Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **455,** 3871–3889 (2016) doi:10.1093/mnras/stv2588

Fourier analysis of multitracer cosmological surveys

L. Raul Abramo,* Lucas F. Secco and Arthur Loureiro

Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, CEP 05314-970 São Paulo, Brazil

Accepted 2015 November 1. Received 2015 October 13; in original form 2015 June 1

ABSTRACT

We present optimal quadratic estimators for the Fourier analysis of cosmological surveys that detect several different types of tracers of large-scale structure. Our estimators can be used to simultaneously fit the matter power spectrum and the biases of the tracers – as well as redshift-space distortions (RSDs), non-Gaussianities (NGs), or any other effects that are manifested through differences between the clusterings of distinct species of tracers. Our estimators reduce to the one by Feldman, Kaiser & Peacock (FKP) in the case of a survey consisting of a single species of tracer. We show that the multitracer estimators are unbiased, and that their covariance is given by the inverse of the multitracer Fisher matrix. When the biases, RSDs and NGs are fixed to their fiducial values, and one is only interested in measuring the underlying power spectrum, our estimators are projected into the estimator found by Percival, Verde & Peacock. We have tested our estimators on simple (lognormal) simulated galaxy maps, and we show that it performs as expected, being either equivalent or superior to the FKP method in all cases we analysed. Finally, we have shown how to extend the multitracer technique to include the one-halo term of the power spectrum.

Key words: cosmology: theory – large-scale structure of Universe.

1 INTRODUCTION

Astrophysical surveys have come to occupy a central role in cosmology (York et al. 2000; Cole et al. 2005; Abbott et al. 2005; Scoville et al. 2007; Adelman-McCarthy et al. 2008; PAN-STARRS; Tony et al. 2012; Dawson et al. 2012; Blake et al. 2011). Percent-level accuracies can now be reached on distance measurements at high redshifts, both across and along the line of sight (Anderson et al. 2012, 2014). This means more and better constraints on the acceleration of the expansion rate of the Universe, on modified gravity (Linder 2005), on non-Gaussian initial conditions (Verde et al. 2000; Bartolo et al. 2004), and about the role of massive neutrinos, among other applications.

The ground-breaking achievements of the Sloan Digital Sky Survey (York et al. 2000) are being surpassed by other surveys with higher completeness, wider wavelength coverage, and a larger range of redshifts (Schlegel et al. 2009; Ellis et al. 2012; Abell et al. 2009). Some surveys will specialize in mapping very large volumes to an extremely high completeness, by employing imaging with narrow-band filters (Benítez et al. 2009, 2014), or by resorting to low-resolution integral-field spectroscopy (Hill et al. 2008). In addition to galaxies, quasars can also serve both as sources of background light to investigate the intervening matter through the Ly α forest (Slosar et al. 2013), or directly as tracers of the large-scale structure (Croom et al. 2005; da Ângela et al. 2005; Yahata et al. 2005; Shen et al. 2007; Ross et al. 2009; Sawangwit et al. 2012; Abramo et al. 2012; Leistedt et al. 2013; Leistedt & Peiris 2014). Another way in which the 3D matter distribution can be mapped is through the 21 cm hyperfine transition of neutral H, and new radiotelescopes dedicated to measuring that line are being deployed or are in the planning stages (Bandura et al. 2014; Battye et al. 2013).

This intense activity points to an exciting future, where vast volumes of the Universe will be increasingly mapped in a variety of ways, and with different types of tracers of the large-scale structure.

However, these maps are not independent, since the different species of galaxies tracer the same distribution of dark matter in slightly different ways (Guzzo et al. 1997). This points to a key obstacle on the way to explore the full power of these overlapping surveys: cosmic variance – a particular case of sample variance, where the sample is the set of modes of the density perturbations which were realized in some region of the Universe from a (nearly) Gaussian random process.

Despite this fundamental limitation, it was pointed out by Seljak (2009) and McDonald & Seljak (2009) that the bounds imposed by cosmic variance do not apply for many key physical quantities of interest. In particular, by comparing the clustering of tracers with different biases one can measure some parameters to an accuracy which is basically unconstrained by cosmic variance. This applies not only to bias itself, but to the matter growth rate of the redshift-space distortions (RSDs), to the non-Gaussianity (NG) parameters $f_{\rm NL}$ and $g_{\rm NL}$, as well as any parameters that affect the relative clusterings of different tracers. The consequences of this extraordinary windfall were explored in many papers – see, e.g. Slosar (2009),

* E-mail: abramo@fma.if.usp.br

Gil-Marín et al. (2010), Cai & Bernstein (2011a), Cai & Bernstein (2011b), Hamaus, Seljak & Desjacques (2011), Smith, Desjacques & Marian (2011), Hamaus, Seljak & Desjacques (2012), Abramo & Leonard (2013), Blake et al. (2013), Bull et al. (2015) and Ferramacho et al. (2014). It is important to stress that this additional information comes from measuring the two-point functions of the different tracers, as opposed to enhancing the signal to noise by employing different statistics such as mass weighting (Seljak, Hamaus & Desjacques 2009; Cai, Bernstein & Sheth 2011; Smith & Marian 2014).

