged over all different directions in the measurement space. The numbers provided here are for the full covariance, but the consistency levels of the jackknife covariance prediction from RASCALC are similar. They demonstrate sub-percent stability in RASCALC integrals and ensure that random noise in importance sampling is not a significant source of error, as these numbers are smaller than deviations observed in further comparisons (Tables 3–6).

Mock no.	$R_{\text{inv}}$ pre	$R_{\text{inv}}$ post
Average G	$1.3 \times 10^{-3}$	$3.0 \times 10^{-3}$
Average NG	$1.2 \times 10^{-3}$	$8.1 \times 10^{-3}$
1	$3.7 \times 10^{-3}$	$2.3 \times 10^{-3}$
2	$3.1 \times 10^{-3}$	$2.2 \times 10^{-3}$
3	$3.8 \times 10^{-3}$	$2.1 \times 10^{-3}$
4	$6.3 \times 10^{-3}$	$1.9 \times 10^{-3}$
5	$4.4 \times 10^{-3}$	$1.3 \times 10^{-3}$
6	$4.4 \times 10^{-3}$	$1.9 \times 10^{-3}$
7	$3.7 \times 10^{-3}$	$2.0 \times 10^{-3}$
8	$3.3 \times 10^{-3}$	$1.7 \times 10^{-3}$
9	$3.5 \times 10^{-3}$	$2.5 \times 10^{-3}$
10	$3.5 \times 10^{-3}$	$2.2 \times 10^{-3}$

### 4.3 Internal convergence checks

We perform an intrinsic diagnostic procedure (as described in Section 3.2) to ensure that RASCALC integrals converged well in each run and exclude importance sampling random noise from significant error factors. We found that pre-recon mock 6 and post-recon mock 5 showed significantly worse consistency than all the rest. Therefore, we have run them twice longer.

After that, all the RASCALC results have reached a high and quite uniform level of internal consistency, as presented in Table 1. We only show  $R_{\text{inv}}$ , which are easier to interpret as the root-mean-square relative deviation between different partial estimates of the covariance matrix (considered over all directions in measurement space).  $\chi_{\text{red}}^2$  are limited via equation (25), and  $D_{\text{KL}}$  values are quite close to estimates from equation (19). Due to high consistency in measurement space, we have not performed projection to parameter space here. The splitting of Monte Carlo subsamples has been done in a few different ways and the non-symmetric metric has been computed both ways ( $\Psi_1 C_2$  and  $\Psi_2 C_1$ ), but all the values were very close,<sup>4</sup> and thus have been averaged to one number for each metric. These low (sub-percent) internal deviations give us confidence that the Monte Carlo integration procedure in RASCALC has converged well, and it will not be a significant error source in further comparison. After ascertaining this, we have not touched the covariance matrix products to be fair – with real survey and no mocks, other validation procedures described in this paper are not available.

### 4.4 Shot-noise rescaling values

Next, we look into the shot-noise rescaling values because the final covariance estimates (with approximate non-Gaussianity) are based on them. The shot-noise rescaling values for single mocks are obtained by fitting the separate RASCALC jackknife covariance

<sup>4</sup>In the case of disjoint running (Section 2.2), there can be a meaningful difference between splittings within or between different random sub-catalogs, due to fluctuations in pair counts between these. But here all the randoms have been concatenated together.

**Table 2.** Shot-noise rescaling values for the 10 mocks, on which the final covariance predictions are based, pre- and post-reconstruction. The ‘Average NG’ value is fit to mock sample covariance, and the 1–10 are fit to jackknife covariance estimates from single mocks, and the resulting numbers are consistent. The standard deviation for single mocks is  $\approx 2$  per cent before and  $\approx 1$  per cent after reconstruction, which translate into a similar effect on relative precision of the final covariance matrices due to rescaling.

Mock no.	Pre-recon $\alpha_{\text{SN}}$	Post-recon $\alpha_{\text{SN}}$
Average NG	1.096	1.038
1	1.096	1.062
2	1.074	1.043
3	1.040	1.034
4	1.077	1.051
5	1.079	1.033
6	1.089	1.030
7	1.080	1.041
8	1.102	1.058
9	1.116	1.033
10	1.080	1.041
1–10 mean $\pm$ std	$1.083 \pm 0.020$	$1.043 \pm 0.011$

prediction to the jackknife covariance estimate for each mock. For the mock-average clustering, the full RASCALC covariance was fit to the mock sample covariance instead, as discussed in Section 4.2.

The shot-noise rescaling values are gathered in Table 2. We note that all of them are greater than one (which corresponds to purely Gaussian covariance), in accordance with our expectation that the non-Gaussianity expands the errorbars. Moreover, the mean shot-noise rescaling of the 10 single mocks is  $\approx 4$  standard deviations larger than 1 both before and after reconstruction. The values obtained from jackknife and mock covariance are consistent. After reconstruction, the shot-noise rescaling decreases for every mock. The pre-recon mean is larger than the post-recon one by  $\approx 1.8$  standard deviations. The scatter after reconstruction is also smaller than before. These deviations can be caused by the random fluctuations in the input 2PCF estimates, noise in jackknife covariances, and differences in shifted randoms (for post-recon only).

The key conclusion is that we have obtained the shot-noise rescaling parameter for a data-like setup (single mock runs) with a per cent-level precision. This maps into a similar or smaller relative deviation in the rescaled covariance matrices since the two-point term has the strongest scaling,  $\propto \alpha_{\text{SN}}^2$ , and the four-point term remains the same [equation (A3)].

### 4.5 Measurement-space validation

Now we proceed to comparison with the sample covariance matrices as reference, keeping in mind they are not devoid of noise so not all the comparison measures can be ideal. We consider the higher-dimensional space of observables first, where the effects of sample variance are quite significant. It consists of 45 bins of 2PCF, spanning 20–200  $h^{-1}$  Mpc linearly with a bin width of 4  $h^{-1}$  Mpc.

Sample covariances for all original bins have been estimated using  $n_s = 999$  2PCF measurements from all the DESI-M2 Firstgen EZ mocks and the standard unbiased estimator [equation (20)]. The procedures have been similar for pre- and post-recon.

The comparison measures between the RASCALC precision matrices [estimated via equation (8)] and the sample covariance matrices have been computed and are presented in Table 3 for pre-reconstruction and Table 4 for post-reconstruction. First of all, there are fluctuations in the comparison measures, involving

**Table 3.** Results of general measurement-space comparison between the RASCALC results ( $R$ ) and mock sample covariance ( $S$ ) before reconstruction. The last row provides the perfect-case reference – expectation values and standard deviations for the three metrics, if the RASCALC precision matrix truly described the distribution of the mock correlation functions.

Mock no.	$D_{\text{KL}}(\Psi_R, C_S)$	$R_{\text{inv}}(\Psi_R, C_S)$	$\chi_{\text{red}}^2(\Psi_R, C_S)$
Average G	0.793	0.2922	1.1422
Average NG	0.537	0.2136	0.9906
1	0.721	0.2252	0.9600
2	0.602	0.2234	0.9992
3	0.571	0.2329	1.0477
4	0.548	0.2191	0.9991
5	0.755	0.2302	0.9830
6	0.727	0.2315	0.9723
7	0.695	0.2254	0.9675
8	0.530	0.2085	0.9695
9	0.610	0.2140	0.9291
10	0.518	0.2106	0.9895
1–10	$0.628 \pm 0.089$	$0.2221 \pm 0.0087$	$0.982 \pm 0.031$
Perfect $\Psi$	$0.519 \pm 0.024$	$0.2147 \pm 0.0049$	$1.0000 \pm 0.0067$

**Table 4.** Results of general measurement-space comparison between RASCALC results ( $R$ ) and mock sample covariance ( $S$ ) after reconstruction. The last row provides the perfect-case reference – expectation values and standard deviations for the three metrics, if the RASCALC precision matrix truly described the distribution of the mock correlation functions.

Mock no.	$D_{\text{KL}}(\Psi_R, C_S)$	$R_{\text{inv}}(\Psi_R, C_S)$	$\chi_{\text{red}}^2(\Psi_R, C_S)$
Average G	0.62	0.247	1.0653
Average NG	0.57	0.225	1.0028
1	0.94	0.256	0.9463
2	0.63	0.229	0.9725
3	0.72	0.266	1.0031
4	0.61	0.222	0.9657
5	0.62	0.229	0.9924
6	0.65	0.231	0.9817
7	0.62	0.226	0.9706
8	0.61	0.222	0.9631
9	0.77	0.281	1.0214
10	0.88	0.314	1.0162
1–10	$0.70 \pm 0.12$	$0.247 \pm 0.031$	$0.983 \pm 0.024$
Perfect $\Psi$	$0.519 \pm 0.024$	$0.2147 \pm 0.0049$	$1.0000 \pm 0.0067$

the single-mock results, stemming from the input 2PCF estimates, jackknife covariances, and differences in shifted randoms (for post-recon only) – the same causes as for scatter in  $\alpha_{\text{SN}}$  discussed in Section 4.4. The individual pre-reconstruction covariances appear to agree with the mock sample covariance better than the post-reconstruction. The covariance with mock-averaged clustering and no shot-noise rescaling, on the contrary, gives a closer agreement after reconstruction. Before reconstruction, any individual shot-noise rescaled covariance shows better agreement than the Gaussian mock-averaged clustering run; after reconstruction, it is very often worse. With mock-averaged clustering, the shot-noise rescaling is clearly beneficial for pre-reconstruction and less so for post-reconstruction. This may be a hint that the shot-noise rescaling might not be doing as well after reconstruction as before.

Compared to the perfect case, RASCALC typically performs worse (higher  $D_{\text{KL}}$  and  $R_{\text{inv}}$ , reduced chi-squared further from one), which we expect since the code (and the mocks) involve (different) approximations. We note that on average the comparison metrics

are within a couple of standard deviations of the expectation value for the true underlying covariance matrix. On the other hand, it is the larger standard deviation in RASCALC results that is allowing this conclusion, and reducing the noise factors causing it (input 2PCF fluctuations, single jackknife covariance) may allow us to reach a closer agreement in the future works.

#### 4.6 Parameter-space validation

In this section, we project the covariance into a lower-dimensional and more physically meaningful space of BAO model parameters. Lower dimensionality makes the reference values for comparison metrics clearer and the results become easier to interpret. In addition, the correlation function modes that are not physically possible or do not affect the parameter constraints are removed from consideration, which leaves only real and important ‘directions’ for consideration.