In Abramo & Leonard (2013), we showed that the enhanced constraints of multitracer cosmological surveys are a straightforward consequence of the multitracer Fisher information matrix (Abramo 2012). In the presence of N different types of tracers, each one with a different bias, there is a simple choice of variables which diagonalizes the multitracer Fisher matrix: in addition to the underlying power spectrum (which is subject to the cosmic variance bounds), there are $N-1$ variables which correspond to the relative clustering strengths between the tracers. These relative clustering strengths are not affected by cosmic variance, and their measurements can be arbitrarily accurate, even if the survey has a finite volume. In the case of a single tracer, the multitracer Fisher matrix reduces to the usual case treated in the seminal paper by Feldman, Kaiser & Peacock (1994, henceforth FKP).

In this paper, we use the multitracer Fisher matrix to derive optimal estimators for the redshift-space power spectra of an arbitrary number of different types of tracers of large-scale structure (see Section 3). These tracers may overlap in some regions but not others, or not at all. The tracers can be galaxies of different types, quasars, Ly α absorption systems, etc. One may also choose to trade individual objects by haloes of different masses – in which case the bias of the tracers become the halo bias (Seljak et al. 2009; Hamaus et al. 2010; Cai et al. 2011; Smith & Marian 2014, 2015).

Our formulas can be used in any of those situations, in real or in redshift space, including effects such as scale-dependent bias. An important cross-check is that a particular combination (or projection) of our estimators leads back to the estimator of Percival, Verde & Peacock (2004, henceforth PVP), – namely, the PVP estimator follows from ours if the biases of the tracers, as well as the RSDs, are fixed to their true values, and one then computes the underlying matter power spectrum using the aggregated clustering information from all the tracers.

We have also incorporated the contribution of the one-halo term to the covariance of the galaxy counts (see Section 6) into the multitracer Fisher matrix. In principle, the full covariance can be computed using the Halo Model (Cooray & Sheth 2002), together with appropriate halo occupation distributions (HODs) for the tracers. Recently, Smith & Marian (2015) presented an optimal estimator for the power spectrum including all Halo Model corrections to the spectrum, bispectrum and trispectrum. However, the Smith & Marian (2015) estimator generalizes the estimator of Percival et al. (2004), while we have obtained estimators not only for the matter power spectrum, but also for the biases, RSDs, NGs, etc. Hence, we are only able to include the simplest correction to galaxy clustering from the Halo Model (the one-halo term), but our framework allows the simultaneous estimation of bias, RSDs and the power spectrum, while Smith & Marian (2015), include all the corrections that can be computed on the basis of the Halo Model, but their estimator applies only to the power spectrum, and assumes prior knowledge of the bias of all the species, as well as of the shape of the RSDs.

The Fisher matrix and the covariance of the counts of the tracers are the basic objects used to construct our estimators – in fact, the optimal estimators are a type of Wiener filtering, in the sense that we are basically weighting the data by the inverse of their covariance. We employ results and notation introduced in Abramo (2012) and Abramo & Leonard (2013), and the construction of the estimators follows the steps outlined in Tegmark (1997), Tegmark et al. (1998), Hamilton (2005a) and Hamilton (2005b).

The estimators were tested using simple mock galaxy catalogues based on lognormal maps (see Section 5). We find that the empirical covariance of the power spectra is well approximated by the theoretical covariance (the inverse of the multitracer Fisher matrix), confirming the optimality of the estimator.

The main formulas of this paper are derived in Section 3, and a practical algorithm for the Fourier analysis of multitracer surveys is summarized in Section 5.2.