We choose a commonly used BAO model<sup>5</sup> (Ross et al. 2017; Ata et al. 2018) with a scalable template  $\xi_0$  and three nuisance polynomial terms:

$$\xi_{\text{mod}}(r) = B\xi_0(\alpha_{\text{BAO}}r) + A_0 + A_1/r + A_2/r^2, \quad (27)$$

comprising  $N_{\text{pars}} = 5$  parameters:  $B, A_0, A_1, A_2, \alpha_{\text{BAO}}$ .

Instead of performing full fits, we use Fisher matrix formalism. This can be seen as less precise than full fits on every mock or MCMC using a 2PCF likelihood. On the other hand, the parameter distribution is not Gaussian when the model is not a linear function of parameters, which makes linear approximation within Fisher matrix formalism more suitable for the comparison methods we have discussed.

We estimate the parameter covariance matrix as the inverse of the Fisher matrix:

$$\mathbf{C}_S^{\text{par}} = \frac{n_S - 1}{n_S - N'_{\text{bins}} + N_{\text{pars}} - 1} \left[ \mathbf{M}(\mathbf{C}_S^{\text{meas}})^{-1} \mathbf{M}^T \right]^{-1} \quad (28)$$

where the measurement-space mock sample covariance matrix  $\mathbf{C}_S^{\text{meas}}$  is cut to the  $N'_{\text{bins}} = 25$  bins spanning separations from 48 to 148  $h^{-1}\text{Mpc}$  used in BAO fits, and  $\mathbf{M}$  is the matrix of derivatives of binned 2PCF vector  $\xi$  with respect to parameters  $\mathbf{p}$ :

$$M_{ca} \equiv \frac{\partial \xi_a}{\partial p_c}. \quad (29)$$

The derivatives have been taken at the best-fit parameters for the mock-averaged clustering measurements (separate before and after reconstruction).

Note that equation (28) is scaled by a correction factor according to equation (B6) in Paillas et al. (2023) to account for biases caused by both matrix inversions. This provides an unbiased (although not noiseless) estimate of the true underlying covariance in parameter space as validated in Appendix B4.

A similar but simpler procedure was performed with RASCALC products:

$$\Psi_{R,cd}^{\text{par}} \approx \mathbf{M}(\mathbf{C}_R^{\text{meas}})^{-1} \mathbf{M}^T, \quad (30)$$

where  $\mathbf{C}_R^{\text{meas}}$  was also cut to the 25 bins spanning separations from 48 to 148  $h^{-1}\text{Mpc}$  used in BAO fits. There is a bias correction matrix  $\mathbf{D}$  for RASCALC [equation (8)], but for the results presented here absolute values of its eigenvalues are  $\lesssim 10^{-3}$  thus we have decided to neglect this correction factor.

<sup>5</sup>[https://github.com/cosmodesi/BAOfit\\_xs/](https://github.com/cosmodesi/BAOfit_xs/) (Ashley J. Ross, Juan Mena Fernandez)



**Table 5.** Results of BAO parameter space ( $B$ ,  $A_0$ ,  $A_1$ ,  $A_2$ , and  $\alpha_{\text{BAO}}$ ) comparison between RASCALC results ( $R$ ) and mock sample covariance ( $S$ ) before reconstruction. The last row provides the perfect-case reference – expectation values and standard deviations for the three metrics, if the RASCALC precision matrix truly described the distribution of the mock correlation functions.

Mock no.	$D_{\text{KL}}(\Psi_R, C_S)$	$R_{\text{inv}}(\Psi_R, C_S)$	$\chi_{\text{red}}^2(\Psi_R, C_S)$
Average G	0.026	0.138	0.992
Average NG	0.027	0.135	0.925
1	0.104	0.236	0.855
2	0.065	0.193	0.911
3	0.022	0.122	0.941
4	0.013	0.099	0.962
5	0.127	0.250	0.883
6	0.091	0.232	0.866
7	0.095	0.229	0.861
8	0.025	0.128	0.918
9	0.042	0.165	0.876
10	0.016	0.107	0.944
1–10	$0.060 \pm 0.042$	$0.176 \pm 0.059$	$0.902 \pm 0.039$
Perfect $\Psi$	$0.0075 \pm 0.0028$	$0.078 \pm 0.014$	$1.000 \pm 0.020$

**Table 6.** Results of BAO parameter space ( $B$ ,  $A_0$ ,  $A_1$ ,  $A_2$ , and  $\alpha_{\text{BAO}}$ ) comparison between RASCALC results ( $R$ ) and mock sample covariance ( $S$ ) after reconstruction. The last row provides the perfect-case reference – expectation values and standard deviations for the three metrics, if the RASCALC precision matrix truly described the distribution of the mock correlation functions.

Mock no.	$D_{\text{KL}}(\Psi_R, C_S)$	$R_{\text{inv}}(\Psi_R, C_S)$	$\chi_{\text{red}}^2(\Psi_R, C_S)$
Average G	0.020	0.134	1.087
Average NG	0.014	0.113	1.049
1	0.122	0.252	0.907
2	0.036	0.171	1.002
3	0.041	0.202	1.082
4	0.015	0.106	0.981
5	0.030	0.149	0.999
6	0.029	0.144	0.974
7	0.021	0.125	0.981
8	0.014	0.105	0.977
9	0.095	0.329	1.154
10	0.096	0.327	1.155
1–10	$0.050 \pm 0.039$	$0.191 \pm 0.085$	$1.021 \pm 0.082$
Perfect $\Psi$	$0.0075 \pm 0.0028$	$0.078 \pm 0.014$	$1.000 \pm 0.020$

The comparison measures have been computed between the projected matrices and are presented in Table 5 for pre-recon and Table 6 for post-recon. Generally, lower expectation values of  $D_{\text{KL}}$  and  $R_{\text{inv}}$  for perfect precision make these numbers for RASCALC easier to interpret. There is a less apparent difference between pre- and post-recon. A notable exception is that all  $\chi_{\text{red}}^2$  for rescaled (with mock-averaged and single-mock clusterings) pre-recon are significantly less than 1 (meaning RASCALC ‘overestimates’ the covariance then). In other cases, mimicking non-Gaussianity gives a slight improvement for mock-averaged clustering, but higher noise in 2PCF and jackknife covariance in single-mock estimates often drives the agreement with mock sample covariance worse than in the mock-averaged Gaussian estimate.

Overall, RASCALC single-mock results are within  $<2\sigma$  (dominated by the standard deviation of the perfect reference values, except the reduced chi-squared before reconstruction, which deviates by

**Table 7.** Pre- and post-reconstruction errorbars on  $\alpha_{\text{BAO}}$  from Fisher forecast. The mean of 10 single pre-reconstruction catalogs agrees with the sample covariance within a standard deviation. For post-reconstruction, the difference is within 2 standard deviations.

$\sigma(\alpha_{\text{BAO}})$	Pre-recon	Post-recon
Sample cov	$0.01590 \pm 0.00036$	$0.01459 \pm 0.00033$
Average G	0.01522	0.01360
Average NG	0.01596	0.01392
1	0.01616	0.01424
2	0.01598	0.01395
3	0.01583	0.01410
4	0.01609	0.01405
5	0.01595	0.01385
6	0.01613	0.01396
7	0.01594	0.01389
8	0.01621	0.01400
9	0.01652	0.01402
10	0.01618	0.01405
1–10 mean $\pm$ std	$0.01610 \pm 0.00019$	$0.01401 \pm 0.00011$

$\approx 2.2$  std (combined). However, the scatter in these numbers is quite significant (e.g. a few percent in root-mean-square relative error  $R_{\text{inv}}$ ), and we should try to reduce it in future work.

#### 4.7 Errorbars on BAO scale parameter

Since the scale parameter  $\alpha_{\text{BAO}}$  is the important output of the current BAO analysis [equation (27)], we have decided to extract its errorbar, marginalized over the other four parameters. This is quite trivial after the previous subsection – we only needed to invert the RASCALC parameter-space precisions

$$C_R^{\text{par}} = (\Psi_R^{\text{par}})^{-1} \quad (31)$$

neglecting the inversion bias, since it is expected to be even smaller than before with the smaller size of the matrices. Then we extract the marginalized errorbars from all the parameter covariances as

$$\sigma(\alpha_{\text{BAO}}) = \sqrt{C_{\alpha\alpha}^{\text{par}}} \quad (32)$$

For the sample covariance, we expect the variance of  $C_{S,\alpha\alpha}$  to nearly follow equation (B2):

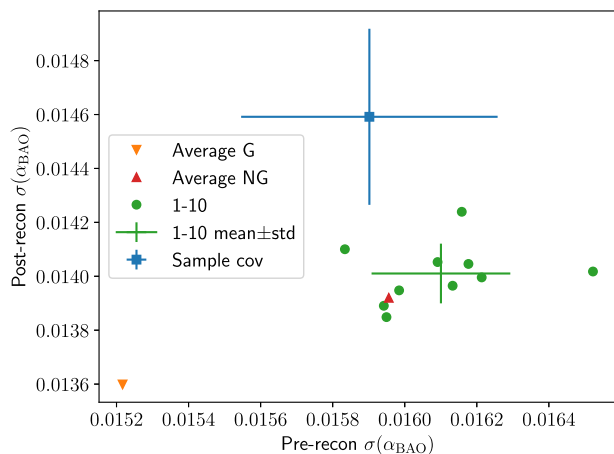
$$\text{Var}[C_{S,\alpha\alpha}^{\text{par}}] \approx \frac{2}{n_S - 1} (C_{S,\alpha\alpha}^{\text{par}})^2 \quad (33)$$

and therefore the standard deviation of  $\sigma(\alpha_{\text{BAO}})$  of

$$\sigma[\sigma(\alpha_{\text{BAO}})] \approx \frac{\sigma(\alpha_{\text{BAO}})}{\sqrt{2(n_S - 1)}}, \quad (34)$$

resulting in relative precision of 2.2 per cent. This has been confirmed in Appendix B4.