This paper is organized as follows: in Section 2 we review the Fisher information matrix for single-tracer and multitracer cosmological surveys. In Section 3, we construct the optimal quadratic estimators on the basis of the covariance matrix for the data. In Section 4, we discuss the relationships between the multitracer estimators and other methods, such as FKP and PVP, as well as the main features of the multitracer technique. In Section 5, we test the estimators in simple simulated maps, showing that the empirical covariance matches closely the theoretical covariance – which establishes that the multitracer estimators are unbiased and optimal. In Section 6, we show how to include the one-halo term in the estimators of the multitracer power spectra. We also show there how to construct estimators for the one-halo term, and how to generalize the procedure to estimate simultaneously the two-halo and the one-halo term of the power spectrum. We conclude in Section 7.

2 THE INFORMATION IN GALAXY SURVEYS

The matter power spectrum is defined through the expectation value $\langle\langle\delta_{\rm m}(\boldsymbol{k},z)\delta^*_{\rm m}(\boldsymbol{k}',z)\rangle\rangle = (2\pi)^3 P_{\rm m}(k,z)\delta_{\rm D}(\boldsymbol{k}-\boldsymbol{k}')$, where $\delta_{\rm m}(\boldsymbol{k})$ is the matter density contrast, and $\delta_{\rm D}$ is the Dirac delta function. However, galaxy surveys actually measure counts of tracers of the large-scale structure (galaxies and other extragalactic objects) in redshift space. It is from those observable that we can then measure derived quantities such as the baryon acoustic oscillations (Eisenstein, Hu & Tegmark 1999; Blake & Glazebrook 2003; Seo & Eisenstein 2003), or the pattern of redshift-space distortions (Kaiser 1987).

For a tracer of type α (greek indices refer to different types of galaxies, or tracers), with counts per unit volume $n_\alpha(\boldsymbol{x})$, the density contrast is $\delta_\alpha(\boldsymbol{x}) = n_\alpha(\boldsymbol{x})/\bar{n}_\alpha(\boldsymbol{x}) - 1$. The mean number densities $\bar{n}_\alpha$ should reflect the spatial modulations of the observed numbers of galaxies which are due to the instrument, the strategy and schedule of observations, as well as any other factors unrelated to the redshift-space cosmological fluctuations.

If we assume that bias is linear and deterministic, then in the distant-observer approximation the redshift-space fluctuations in the counts of the α-type galaxies are related to the underlying mass fluctuations by the relation $\delta_\alpha(\boldsymbol{k},z) \simeq [b_\alpha + f\,\mu_k^2]\delta_{\rm m}(\boldsymbol{k},z)$. Here b_α is the bias of the tracer species α, $f(z)$ is the matter growth rate, and $\mu_k = \hat{k}\cdot\hat{r}$ is the cosine of the angle between the Fourier mode and the line of sight.

The index α (we employ greek letters to denote different tracer species) can be any kind of discriminant of the types of tracers of large-scale structure: it may stand for luminosity, morphological

type, star formation rate, equivalent width of some emission line, or a combination of those. One may also regard the dark matter haloes themselves as the tracers, in which case α would stand for the halo mass (or some proxy for it, such as richness), and the bias b_α becomes halo bias.

There are some complicating factors in this description. First, structure formation should introduce a scale dependence for the bias, as well as some degree of stochasticity (Benson et al. 2000; Dekel & Lahav 1999; Berlind & Weinberg 2002; Smith, Scoccimarro & Sheth 2009). Secondly, the initial conditions may contain non-Gaussian features, which would manifest themselves as an additional scale-dependent bias (Bartolo et al. 2004; Sefusatti & Komatsu 2007; Dalal et al. 2008). Thirdly, the velocity dispersion from random motions inside haloes will smear the galaxy density contrast, affecting the shape of RSDs. In fact, the RSD parameters and angular dependence can inherit scale-dependent non-linear corrections (Raccanelli et al. 2012)

Hence, in practice it is more useful to regard 'bias' as a more general function of redshift, scale, and angle with the line of sight that should be determined by observations, and *define* the clustering of a species α as

$$P_\alpha \equiv B_\alpha^2 \, P_{\rm m}, \tag{1}$$

where $B_\alpha = b_\alpha + f\mu_k^2 + \Delta b_{\rm NG}$ is an *effective bias*. This effective bias includes the tracer bias, RSDs, NGs, as well as any mechanism that distorts the power spectrum of the tracers relative to the spectrum of the underlying matter distribution.

In principle, everything depends on $\boldsymbol{x}$ and $\boldsymbol{k}$, but one can regard x (i.e. the radial position) as standing in for redshift, so we could also write $B_\alpha = B_\alpha(z, k, \mu_k)$, and $P_{\rm m} = P_{\rm m}(z, k)$. Since the matter power spectrum, bias, RSDs, as well as NGs and other corrections, are just subproducts of clustering measurements for all the available tracers in a given survey, the problem we must address is how one can optimally estimate the power spectra of the tracers, $P_\alpha(\boldsymbol{x}, \boldsymbol{k})$.

Our approach also means that, as a first approximation, cross-correlations are expressed as $P_{\alpha\beta} = B_\alpha B_\beta P_{\rm m}$. This is always valid on large ($k \lesssim 1\,h\,{\rm Mpc}^{-1}$) scales, where the contribution of the two-halo term dominates, but on small scales the one-halo term can invalidate that assumption. That case is considered in Section 6.

2.1 Optimal estimators and the Fisher matrix

The Fisher information matrix can be constructed from the data covariance after a series of simple steps – for a review in the specific context of cosmological surveys, see, e.g. Tegmark, Taylor & Heavens (1997), Tegmark (1997) and Tegmark et al. (1998).

Consider some data d_x (where for the moment we can consider x to be a discrete index) with vanishing expectation values, $\langle\langle d_x\rangle\rangle = 0$, and covariance ${\rm Cov}(d_x, d_{x'}) = \langle\langle d_x d_{x'}\rangle\rangle$. From this data covariance we would like to estimate a set of parameters θ^i given, some initial (fiducial) guess for the parameters $\bar\theta^i$.

If we have a model for the data covariance in terms of the parameters, ${\rm Cov}(d_x, d_{x'}) \to C_{xx'}(\theta^i)$, then we can compute the Fisher information matrix for this model. Assuming that the likelihood is approximately described by a multivariate Gaussian, and that the maximum of the likelihood lies near the fiducial values of the parameters, the Fisher information matrix is given by the expectation value:

$$\begin{aligned} F_{ij} \equiv F[\theta^i, \theta^j] &= -\left\langle\left\langle \frac{\partial^2 \log\mathcal{L}}{\partial\theta^i\partial\theta^j}\right\rangle\right\rangle \\ &\simeq \frac{1}{2}\sum_{xyx'y'} C^{-1}_{xy}\frac{\partial C_{yy'}}{\partial\theta^i}C^{-1}_{y'x'}\frac{\partial C_{x'x}}{\partial\theta^j}, \\ &= \frac{1}{2}\,{\rm Tr}\left[C^{-1}\frac{\partial C}{\partial\theta^i}C^{-1}\frac{\partial C}{\partial\theta^j}\right]. \end{aligned} \tag{2}$$

Now suppose that *estimators* $\hat\theta^i$ can be constructed, such that their covariance is given by ${\rm Cov}(\hat\theta^i, \hat\theta^j) = F^{-1}_{ij}$. These estimators must then be optimal, in the sense that they saturate the Cramér-Rao bound: ${\rm Cov}(\hat\theta^i, \hat\theta^j) \geq F^{-1}_{ij}$.

There are in fact such estimators, which can be constructed after a few simple steps, and which employ the same basic objects that appear in the Fisher matrix. The first step is to create the quadratic form:

$$\hat q^i \equiv \sum_{xx'} d_x \, E^i_{xx'} \, d_{x'} - \Delta q^i, \tag{3}$$

where the weighting matrix is

$$E^i_{xx'} = \frac{1}{2}\sum_{yy'} C^{-1}_{xy}\frac{\partial C_{yy'}}{\partial\theta^i}C^{-1}_{y'x'}, \tag{4}$$

and Δq^i subtracts the *bias of the estimator*.

Gaussianity of the data (a key underlying assumption) implies that the four-point function $\langle\langle d_x d_y d_{x'} d_{y'}\rangle\rangle \to C_{xy}C_{x'y'} + C_{xx'}C_{yy'} + C_{xy'}C_{x'y}$, and from it follows, after some algebra, that the covariance of the quadratic form above is ${\rm Cov}(\hat q^i, \hat q^j) = F_{ij}$.

The final step is to define the optimal quadratic estimators in such a way that their covariance is the *inverse* of the Fisher matrix. Clearly, then, the estimators:

$$\hat\theta^i = \sum_j F^{-1}_{ij}\,\hat q^j, \tag{5}$$

satisfy that condition. Finally, with the definition:

$$\begin{aligned} \Delta q^i &\equiv \sum_{xx'} E^i_{xx'} C_{x'x} - \sum_j F_{ij}\,\theta^j \\ &= \frac{1}{2}\sum_{xy} C^{-1}_{xy}\frac{\partial C_{yx}}{\partial\theta^i} - \sum_j F_{ij}\,\theta^j, \end{aligned} \tag{6}$$

we obtain estimators which are also *unbiased*. This means that if the estimators are evaluated assuming fiducial values for the parameters $\theta^i \to \bar\theta^i$ which are not far from the true values, and as long as the data covariance is correctly modelled by the theoretical covariance, then the expectation values of the estimators converge to the fiducial values in the sense that $\langle\langle\,(\hat\theta^i - \bar\theta^i)^2\,\rangle\rangle \to 0$.