The resulting errorbar (Fisher) forecasts are provided in Table 7 and also presented as a scatter plot in Fig. 1. We can notice that in any case the post-recon precision is expected to be higher than pre-recon. In both pre-recon and post-recon,  $\sigma(\alpha_{\text{BAO}})$  in the mock-averaged clustering run without shot-noise rescaling are noticeably smaller than predicted from the sample covariance and are brought closer in the rescaled results. Mock-averaged clustering with fit shot noise and single mock runs give very similar numbers. The key conclusion is that single-mock runs are in good agreement with the sample covariance on  $\sigma(\alpha_{\text{BAO}})$ , with a remarkably close match before reconstruction (just fractions of standard deviation) and a difference of  $\approx 2$  standard deviation after reconstruction. This gives assurance



**Figure 1.** Pre- and post-reconstruction errorbars on  $\alpha_{\text{BAO}}$  from Fisher forecast plotted against each other. Compared to the mock-averaged Gaussian run (Average G), the mock-averaged rescaled run (Average NG) and single-mock predictions with rescaling are closer to the sample covariance. The horizontal (pre-reconstruction) agreement is closer than the vertical (post-reconstruction) one, but even the latter falls within  $\approx 2\sigma$ . Note that the range of the axes is quite narrow, comprising only  $\approx 9$  per cent relative difference in errorbar before reconstruction and  $\approx 7$  per cent after.

that data-based RASCALC covariances are on par with mock sample one for isotropic BAO fits.

## 5 SUMMARY AND OUTLOOK

This work continues a series of papers (O’Connell et al. 2016; O’Connell & Eisenstein 2019; Philcox & Eisenstein 2019; Philcox et al. 2020) developing a semi-empirical approach for estimating covariances of 2PCFs, combining analytical methods with the usage of measured clustering and calibration on jackknives. The former brings smoothness and reliability, and the latter allows for flexibility of the results while being independent of mock galaxy catalogs. We should note that the method is expected to be applicable at intermediate scales – as analytical methods tend to fail on the smallest scales, while on the largest scales, the number of configurations increases making the computation longer, and the signal to noise in correlation function measurement decreases. The latter issue could be alleviated by a smooth transition to a theoretically modelled 2PCF, which we leave for future work.

We have discussed the implications of split random-random counts computation and made a slight modification to the formalism to cover the reconstructed 2PCF estimates. Then, we reconsidered the methods for covariance matrix comparison, paying great attention to their meaning, interpretation, and noise stemming from mock sample variance.

Finally, we have applied the selected approaches to the validation of RASCALC on single DESI-M2 Firstgen EZ mock catalogs (using their individual clustering measurements and shifted random catalogs after reconstruction), each representing a reasonable proxy for DESI-M2 data, by comparison with full mock sample covariance. We find a close agreement (maximum deviation  $\approx 2.2\sigma$ ) with a perfect case, although much of this deviation is due to scatter in RASCALC results. The preceding discussion about the interpretations of the metrics, focusing on a smaller number of observables and even fewer parameters allowed us to obtain a clearer quantitative

assessment of the precision and accuracy of RASCALC results than in previous works. One should keep in mind the mocks are approximate and this can partially account for the imperfection of the match with the reference statistics.

Focusing on the errorbar of the BAO scale, we found a very close, percent-level agreement with the sample covariance from mocks. It is on par with the accuracy that a set of  $\approx 1000$  simulations can provide. The number of available mocks thus limits the precision of the validation at the current level.

The comparison suggests that noise in the input 2PCF might be a significant limiting factor for the accuracy of our covariance matrices in higher-dimensional spaces. Smoothing this input or complementing it with a theoretical best-fit model could help to mitigate this issue without introducing many additional assumptions. This marks an important topic for follow-up studies.

In the full measurement space before reconstruction, using a shot-noise rescaling is particularly clearly beneficial compared to the pure Gaussian estimate even with less noisy (mock sample average) input clustering. The discrepancies and fluctuations are likely to be impacted by the precision of correlation function estimates from data, which will improve with its size in the future. Further validations with larger mocks, corresponding to a year and/or full 5 yr of DESI data, will follow.

During the comparison, we have seen indications that the reconstructed extension may not be working better than with the pre-reconstructed data, contrary to the expectation. This might be related to the subtleties of small-scale behaviour of reconstructed points and will be investigated in further detail in future work.

Another perspective direction is the development of alternatives to shot-noise rescaling within the more generic semi-analytic configuration-space formalism. Usage of fully empirical higher-point functions is likely to be not viable, due to a significantly higher number of bins and accordingly lower signal to noise. Precise theoretical modeling of non-Gaussian correlation functions is also very challenging. Instead, we might include a basic prescription for non-Gaussian covariance contribution inferred from a set of detailed simulations, or use approximate expressions for higher-point functions like  $\zeta(\mathbf{r}_1, \mathbf{r}_2) = Q[\xi(r_1)\xi(r_2) + \xi(r_1)\xi(|\mathbf{r}_1 - \mathbf{r}_2|) + \xi(r_2)\xi(|\mathbf{r}_1 - \mathbf{r}_2|)]$  motivated by hierarchical models (Peebles & Groth 1975), and possibly a similar structure for the 4PCF. This might provide better accuracy than rescaling the Gaussian terms while keeping the number of parameters low and thus still allowing us to fit them to a reference (e.g. jackknife) covariance. On the other hand, the aforementioned 3PCF prescription is known to be far from exact with constant  $Q$  (Takada & Jain 2003), and the computations may suffer from slower convergence due to additional large values of small-scale 2PCF compared to Gaussian parts.

The key advantage of the approaches considered in this paper is that a covariance can be based on the data itself and does not require matching mock catalogs each time. This alleviates the concern about the accuracy of such approximate simulations, not completely removing it since we still need some as references for validation of the prescriptions. Perhaps more importantly, covariance computation becomes much more flexible. Relevant cases are when the clustering signal changes after more data are gathered, or when alternative assumptions (for instance, the base cosmology) are tested. For such updates, calibration, generation, and processing of a suite of mocks large enough for good sample covariance consumes great resources. The computation of one proxy-data covariance in this paper took about 100 core-hours for pre-recon and about

300 core-hours for post-recon.<sup>6</sup> This could be optimized further by noting that the intrinsic consistency in each run was much higher than the resemblance of reference sample covariance – time of computation helps with the former but the latter is fundamentally limited by the accuracy of approximations and precision of the input correlation function. Moreover, the number of mocks limits the level of accuracy of validation. Fast covariance matrix computation without mocks can also allow shifting the balance from quantity to quality of the simulations, freeing resources for more detailed ones.

## ACKNOWLEDGEMENTS

We thank Daniel Forero-Sanchez, Violeta Gonzalez-Perez, Nikhil Padmanabhan, Will Percival, Oliver Philcox, and Martin White for useful comments and fruitful discussions. MR and DJE are supported by the US Department of Energy (DOE) grant DE-SC0007881 and the Simons Foundation Investigator program.

This material is based upon work supported by the US DOE, Office of Science, Office of High-Energy Physics, under Contract No. DE-AC02-05CH11231, and by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract. Additional support for DESI was provided by the US National Science Foundation (NSF), Division of Astronomical Sciences under Contract No. AST-0950945 to the NSF’s National Optical-Infrared Astronomy Research Laboratory; the Science and Technology Facilities Council of the United Kingdom; the Gordon and Betty Moore Foundation; the Heising-Simons Foundation; the French Alternative Energies and Atomic Energy Commission (CEA); the National Council of Science and Technology of Mexico (CONACYT); the Ministry of Science and Innovation of Spain (MICINN), and by the DESI Member Institutions: <https://www.desi.lbl.gov/collaborating-institutions>. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the US NSF, the US DOE, or any of the listed funding agencies.

The authors are honoured to be permitted to conduct scientific research on Iolkam Du’ag (Kitt Peak), a mountain with particular significance to the Tohono O’odham Nation.

<sup>6</sup>In one case out of 10 (both pre- and post-recon), a repeated computation taking twice longer was performed.

## DATA AVAILABILITY

The code is openly accessible at <https://github.com/oliverphilcox/RascalC>. All data from the tables and figures are available in machine-readable format at [doi:10.5281/zenodo.7750637](https://doi.org/10.5281/zenodo.7750637) in compliance with the DESI data management plan. The DESI-M2 Firstgen EZ mocks used in this paper will be made public with the DESI Y1 data release (DR1), and all the covariance matrices used in this work will be released in the same supplementary material (Rashkovetskyi et al. 2023).

## REFERENCES

- Alam S. et al., 2021, *Phys. Rev. D*, 103, 083533  
 Ata M. et al., 2018, *MNRAS*, 473, 4773  
 Bernstein G. M., 1994, *ApJ*, 424, 569  
 Burden A., Percival W. J., Howlett C., 2015, *MNRAS*, 453, 456  
 Chuang C.-H., Kitaura F.-S., Prada F., Zhao C., Yepes G., 2015, *MNRAS*, 446, 2621  
 DESI Collaboration, 2016, preprint ([arXiv:1611.00036](https://arxiv.org/abs/1611.00036))  
 DESI Collaboration, 2022, *AJ*, 164, 207  
 DESI Collaboration, 2023a, preprint ([arXiv:2306.06307](https://arxiv.org/abs/2306.06307))  
 DESI Collaboration, 2023b, preprint ([arXiv:2306.06308](https://arxiv.org/abs/2306.06308))  
 Eisenstein D. J. et al., 2005, *ApJ*, 633, 560  
 Eisenstein D. J., Seo H.-J., Sirko E., Spergel D. N., 2007, *ApJ*, 664, 675  
 Feldman H. A., Kaiser N., Peacock J. A., 1994, *ApJ*, 426, 23  
 Hartlap J., Simon P., Schneider P., 2007, *A&A*, 464, 399  
 Isserlis L., 1918, *Biometrika*, 12, 134  
 Keihänen E. et al., 2019, *A&A*, 631, A73  
 Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64  
 Mohammad F. G., Percival W. J., 2022, *MNRAS*, 514, 1289  
 Moon J. et al., 2023, preprint ([arXiv:2304.08427](https://arxiv.org/abs/2304.08427))  
 O’Connell R., Eisenstein D. J., 2019, *MNRAS*, 487, 2701  
 O’Connell R., Eisenstein D., Vargas M., Ho S., Padmanabhan N., 2016, *MNRAS*, 462, 2681  
 Paillas E. et al., 2023, *MNRAS*  
 Peebles P. J. E., Groth E. J., 1975, *ApJ*, 196, 1  
 Philcox O. H. E., Eisenstein D. J., 2019, *MNRAS*, 490, 5931  
 Philcox O. H. E., Eisenstein D. J., O’Connell R., Wiegand A., 2020, *MNRAS*, 491, 3290  
 Rashkovetskyi M., Eisenstein D. et al. 2023, Validation of semi-analytical, semi-empirical covariance matrices for two-point correlation function for Early DESI data, , <https://doi.org/10.5281/zenodo.7750637>  
 Ross A. J. et al., 2017, *MNRAS*, 464, 1168  
 Takada M., Jain B., 2003, *MNRAS*, 340, 580  
 Trusov S. et al., 2023, preprint ([arXiv:2306.16332](https://arxiv.org/abs/2306.16332))  
 Vargas-Magaña M. et al., 2018, *MNRAS*, 477, 1153  
 Wadekar D., Scoccimarro R., 2020, *Phys. Rev. D*, 102, 123517  
 Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, *Phys. Rep.*, 530, 87  
 Xu X. et al., 2010, *ApJ*, 718, 1224  
 Zhao C. et al., 2021, *MNRAS*, 503, 1149  
 Zhou R. et al., 2020, *Res. Notes Am. Astron. Soc.*, 4, 181

## APPENDIX A: COVARIANCE ESTIMATORS

### A1 Full covariance in radial and angular bins

The expression for full covariance in radial and angular bins ( $\text{cov} \left[ (\hat{\xi}^{XY})_a^c, (\hat{\xi}^{ZW})_b^d \right]$ ) is

$$\begin{aligned} (\tilde{C}^{XY,ZW})_{ab}^{cd}(\alpha_{\text{SN}}) &= ({}^4C^{XY,ZW})_{ab}^{cd} + \frac{\alpha_{\text{SN}}^X}{4} \left[ \delta^{XW} ({}^3C^{X,YZ})_{ab}^{cd} + \delta^{XZ} ({}^3C^{X,YW})_{ab}^{cd} \right] \\ &+ \frac{\alpha_{\text{SN}}^Y}{4} \left[ \delta^{YW} ({}^3C^{Y,XZ})_{ab}^{cd} + \delta^{YZ} ({}^3C^{Y,XW})_{ab}^{cd} \right] + \frac{\alpha_{\text{SN}}^X \alpha_{\text{SN}}^Y}{2} [\delta^{XW} \delta^{YZ} + \delta^{XZ} \delta^{YW}] ({}^2C^{XY})_{ab}^{cd} \end{aligned} \quad (\text{A1})$$

with

$$\begin{aligned} ({}^4C^{XY,ZW})_{ab}^{cd} &= \frac{1}{(R^X R^Y)_a^c (R^Z R^W)_b^d} \sum_{i \neq j \neq k \neq l} n_i^X n_j^Y n_k^Z n_l^W w_i^X w_j^Y w_k^Z w_l^W \Theta^a(r_{ij}) \Theta^c(\mu_{ij}) \Theta^b(r_{kl}) \Theta^d(\mu_{kl}) \\ &\times \left[ \eta_{ijkl}^{(c),XYWZ} + 2\xi_{ik}^{XZ} \xi_{jl}^{YW} \right] \\ ({}^3C^{Y,XZ})_{ab}^{cd} &= \frac{4}{(R^X R^Y)_a^c (R^Y R^Z)_b^d} \sum_{i \neq j \neq k} n_i^X n_j^Y n_k^Z w_i^X (w_j^Y)^2 w_k^Z \Theta^a(r_{ij}) \Theta^c(\mu_{ij}) \Theta^b(r_{jk}) \Theta^d(\mu_{jk}) [\xi_{ijk}^{XYZ} + \xi_{ik}^{XZ}] \\ ({}^2C^{XY})_{ab}^{cd} &= \frac{2\delta^{ab} \delta^{cd}}{(R^X R^Y)_a^c (R^X R^Y)_b^d} \sum_{i \neq j} n_i^X n_j^Y (w_i^X w_j^Y)^2 \Theta^a(r_{ij}) \Theta^c(\mu_{ij}) [1 + \xi_{ij}^{XY}] \end{aligned} \quad (\text{A2})$$

where  $\delta^{XY}$ ,  $\delta^{ab}$ , and  $\delta^{cd}$  are Kronecker deltas. Similarly to Philcox et al. (2020), we have written  $2\xi_{ik}^{XZ} \xi_{jl}^{YW}$  instead of  $\xi_{ik} \xi_{jl} + \xi_{il} \xi_{jk}$ , allowing to optimize the computation using the symmetries of the term but requiring to compute a few more distinct terms in multitracer setup. Computing full sums (beyond the pair one) is not feasible, so are estimated by the Monte Carlo method instead, by randomly sampling a subset of two/three/four-point configurations from the random catalog (Philcox et al. 2020).

The most practical form for single tracer  $X$  is simpler:

$$(C^{XX,XX})_{ab}^{cd}(\alpha_{\text{SN}}^X) = ({}^4C^{XX,XX})_{ab}^{cd} + \alpha_{\text{SN}}^X ({}^3C^{X,XX})_{ab}^{cd} + (\alpha_{\text{SN}}^X)^2 ({}^2C^{XX})_{ab}^{cd}. \quad (\text{A3})$$

### A2 Jackknife covariance in radial and angular bins

The jackknife pair counts are

$$\begin{aligned} (N^X N_A^Y)_a^c &= \sum_{i \neq j} n_i^X n_j^Y w_i^X w_j^Y \Theta^a(r_{ij}) \Theta^c(\mu_{ij}) \delta_i^X \delta_j^Y q_{ij}^A \\ (R^X R_A^Y)_a^c &= \sum_{i \neq j} n_i^X n_j^Y w_i^X w_j^Y \Theta^a(r_{ij}) \Theta^c(\mu_{ij}) q_{ij}^A, \end{aligned} \quad (\text{A4})$$

where  $q_{ij}^A$  is the jackknife weighting factor for the pair of particles, and for any pair,  $\sum_A q_{ij}^A = 1$  in the unrestricted jackknife formalism.

The binned correlation function estimate  $(\hat{\xi}_A^{XY})_a^c$  is their ratio. It is sensible to weight the regions by the  $RR$  pair counts, which roughly correspond to the volume fraction of each region:

$$(w_A^{XY})_a^c = \frac{(R^X R_A^Y)_a^c}{(R^X R^Y)_a^c}, \quad (\text{A5})$$

and then the weighted average of the jackknife 2PCF is identical to full-survey estimate [equation (2)].

The jackknife covariance estimate is

$$(C_J^{XY,ZW})_{ab}^{cd} = \frac{1}{1 - \sum_B (w_B^{XY})_a^c (w_B^{ZW})_b^d} \sum_A \left\{ (w_A^{XY})_a^c (w_A^{ZW})_b^d \left[ (\hat{\xi}_A^{XY})_a^c - (\hat{\xi}^{XY})_a^c \right] \left[ (\hat{\xi}_A^{ZW})_b^d - (\hat{\xi}^{ZW})_b^d \right] \right\}. \quad (\text{A6})$$

By substituting equation (2), expanding using equations (A4) and (3) and simplifying through equation (6) one can arrive to

$$\begin{aligned} (\tilde{C}_J^{XY,ZW})_{ab}^{cd} &= ({}^4C_J^{XY,ZW})_{ab}^{cd} + \frac{\alpha^X}{4} \left[ \delta^{XW} ({}^3C_J^{X,YZ})_{ab}^{cd} + \delta^{XZ} ({}^3C_J^{X,YW})_{ab}^{cd} \right] + \frac{\alpha^Y}{4} \left[ \delta^{YW} ({}^3C_J^{Y,XZ})_{ab}^{cd} + \delta^{YZ} ({}^3C_J^{Y,XW})_{ab}^{cd} \right] \\ &+ \frac{\alpha^X \alpha^Y}{2} [\delta^{XW} \delta^{YZ} + \delta^{XZ} \delta^{YW}] ({}^2C_J^{XY})_{ab}^{cd} \end{aligned} \quad (\text{A7})$$



with

$$\begin{aligned}
\left({}^4C_J^{XY,ZW}\right)_{ab}^{cd} &= \frac{1}{\left(R^X R^Y\right)_a^c \left(R^Z R^W\right)_b^d \left[1 - \sum_B \left(w_B^{XY}\right)_a^c \left(w_B^{ZW}\right)_b^d\right]_{i \neq j \neq k \neq l}} \sum n_i^X n_j^Y n_k^Z n_l^W w_i^X w_j^Y w_k^Z w_l^W \Theta^a(r_{ij}) \Theta^c(\mu_{ij}) \\
&\quad \times \Theta^b(r_{kl}) \Theta^d(\mu_{kl}) \left[ \eta_{ijkl}^{(c),XYZW} + \xi_{ij}^{XY} \xi_{kl}^{ZW} + 2\xi_{ik}^{XZ} \xi_{jl}^{YW} \right] \left(\omega_{ijkl}^{XY,ZW}\right)_{ab}^{cd} \\
\left({}^3C_J^{Y,XZ}\right)_{ab}^{cd} &= \frac{4}{\left(R^X R^Y\right)_a^c \left(R^Y R^Z\right)_b^d \left[1 - \sum_B \left(w_B^{XY}\right)_a^c \left(w_B^{YZ}\right)_b^d\right]_{i \neq j \neq k}} \sum n_i^X n_j^Y n_k^Z w_i^X \left(w_j^Y\right)^2 w_k^Z \Theta^a(r_{ij}) \Theta^c(\mu_{ij}) \Theta^b(r_{jk}) \Theta^d(\mu_{jk}) \\
&\quad \times \left[\zeta_{ijk}^{XYZ} + \xi_{ik}^{XZ}\right] \left(\omega_{ijk}^{XY,YZ}\right)_{ab}^{cd} \\
\left({}^2C_J^{XY}\right)_{ab}^{cd} &= \frac{2\delta^{ab}\delta^{cd}}{\left(R^X R^Y\right)_a^c \left(R^X R^Y\right)_b^d \left[1 - \sum_B \left(w_B^{XY}\right)_a^c \left(w_B^{XY}\right)_b^d\right]_{i \neq j}} \sum n_i^X n_j^Y \left(w_i^X w_j^Y\right)^2 \Theta^a(r_{ij}) \Theta^c(\mu_{ij}) \left[1 + \xi_{ij}^{XY}\right] \left(\omega_{ijij}^{XY,XY}\right)_{ab}^{cd}, \tag{A8}
\end{aligned}$$

where  $(\omega_{ijkl})_{ab}^{cd}$  is an additional weight tensor:

$$\left(\omega_{ijkl}^{XY,ZW}\right)_{ab}^{cd} = \sum_A \left\{ \left[ q_{ij}^A - \left(w_A^{XY}\right)_a^c \right] \left[ q_{kl}^A - \left(w_A^{ZW}\right)_b^d \right] \right\}. \tag{A9}$$

In practice, shot-noise rescaling for each tracer has been obtained by fitting the prediction for jackknife covariance of its auto-correlation function

$$\left(C_J^{XX,XX}\right)_{ab}^{cd} = \left({}^4C_J^{XX,XX}\right)_{ab}^{cd} + \alpha^X \left({}^3C_J^{XX,XX}\right)_{ab}^{cd} + \left(\alpha^X\right)^2 \left({}^2C_J^{XX}\right)_{ab}^{cd} \tag{A10}$$

to the data-based jackknife estimate [computed via equation (A6)]. The resulting  $\alpha^X$  value(s) should be plugged into equation (A1) to obtain the full survey covariance.

### A3 Covariance of Legendre multipoles in radial bins

Philcox & Eisenstein (2019) further derive direct covariance estimators for Legendre moments of the anisotropic 2PCF, which is related to  $\xi(r, \mu)$  through

$$\xi^{XY}(r, \mu) = \sum_{\ell=0}^{\infty} \left(\xi^{XY}\right)_{\ell}(r) L_{\ell}(\mu), \tag{A11}$$

$$\left(\xi^{XY}\right)_{\ell}(r) = \frac{2\ell+1}{2} \int_{-1}^1 d\mu \xi^{XY}(r, \mu) L_{\ell}(\mu) = \frac{(-1)^{\ell}+1}{2} (2\ell+1) \int_0^1 d\mu \xi^{XY}(r, \mu) L_{\ell}(\mu), \tag{A12}$$

$L_{\ell}(\mu)$  being the Legendre polynomial of order  $\ell$ , and the second equality in last line assumes symmetry  $\xi^{XY}(r, \mu) = \xi^{XY}(r, -\mu)$  (necessarily true for auto-correlation,  $X = Y$ , and violation in cross-correlations is debatable).

While one could estimate the angularly binned 2PCF first and then transform it to Legendre moments, a direct computation of the latter allows to evade inaccuracies caused by the replacement of integral by sum over bins and removes the need for very fine splitting in  $\mu$ , which would make the covariance integrals slower to converge.

Averaging equation (A12) between the boundaries of radial bin  $a$ , we obtain

$$\left(\xi^{XY}\right)_a^{\ell} = (2\ell+1) \int_0^1 d\mu \left(\xi^{XY}\right)_a(\mu) L_{\ell}(\mu). \tag{A13}$$

Here  $\xi_a^{XY}(\mu)$ , binned radially but not angularly, is given by the reduced Landy-Szalay estimator

$$\left(\xi^{XY}\right)_a(\mu) = \frac{\left(N^X N^Y\right)_a(\mu)}{\left(R^X R^Y\right)_a(\mu)} \tag{A14}$$

similarly to equations (1) & (2). The continuous counts  $PP_a(\mu)$  (where each  $P$  can stand for  $D^X, D^Y, R^X$  or  $R^Y$ ) are the limit of the ratio of conventional pair counts in infinitesimally small angular bins  $c$  to its width  $\delta\mu$ ,  $PP_a^c = \int d\mu PP_a(\mu) \Theta^c(\mu) \approx PP_a(\mu_c) \delta\mu$ ,  $\mu_c$  being the bin center.

A functional form for the continuous random pair counts  $(R^X R^Y)_a(\mu)$  is needed to go further due to integration in equation (A13). This is done via a *survey correction function*  $\Phi^{XY}(r_a, \mu)$  [following Xu et al. (2010)], which accounts for survey boundaries and selection, defined through

$$\left(R^X R^Y\right)_a(\mu) \equiv \frac{V^X \bar{n}^X \bar{n}^Y \bar{w}^X \bar{w}^Y v_a}{\Phi^{XY}(r_a, \mu)}, \tag{A15}$$

where  $V^X$  is the total volume of survey for tracer  $X$ , and  $\bar{n}^X$  and  $\bar{w}^X$  are the survey-averaged number density and weight for tracer  $X$ , respectively.  $v_a = \frac{4\pi}{3} (r_{a,\max}^3 - r_{a,\min}^3)$  is the volume of radial bin  $a$  ( $r_{a,\min}$  and  $r_{a,\max}$  being its lower and upper boundaries). The numerator is the expression for a periodic box with uniform number density, weights, and bins in  $|\mu|$  (i.e. negative values reversed into  $[0,1]$  interval); thus, in this simple case, the survey correction function  $\Phi^{XY}(r_a, \mu) = 1$ . For non-trivial geometry, varying density and/or weights, it may become different but not too far by order of magnitude. However, note that only the angular variable value is arbitrary, while the radius is taken as representative of the bin. In practice, a piecewise-polynomial functional fit to empirical data was used by Philcox & Eisenstein (2019).

Then the 2PCF Legendre moment can be estimated and the covariance can be computed by definition using equation (6), resulting in

$$\begin{aligned} (C^{XY,ZW})_{ab}^{pq} &= ({}^4C^{XY,ZW})_{ab}^{pq} + \frac{\alpha^X}{4} \left[ \delta^{XW} ({}^3C^{X,YZ})_{ab}^{pq} + \delta^{XZ} ({}^3C^{X,YW})_{ab}^{pq} \right] \\ &+ \frac{\alpha^Y}{4} \left[ \delta^{YW} ({}^3C^{Y,XZ})_{ab}^{pq} + \delta^{YZ} ({}^3C^{Y,XW})_{ab}^{pq} \right] \frac{\alpha^X \alpha^Y}{2} (\delta^{XW} \delta^{YZ} + \delta^{XZ} \delta^{YW}) ({}^2C^{XY})_{ab}^{pq}. \end{aligned} \quad (A16)$$

$$\begin{aligned} ({}^4C^{XY,ZW})_{ab}^{pq} &= \frac{(2p+1)(2q+1)}{V^X V^Z \bar{n}^X \bar{n}^Y \bar{n}^Z \bar{n}^W \bar{w}^X \bar{w}^Y \bar{w}^Z \bar{w}^W v_a v_b} \sum_{i \neq j \neq k \neq l} n_i^X n_j^Y n_k^Z n_l^W w_i^X w_j^Y w_k^Z w_l^W \Theta^a(r_{ij}) \Theta^b(r_{kl}) \\ &\times \Phi^{XY}(r_a, \mu_{ij}) \Phi^{ZW}(r_b, \mu_{kl}) L_p(\mu_{ij}) L_q(\mu_{kl}) \left[ \xi_{ijkl}^{(4),XYZW} + \xi_{ik}^{XZ} \xi_{jl}^{YW} + \xi_{il}^{XW} \xi_{jk}^{YZ} \right] \\ ({}^3C^{Y,XZ})_{ab}^{pq} &= 4 \times \frac{(2p+1)(2q+1)}{V^X V^Z \bar{n}^X (\bar{n}^Y)^2 n^Z \bar{w}^X (\bar{w}^Y)^2 w^Z v_a v_b} \sum_{i \neq j \neq k} n_i^X n_j^Y n_k^Z w_i^X (w_j^Y)^2 w_k^Z \Theta^a(r_{ij}) \Theta^b(r_{jk}) \\ &\times \Phi^{XY}(r_a, \mu_{ij}) \Phi^{ZY}(r_b, \mu_{jk}) L_p(\mu_{ij}) L_q(\mu_{jk}) \left[ \xi_{ijk}^{XYZ} + \xi_{ik}^{XZ} \right] \\ ({}^2C^{XY})_{ab}^{pq} &= 2\delta_{ab} \times \frac{(2p+1)(2q+1)}{(V^X \bar{n}^X \bar{n}^Y \bar{w}^X \bar{w}^Y v_a)^2} \sum_{i \neq j} n_i^X n_j^Y (w_i^X w_j^Y)^2 \Theta^a(r_{ij}) \\ &\times (\Phi^{XY}(r_a, \mu_{kl}))^2 L_p(\mu_{ij}) L_q(\mu_{ij}) [1 + \xi_{ij}^{XY}]. \end{aligned} \quad (A17)$$

Jackknife poses certain challenges to direct computation for the Legendre moments of the 2PCF. So far, it has been suggested that the shot-noise rescaling value(s) should be optimized with a jackknife estimate on angularly binned 2PCF (Philcox & Eisenstein 2019).

## APPENDIX B: STATISTICS OF COMPARISON METRICS FOR NOISY SAMPLE COVARIANCE MATRIX

Here, we provide derivations of the expectation values for comparison metrics between a noisy sample covariance and the true covariance/precision matrix. This is useful for testing how close RASCALC results are to the latter.

### B1 KL divergence mean and variance

A more generic setup – two sample covariance matrices based on draws from a multivariate normal distribution – has been considered in appendix D of Philcox et al. (2020). However, the derivation was limited to the expectation value of the KL divergence between them, and we have not been able to find a reference about the metric’s scatter around the mean (variance or standard deviation). In addition, we believe the final result there is slightly incorrect, namely 1 should be subtracted from the number of samples. This is because for the estimate of sample covariance commonly used with mocks

$$X_{ab} = \frac{1}{n_S - 1} \sum_{i=1}^{n_S} (x_{a,i} - \bar{x}_a)(x_{b,i} - \bar{x}_b) \quad (B1)$$

the mean is not known beforehand but estimated from the sample as well:  $\bar{x}_a \equiv \frac{1}{n_S} \sum_{i=1}^{n_S} x_{a,i}$ . This reduces the number of degrees of freedom by one. Then the covariance of sample covariance matrix elements is

$$\text{cov}(X_{ab}, X_{cd}) = \frac{C_{0,ac} C_{0,bd} + C_{0,ad} C_{0,bc}}{n_S - 1} \quad (B2)$$

instead of

$$\text{cov}(X_{ab}, X_{cd}) = \frac{C_{0,ac} C_{0,bd} + C_{0,ad} C_{0,bc}}{n_S}$$

as in Philcox et al. (2020).  $C_0$  is the true underlying covariance matrix of the Gaussian distribution the samples are drawn from. The sample covariance estimate is unbiased, meaning that the expectation value is the true covariance:  $\langle \mathbf{X} \rangle = C_0$ . Then, considering two sample covariance matrices  $\mathbf{X}_i$  obtained from  $n_S^{(i)}$  samples each, decomposing them as  $\mathbf{X}_i = C_0 + \delta \mathbf{X}_i$ , Taylor expanding and only leaving the leading non-trivial (quadratic) order in  $\delta \mathbf{X}_i$ , we obtain

$$\langle D_{\text{KL}}(\mathbf{X}_1^{-1}, \mathbf{X}_2) \rangle \approx \frac{N_{\text{bins}}(N_{\text{bins}} + 1)}{4} \left( \frac{1}{n_S^{(1)} - 1} + \frac{1}{n_S^{(2)} - 1} \right) \quad (B3)$$

instead of

$$\langle D_{\text{KL}}(\mathbf{X}_1^{-1}, \mathbf{X}_2) \rangle \approx \frac{N_{\text{bins}}(N_{\text{bins}} + 1)}{4} \left( \frac{1}{n_S^{(1)}} + \frac{1}{n_S^{(2)}} \right)$$

as in Philcox et al. (2020).

Since in this work we only consider one sample covariance matrix, while the RASCALC results are not expected to follow the Wishart distribution, the more relevant result is for the true precision matrix  $\Psi_0$ :

$$\langle D_{\text{KL}}(\Psi_0, \mathbf{X}) \rangle \approx \frac{N_{\text{bins}}(N_{\text{bins}} + 1)}{4(n_S - 1)}, \quad (B4)$$

which can be obtained from equation (B3) by setting the first number of samples to infinity, reducing the noise in  $\mathbf{X}_1^{-1}$  to zero.

For the further derivations, it is convenient to ‘normalize’ the covariance. Let us take

$$\mathbf{y}_i = \Psi_0^{1/2} \mathbf{x}_i, \quad (\text{B5})$$

where  $\Psi_0^{1/2}$  means the matrix square root of  $\Psi_0$  – a matrix with the same eigenvectors and eigenvalues equal to the square roots of corresponding eigenvalues of the original matrix. Then

$$\text{cov}(y_{a,i}, y_{b,j}) = \delta^{ij} \delta^{ab}. \quad (\text{B6})$$

Let us also introduce  $\bar{y}_a \equiv \frac{1}{n_S} \sum_{i=1}^{n_S} y_{a,i}$ , define  $\delta y_{a,i} \equiv y_{a,i} - \bar{y}_a$  (so that  $\langle \delta y_{a,i} \rangle = 0$ ) and finally compute the ‘normalized’ covariance matrix:

$$Y_{ab} \equiv \frac{1}{n_S - 1} \sum_{i=1}^{n_S} \delta y_{a,i} \delta y_{b,i}. \quad (\text{B7})$$

Then also

$$\mathbf{Y} = \Psi_0^{1/2} \mathbf{X} \Psi_0^{1/2}. \quad (\text{B8})$$

Let us compute

$$\text{cov}(\delta y_{a,i}, \delta y_{b,j}) = \langle \delta y_{a,i} \delta y_{b,j} \rangle = \delta^{ij} \delta^{ab} - 2 \times \frac{1}{n_S} \delta^{ab} + n_S \times \frac{1}{n_S^2} \delta^{ab} = \left( \delta^{ij} - \frac{1}{n_S} \right) \delta^{ab}. \quad (\text{B9})$$

As a consequence,  $\langle \mathbf{Y} \rangle = \mathbb{I}$ , and we can expand

$$\mathbf{Y} = \mathbb{I} + \delta \mathbf{Y} \quad (\text{B10})$$

while  $\langle \delta \mathbf{Y} \rangle = 0$ . Also,

$$\text{cov}(Y_{ab}, Y_{cd}) = \langle \delta Y_{ab} \delta Y_{cd} \rangle = \frac{\delta^{ac} \delta^{bd} + \delta^{ad} \delta^{bc}}{n_S - 1}. \quad (\text{B11})$$

Now let us expand the KL divergence using the ‘normalized’ covariance matrix, starting from

$$2D_{\text{KL}}(\Psi_0, \mathbf{X}) = \text{tr}(\Psi_0 \mathbf{X}) - N_{\text{bins}} - \ln \det(\Psi_0 \mathbf{X}), \quad (\text{B12})$$

we can write  $\Psi_0 = \Psi_0^{1/2} \Psi_0^{1/2}$ , use the cyclic property of trace and determinant to arrive to

$$2D_{\text{KL}}(\Psi_0, \mathbf{X}) = \text{tr}(\mathbf{Y}) - N_{\text{bins}} - \ln \det(\mathbf{Y}), \quad (\text{B13})$$

remembering equation (B8). Then we expand in  $\delta \mathbf{Y}$  (equation (B10)):

$$2D_{\text{KL}}(\Psi_0, \mathbf{X}) = \text{tr}(\delta \mathbf{Y}) - \ln \det(\mathbb{I} + \delta \mathbf{Y}). \quad (\text{B14})$$

Now using  $\ln \det \mathbf{A} = \text{tr} \ln \mathbf{A}$  and expanding the second term in Taylor series up to quadratic order in  $\delta \mathbf{Y}$  we obtain

$$2D_{\text{KL}}(\Psi_0, \mathbf{X}) \approx \frac{1}{2} \text{tr} [(\delta \mathbf{Y})^2] = \frac{1}{2} \sum_{a,b=1}^{N_{\text{bins}}} \delta Y_{ab} \delta Y_{ab}. \quad (\text{B15})$$

Taking the expectation value of equation (B15) and using equation (B11), one can re-derive equation (B4). We will proceed to compute the variance:

$$\text{Var} \{ \text{tr} [(\delta \mathbf{Y})^2] \} = \left\langle (\text{tr} [(\delta \mathbf{Y})^2])^2 \right\rangle - \langle \text{tr} [(\delta \mathbf{Y})^2] \rangle^2 = \sum_{a,b,c,d=1}^{N_{\text{bins}}} [(\delta Y_{ab} \delta Y_{ab} \delta Y_{cd} \delta Y_{cd}) - \langle \delta Y_{ab} \delta Y_{ab} \rangle \langle \delta Y_{cd} \delta Y_{cd} \rangle]. \quad (\text{B16})$$

Full expansion gives

$$Y_{ab} Y_{ab} Y_{cd} Y_{cd} = \frac{1}{(n_S - 1)^4} \sum_{i,j,k,l=1}^{n_S} \delta y_{a,i} \delta y_{b,i} \delta y_{a,j} \delta y_{b,j} \delta y_{c,k} \delta y_{d,k} \delta y_{c,l} \delta y_{d,l}. \quad (\text{B17})$$

$\delta y$  are normally distributed and have zero means, so for them, we can use Wick’s theorem to split this into all possible pairs. ( $\delta \mathbf{Y}$  also has zero mean, but not Gaussian distribution, this is why we need to go to a deeper level.) The total number of pairs is  $8!/(4! \times 2^4) = 105$ , so it is easy to go over them in a computer program. Additionally, it is useful to check which are similar. It is apparent that the following five index permutations leave the expression unchanged:  $a \leftrightarrow b$ ,  $c \leftrightarrow d$ ,  $i \leftrightarrow j$ ,  $k \leftrightarrow l$  and  $(a, b, i, j) \leftrightarrow (c, d, k, l)$ . Finally, some of the pairs will not contribute to variance and can be excluded: contraction of a pair inside the same  $Y$  contribute to  $\langle Y \rangle = \mathbb{I}$  and must be subtracted; and if all the contracted pairs correspond to  $Y$ ’s with the same indices, it contributes to the mean of  $D_{\text{KL}}$  and has to be subtracted too.

We find there are 56 pair assignments contributing to  $\approx 16(n_S - 1)^4 \text{Var}[D_{\text{KL}}(\Psi_0, X)]$ , but no more than 8 of them are distinct after using symmetries:

$$8 \sum_{a,b,c,d=1}^{N_{\text{bins}}} \sum_{i,j,k,l=1}^{n_S} \langle \delta y_{a,i} \delta y_{a,j} \rangle \langle \delta y_{b,i} \delta y_{c,k} \rangle \langle \delta y_{b,j} \delta y_{c,l} \rangle \langle \delta y_{d,k} \delta y_{d,l} \rangle = 8 \sum_{a,b,c,d=1}^{N_{\text{bins}}} \sum_{i,j,k,l=1}^{n_S} \left( \delta^{ij} - \frac{1}{n_S} \right) \left( \delta^{ik} - \frac{1}{n_S} \right) \delta^{bc} \left( \delta^{jl} - \frac{1}{n_S} \right) \delta^{bc} \left( \delta^{kl} - \frac{1}{n_S} \right) = 8N_{\text{bins}}^3 (n_S - 1) \quad (\text{B18})$$

$$16 \sum_{a,b,c,d=1}^{N_{\text{bins}}} \sum_{i,j,k,l=1}^{n_S} \langle \delta y_{a,i} \delta y_{a,j} \rangle \langle \delta y_{b,i} \delta y_{c,k} \rangle \langle \delta y_{b,j} \delta y_{d,l} \rangle \langle \delta y_{d,k} \delta y_{c,l} \rangle =$$

$$16 \sum_{a,b,c,d=1}^{N_{\text{bins}}} \sum_{i,j,k,l=1}^{n_S} \left( \delta^{ij} - \frac{1}{n_S} \right) \left( \delta^{ik} - \frac{1}{n_S} \right) \delta^{bc} \left( \delta^{jl} - \frac{1}{n_S} \right) \delta^{bd} \left( \delta^{kl} - \frac{1}{n_S} \right) \delta^{dc} = 16 N_{\text{bins}}^2 (n_S - 1) \quad (\text{B19})$$

$$8 \sum_{a,b,c,d=1}^{N_{\text{bins}}} \sum_{i,j,k,l=1}^{n_S} \langle \delta y_{a,i} \delta y_{b,j} \rangle \langle \delta y_{b,i} \delta y_{c,k} \rangle \langle \delta y_{a,j} \delta y_{d,l} \rangle \langle \delta y_{d,k} \delta y_{c,l} \rangle =$$

$$8 \sum_{a,b,c,d=1}^{N_{\text{bins}}} \sum_{i,j,k,l=1}^{n_S} \left( \delta^{ij} - \frac{1}{n_S} \right) \delta^{ab} \left( \delta^{ik} - \frac{1}{n_S} \right) \delta^{bc} \left( \delta^{jl} - \frac{1}{n_S} \right) \delta^{ad} \left( \delta^{kl} - \frac{1}{n_S} \right) \delta^{cd} = 8 N_{\text{bins}} (n_S - 1) \quad (\text{B20})$$

$$4 \sum_{a,b,c,d=1}^{N_{\text{bins}}} \sum_{i,j,k,l=1}^{n_S} \langle \delta y_{a,i} \delta y_{c,k} \rangle \langle \delta y_{b,i} \delta y_{d,k} \rangle \langle \delta y_{a,j} \delta y_{c,l} \rangle \langle \delta y_{b,j} \delta y_{d,l} \rangle =$$

$$4 \sum_{a,b,c,d=1}^{N_{\text{bins}}} \sum_{i,j,k,l=1}^{n_S} \left( \delta^{ik} - \frac{1}{n_S} \right) \delta^{ac} \left( \delta^{ik} - \frac{1}{n_S} \right) \delta^{bd} \left( \delta^{jl} - \frac{1}{n_S} \right) \delta^{ac} \left( \delta^{jl} - \frac{1}{n_S} \right) \delta^{bd} = 4 N_{\text{bins}}^2 (n_S - 1)^2 \quad (\text{B21})$$

$$4 \sum_{a,b,c,d=1}^{N_{\text{bins}}} \sum_{i,j,k,l=1}^{n_S} \langle \delta y_{a,i} \delta y_{c,k} \rangle \langle \delta y_{b,i} \delta y_{d,k} \rangle \langle \delta y_{a,j} \delta y_{d,l} \rangle \langle \delta y_{b,j} \delta y_{c,l} \rangle =$$

$$4 \sum_{a,b,c,d=1}^{N_{\text{bins}}} \sum_{i,j,k,l=1}^{n_S} \left( \delta^{ik} - \frac{1}{n_S} \right) \delta^{ac} \left( \delta^{ik} - \frac{1}{n_S} \right) \delta^{bd} \left( \delta^{jl} - \frac{1}{n_S} \right) \delta^{ad} \left( \delta^{jl} - \frac{1}{n_S} \right) \delta^{bc} = 4 N_{\text{bins}} (n_S - 1)^2 \quad (\text{B22})$$

$$4 \sum_{a,b,c,d=1}^{N_{\text{bins}}} \sum_{i,j,k,l=1}^{n_S} \langle \delta y_{a,i} \delta y_{c,k} \rangle \langle \delta y_{b,i} \delta y_{c,l} \rangle \langle \delta y_{a,j} \delta y_{d,k} \rangle \langle \delta y_{b,j} \delta y_{d,l} \rangle =$$

$$4 \sum_{a,b,c,d=1}^{N_{\text{bins}}} \sum_{i,j,k,l=1}^{n_S} \left( \delta^{ik} - \frac{1}{n_S} \right) \delta^{ac} \left( \delta^{il} - \frac{1}{n_S} \right) \delta^{bc} \left( \delta^{jk} - \frac{1}{n_S} \right) \delta^{ad} \left( \delta^{jl} - \frac{1}{n_S} \right) \delta^{bd} = 4 N_{\text{bins}} (n_S - 1) \quad (\text{B23})$$

$$8 \sum_{a,b,c,d=1}^{N_{\text{bins}}} \sum_{i,j,k,l=1}^{n_S} \langle \delta y_{a,i} \delta y_{c,k} \rangle \langle \delta y_{b,i} \delta y_{c,l} \rangle \langle \delta y_{a,j} \delta y_{d,l} \rangle \langle \delta y_{b,j} \delta y_{d,k} \rangle =$$

$$8 \sum_{a,b,c,d=1}^{N_{\text{bins}}} \sum_{i,j,k,l=1}^{n_S} \left( \delta^{ik} - \frac{1}{n_S} \right) \delta^{ac} \left( \delta^{il} - \frac{1}{n_S} \right) \delta^{bc} \left( \delta^{jl} - \frac{1}{n_S} \right) \delta^{ad} \left( \delta^{jk} - \frac{1}{n_S} \right) \delta^{bd} = 8 N_{\text{bins}} (n_S - 1) \quad (\text{B24})$$

$$4 \sum_{a,b,c,d=1}^{N_{\text{bins}}} \sum_{i,j,k,l=1}^{n_S} \langle \delta y_{a,i} \delta y_{c,k} \rangle \langle \delta y_{b,i} \delta y_{d,l} \rangle \langle \delta y_{a,j} \delta y_{c,l} \rangle \langle \delta y_{b,j} \delta y_{d,k} \rangle =$$

$$4 \sum_{a,b,c,d=1}^{N_{\text{bins}}} \sum_{i,j,k,l=1}^{n_S} \left( \delta^{ik} - \frac{1}{n_S} \right) \delta^{ac} \left( \delta^{il} - \frac{1}{n_S} \right) \delta^{bd} \left( \delta^{jl} - \frac{1}{n_S} \right) \delta^{ac} \left( \delta^{jk} - \frac{1}{n_S} \right) \delta^{bd} = 4 N_{\text{bins}}^2 (n_S - 1) \quad (\text{B25})$$

Gathering all together gives

$$\text{Var} \{ \text{tr} [(\delta \mathbf{Y})^2] \} \approx \frac{N_{\text{bins}}}{(n_S - 1)^3} [8 N_{\text{bins}}^2 + 16 N_{\text{bins}} + 8 + 4 N_{\text{bins}} (n_S - 1) + 4 (n_S - 1) + 4 + 8 + 4 N_{\text{bins}}] =$$

$$\frac{4 N_{\text{bins}}}{(n_S - 1)^3} [2 N_{\text{bins}}^2 + 4 N_{\text{bins}} + 4 + (N_{\text{bins}} + 1) n_S] = \frac{4 N_{\text{bins}}}{(n_S - 1)^3} [(N_{\text{bins}} + 1) (n_S + 2 N_{\text{bins}} + 2) + 2]. \quad (\text{B26})$$

Then, according to equation (B15), the variance of the KL divergence is approximately 16 times smaller:

$$\text{Var}[D_{\text{KL}}(\Psi_0, \mathbf{X})] \approx \frac{N_{\text{bins}} [(N_{\text{bins}} + 1) (n_S + 2 N_{\text{bins}} + 2) + 2]}{4 (n_S - 1)^3}. \quad (\text{B27})$$

This is an approximation because in equation (B15), the Taylor expansion was truncated at the quadratic order in  $\delta \mathbf{Y}$ . Other derivation steps are exact.

## B2 Inverse test

We considered how different  $\chi^2$  would result from one matrix compared to the other. If  $\mathbf{u}$  is a unit vector in  $N_{\text{bins}}$ -dimensional space ( $\mathbf{u}^T \mathbf{u} = 1$ ), then  $\mathbf{w} = \mathbf{C}_2^{1/2} \mathbf{u}$  gives a unit  $\chi^2$  according to  $\mathbf{C}_2$ :  $\mathbf{w}^T \mathbf{C}_2^{-1} \mathbf{w} = 1$ . Here,  $\mathbf{C}_2^{1/2}$  means the matrix square root of  $\mathbf{C}_2$  – a matrix with the same eigenvectors and eigenvalues equal to the square roots of corresponding eigenvalues of the original matrix. Then we consider the  $\chi^2$  with respect to  $\mathbf{C}_1$  ( $\Psi_1$ ):  $\mathbf{w}^T \Psi_1 \mathbf{w}$  and subtract the expected value of 1:

$$\mathbf{w}^T \Psi_1 \mathbf{w} - 1 = \mathbf{u}^T \left[ \mathbf{C}_2^{1/2} \Psi_1 \mathbf{C}_2^{1/2} - \mathbb{1} \right] \mathbf{u}. \quad (\text{B28})$$

Taking the RMS over all directions of  $\mathbf{u}$ , one arrives at the RMS eigenvalue of this matrix, which can be expressed through the Frobenius norm:

$$R_{\text{inv}}(\Psi_1, \mathbf{C}_2) = \frac{1}{\sqrt{N_{\text{bins}}}} \left\| \mathbf{C}_2^{1/2} \Psi_1 \mathbf{C}_2^{1/2} - \mathbb{1} \right\|_F. \quad (\text{B29})$$

The Frobenius norm can be recast as a trace and simplified further using its cyclic property:

$$\left\| \mathbf{C}_2^{1/2} \Psi_1 \mathbf{C}_2^{1/2} - \mathbb{1} \right\|_F^2 = \text{tr} \left[ \left( \mathbf{C}_2^{1/2} \Psi_1 \mathbf{C}_2^{1/2} - \mathbb{1} \right)^T \left( \mathbf{C}_2^{1/2} \Psi_1 \mathbf{C}_2^{1/2} - \mathbb{1} \right) \right] = \text{tr} \left[ \left( \mathbf{C}_2^{1/2} \Psi_1 \mathbf{C}_2^{1/2} - \mathbb{1} \right)^2 \right]$$

$$= \text{tr} \left( \mathbf{C}_2^{1/2} \Psi_1 \mathbf{C}_2 \Psi_1 \mathbf{C}_2^{1/2} - 2 \mathbf{C}_2^{1/2} \Psi_1 \mathbf{C}_2^{1/2} + \mathbb{1} \right) = \text{tr} (\Psi_1 \mathbf{C}_2 \Psi_1 \mathbf{C}_2 - 2 \Psi_1 \mathbf{C}_2 + \mathbb{1}) = \text{tr} \text{tr} [(\Psi_1 \mathbf{C}_2 - \mathbb{1})^2]. \quad (\text{B30})$$

This allows us to compute the quantity as

$$R_{\text{inv}}(\Psi_1, \mathbf{C}_2) = \sqrt{\frac{\text{tr} [(\Psi_1 \mathbf{C}_2 - \mathbb{1})^2]}{N_{\text{bins}}}}, \quad (\text{B31})$$



which is more computationally robust – it is better to avoid factorizing (and inverting) covariance matrices (especially the sample one) when possible.

It is notable that this metric is related to the discriminant matrix

$$\mathbf{P} = \sqrt{\Psi_R}^T \mathbf{C}_S \sqrt{\Psi_R} - \mathbb{I}, \quad (\text{B32})$$

where  $\sqrt{\Psi_R}$  is the lower Cholesky decomposition, used in Philcox & Eisenstein (2019) and Philcox et al. (2020). Through a similar procedure, one can find that its Frobenius norm is the same as above:

$$\|\mathbf{P}\|_F^2 = \text{tr}[(\Psi_1 \mathbf{C}_2 - \mathbb{I})^2]. \quad (\text{B33})$$

However, the interpretation of elements of the discriminant matrix is less clear.

Now let us consider the square of this metric (to remove the root):

$$R_{\text{inv}}^2(\Psi_0, \mathbf{X}) = \frac{1}{N_{\text{bins}}} \text{tr}[\Psi_0 X \Psi_0 X - 2\Psi_0 X + \mathbb{I}]. \quad (\text{B34})$$

Remembering equation (B8), we arrive to

$$R_{\text{inv}}^2(\Psi_0, \mathbf{X}) = \frac{1}{N_{\text{bins}}} \text{tr}[\mathbf{Y}^2 - 2\mathbf{Y} + \mathbb{I}]. \quad (\text{B35})$$

Furthermore, expanding in  $\delta\mathbf{Y}$  [according to equation (B10)], we get

$$R_{\text{inv}}^2(\Psi_0, \mathbf{X}) = \frac{1}{N_{\text{bins}}} \text{tr}[(\delta\mathbf{Y})^2], \quad (\text{B36})$$

which is similar to equation (B15) up to a constant factor of  $1/N_{\text{bins}}$  and lack of approximations. We can obtain the expectation value by plugging in equation (B11):

$$\langle R_{\text{inv}}^2(\Psi_0, \mathbf{X}) \rangle = \frac{N_{\text{bins}} + 1}{n_S - 1}. \quad (\text{B37})$$

For variance, we can use equation (B26):

$$\text{Var}[R_{\text{inv}}^2(\Psi_0, \mathbf{X})] = \frac{4[(N_{\text{bins}} + 1)(n_S + 2N_{\text{bins}} + 2) + 2]}{N_{\text{bins}}(n_S - 1)^3}. \quad (\text{B38})$$

Assuming  $\text{Var}[R_{\text{inv}}^2(\Psi_0, \mathbf{X})] \ll \langle R_{\text{inv}}^2(\Psi_0, \mathbf{X}) \rangle^2$ , we can take the square root to estimate the mean and variance of not-squared metric as

$$\langle R_{\text{inv}}(\Psi_0, \mathbf{X}) \rangle \approx \sqrt{\langle R_{\text{inv}}^2(\Psi_0, \mathbf{X}) \rangle} = \sqrt{\frac{N_{\text{bins}} + 1}{n_S - 1}}, \quad (\text{B39})$$

$$\text{Var}[R_{\text{inv}}(\Psi_0, \mathbf{X})] \approx \frac{\text{Var}[R_{\text{inv}}^2(\Psi_0, \mathbf{X})]}{4\langle R_{\text{inv}}^2(\Psi_0, \mathbf{X}) \rangle} = \frac{(N_{\text{bins}} + 1)(n_S + 2N_{\text{bins}} + 2) + 2}{N_{\text{bins}}(N_{\text{bins}} + 1)(n_S - 1)^2}. \quad (\text{B40})$$

Also, it is useful to note that  $R_{\text{inv}}$  is related to the  $\chi^2$  approximation of minus log-likelihood, considering the covariance of all covariance matrix elements. This is rather easy to see with  $\delta\mathbf{Y}$  expression (equation (B36)). Since  $\delta\mathbf{Y}$  is real and symmetric, we obtain

$$N_{\text{bins}} \times R_{\text{inv}}^2(\Psi_0, \mathbf{X}) = \|\delta\mathbf{Y}\|_F^2 = \sum_{a,b=1}^{N_{\text{bins}}} (\delta Y_{ab})^2. \quad (\text{B41})$$

From equation (B11), we conclude that distinct elements of  $\mathbf{Y}$  matrix have zero covariance (excluding the pairs symmetric with respect to the diagonal), its diagonal elements have a variance of  $2/(n_S - 1)$  and off-diagonal elements have a variance of  $1/(n_S - 1)$ . Then this covariance is trivial to invert and we have to sum squares of deviation of independent elements divided by their variance to get the  $\chi^2$ :

$$\chi^2 = (n_S - 1) \left[ \frac{1}{2} \sum_{a=1}^{N_{\text{bins}}} (\delta Y_{aa})^2 + \sum_{a=1}^{N_{\text{bins}}} \sum_{b=a+1}^{N_{\text{bins}}} (\delta Y_{ab})^2 \right] = \frac{n_S - 1}{2} \sum_{a,b=1}^{N_{\text{bins}}} (\delta Y_{ab})^2 = \frac{(n_S - 1)N_{\text{bins}}}{2} \times R_{\text{inv}}^2(\Psi_0, \mathbf{X}). \quad (\text{B42})$$

Since elements of  $\mathbf{Y}$  are independent linear combinations of elements of sample covariance matrix  $\mathbf{X}$  (via equation (B8), since  $\Psi_0$  is not degenerate), the same holds for  $\mathbf{X}$ , but a direct computation without the ‘rotation’ into  $\mathbf{Y}$  would be significantly longer as the covariance of  $X_{ab}$  elements [equation (B2)] has a more generic and complex structure.

### B3 Mean chi-squared

We consider the sum of  $\chi^2$  associated with the deviation of data vectors in individual samples from the estimate of average:

$$\sum_{i=1}^{n_S} \sum_{a,b=1}^{N_{\text{bins}}} (x_{a,i} - \bar{x}_a) \Psi_{0,ab} (x_{b,i} - \bar{x}_b) = (n_S - 1) \sum_{a,b=1}^{N_{\text{bins}}} \Psi_{0,ab} X_{ab} = (n_S - 1) \text{tr}(\Psi_0 X). \quad (\text{B43})$$

**Table B1.** Theoretical versus sampled mean  $\pm$  std of covariance matrix comparison metrics with the true precision matrix. A close agreement can be seen.

		$D_{\text{KL}}(\Psi_0, X)$	$R_{\text{inv}}(\Psi_0, X)$	$\chi_{\text{red}}^2(\Psi_0, X)$
Measurement space (45 bins)	Theoretical	$0.519 \pm 0.024$	$0.2147 \pm 0.0049$	$1.0000 \pm 0.0067$
	Sampled	$0.526 \pm 0.023$	$0.2146 \pm 0.0050$	$1.0000 \pm 0.0067$
Parameter space (5 quantities projected from 25 bins)	Theoretical	$0.0075 \pm 0.0028$	$0.078 \pm 0.014$	$1.000 \pm 0.020$
	Sampled	$0.0077 \pm 0.0028$	$0.077 \pm 0.014$	$1.000 \pm 0.020$

Remembering equation (B5), we can rewrite the LHS as

$$\sum_{i=1}^{n_S} \sum_{a,b=1}^{N_{\text{bins}}} (y_{a,i} - \bar{y}_a) \delta_{ab} (y_{b,i} - \bar{y}_b) = \sum_{a=1}^{N_{\text{bins}}} \sum_{i=1}^{n_S} (y_{a,i} - \bar{y}_a)^2 \sim \sum_{a=1}^{N_{\text{bins}}} \chi^2(n_S - 1) \sim \chi^2[N_{\text{bins}} \times (n_S - 1)]. \quad (\text{B44})$$

Therefore the corresponding reduced  $\chi^2$  is

$$\chi_{\text{red}}^2 = \frac{1}{N_{\text{bins}}} \text{tr}(\Psi_0 \mathbf{X}). \quad (\text{B45})$$

#### B4 Validation of means and standard deviations

To check the theoretical results from the previous sections, we have performed a quick Monte Carlo validation. 10 000 batches of 999 samples having 45 bins each have been generated. For simplicity, we have taken the true covariance and precision to be unity matrices  $C_0 = \Psi_0 = \mathbb{I}$ . This should not affect the results, save for numerical instabilities in matrix operations. Full 45-bin sample covariance and precision matrices have been estimated in each batch. Then, 25 bins were selected and projected into 5 quantities, for simplicity using 5 random orthonormal vectors as parameter derivatives. Comparison between theoretical and sampled means and standard deviations are presented in Table B1. Differences are most pronounced in  $D_{\text{KL}}$ , but the disagreement is only in the second digit of standard deviation and fractions of standard deviation on the mean. Therefore, we report a close agreement, more than enough for the main part of the paper, where the scatter of RASCALC results is significantly larger than these standard deviations. We note that the results for  $D_{\text{KL}}$  and  $R_{\text{inv}}$  [equations (B4), (B27), (B39), and (B40)] are approximate and we expect meaningful deviations from them, especially as  $N_{\text{bins}}$  increases, while the derivations for  $R_{\text{inv}}^2$  and  $\chi_{\text{red}}^2$  [equations (B37), (B37), and (B44)] are exact.

We have repeated this test with a realistic covariance matrix (RASCALC Average NG for pre-reconstruction) and derivatives of the observables with respect to the parameters (accordingly, for the BAO model before reconstruction) to confirm whether the comparison measures are indeed not affected. We have obtained the same numbers as in simpler test (Table B1), and close agreement of  $\sigma[\sigma(\alpha_{\text{BAO}})]/\sigma(\alpha_{\text{BAO}})$  with  $[2(n_S - 1)]^{-1/2}$  according to equation (34).

<sup>1</sup>Center for Astrophysics| Harvard & Smithsonian, 60 Garden St, Cambridge, MA 02138, USA

<sup>2</sup>Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

<sup>3</sup>Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK

<sup>4</sup>Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK

<sup>5</sup>Department of Physics and Astronomy, The University of Utah, 115 South 1400 East, Salt Lake City, UT 84112, USA

<sup>6</sup>Instituto de Física, Universidad Nacional Autónoma de México, Cd. de México C.P. 04510, México

<sup>7</sup>Center for Cosmology and AstroParticle Physics, The Ohio State University, 191 West Woodruff Avenue, Columbus, OH 43210, USA

<sup>8</sup>Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra Barcelona, Spain

<sup>9</sup>Departamento de Física & Observatorio Astronómico, Universidad de los Andes, Cra. 1 No. 18A-10, Edificio Ip, CP 111711, Bogotá, Colombia

<sup>10</sup>Department of Astrophysical Sciences, Princeton University, Princeton NJ 08544, USA

<sup>11</sup>Department of Physics, Southern Methodist University, 3215 Daniel Avenue, Dallas, TX 75275, USA

<sup>12</sup>Institució Catalana de Recerca i Estudis Avançats, Passeig de Lluís Companys, 23, E-08010 Barcelona, Spain

<sup>13</sup>Department of Physics and Astronomy, Sejong University, Seoul 143-747, Korea

<sup>14</sup>Max Planck Institute for Astrophysics, Karl-Schwarzschild-Str 1, D-85748 Garching, Germany

<sup>15</sup>Institute of Cosmology & Gravitation, University of Portsmouth, Dennis Sciama Building, Portsmouth PO1 3FX, UK

<sup>16</sup>National Astronomical Observatories, Chinese Academy of Sciences, A20 Datun Rd., Chaoyang District, Beijing, 100012, PR China

<sup>17</sup>Space Sciences Laboratory, University of California, Berkeley, 7 Gauss Way, Berkeley, CA 94720, USA

<sup>18</sup>CIEMAT, Avenida Complutense 40, E-28040 Madrid, Spain

<sup>19</sup>Korea Astronomy and Space Science Institute, 776, Daedeokdae-ro, Yuseong-gu, Daejeon 34055, Republic of Korea

<sup>20</sup>Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA

<sup>21</sup>Department of Physics & Astronomy, Ohio University, Athens, OH 45701, USA

<sup>22</sup>NSF's NOIRLab, 950 N. Cherry Ave., Tucson, AZ 85719, USA

<sup>23</sup>Department of Astronomy, Tsinghua University, 30 Shuangqing Road, Haidian District, Beijing, 100190, China

<sup>24</sup>Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

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