2.2 Single-species Fisher matrix

The most basic sources of uncertainty in galaxy surveys are cosmic variance and shot noise. In cosmological surveys which target a single species of tracer, the optimal estimator for the galaxy power spectrum was derived by FKP (Feldman et al. 1994). The corresponding Fisher information matrix was derived by Tegmark (1997) and Tegmark et al. (1998), who also showed that the FKP estimator follows from the construction presented in the previous section.

As will be shown in the next section, the FKP Fisher matrix for a survey of a galaxy of type α can be written as

$$F[\theta^i, \theta^j] = \int \frac{\mathrm{d}^3x\,\mathrm{d}^3k}{(2\pi)^3}\,\frac{\partial \log P_\alpha}{\partial\theta^i}\,\mathcal{F}_\alpha\,\frac{\partial \log P_\alpha}{\partial\theta^j}, \tag{7}$$

where the Fisher *information density* in phase space associated with the tracer α is

$$\mathcal{F}_\alpha(\boldsymbol{x},\boldsymbol{k}) = \frac{1}{2}\left(\frac{\mathcal{P}_\alpha}{1+\mathcal{P}_\alpha}\right)^2. \tag{8}$$

In the expression above we have defined a dimensionless 'clustering strength' of the tracer α, which is just the power spectrum in units of (Poissonian) shot noise:

$$\mathcal{P}_\alpha(\boldsymbol{x},\boldsymbol{k}) \equiv \bar{n}_\alpha(\boldsymbol{x})P_\alpha(\boldsymbol{x},\boldsymbol{k}). \tag{9}$$

In the limit of arbitrarily high clustering strength ($1/\bar{n}_\alpha \to 0$, $\mathcal{P}_\alpha \to \infty$), the Fisher information density saturates the limit $\mathcal{F}_\alpha \to \frac{1}{2}$. Hence, for a survey of a single species of tracer there is an upper limit to the information which can be extracted from a finite volume and from a finite range of Fourier modes. This is nothing but a restatement of the limits imposed by cosmic variance.

At this point it is useful to recall how, in practice, one can extract limits on the amplitude of the power spectrum out of the Fisher matrix. In that case, the parameters of the Fisher matrix are the values of the matter power spectrum of the tracer α at a given bandpower (k_i, μ_i):

$$\theta^i \to P^i_\alpha \equiv \langle P_\alpha(\boldsymbol{x},\boldsymbol{k})\rangle^i = \frac{1}{V_i\,\tilde{V}_i}\int_{V_i}\mathrm{d}^3x\int_{\tilde{V}_i}\frac{\mathrm{d}^3k}{(2\pi)^3}\,P_\alpha(\boldsymbol{x},\boldsymbol{k}), \tag{10}$$

where V_i is the volume of the survey (e.g. a particular redshift slice z_i), and $\tilde{V}_i = 2\pi k_i^2\Delta k_i\,\Delta\mu_i$ is the volume associated with the ith bandpower (i.e. the bin in Fourier space). Hereafter, $\langle\cdots\rangle^i$ denotes an *average* over the phase space bin, and $\langle\cdots\rangle^{k_i}$ denotes an average over the Fourier-space bin (the bandpower) – which should not be confused with *expectation values*, expressed here as $\langle\langle\cdots\rangle\rangle$.

In this case we must compute the Jacobian $\partial P_\alpha(\boldsymbol{x},\boldsymbol{k})/\partial P^i_\alpha$ inside equation (7). It is useful to regard such an object in terms of functional derivatives.[1] Using:

$$\frac{\partial f(\boldsymbol{x},\boldsymbol{k})}{\partial f(\boldsymbol{x}',\boldsymbol{k}')} = (2\pi)^3\delta_{\rm D}(\boldsymbol{x}-\boldsymbol{x}')\delta_{\rm D}(\boldsymbol{k}-\boldsymbol{k}'), \tag{11}$$

one can see that the inverse of the Jacobian is

$$\frac{\partial P^i_\alpha}{\partial P_\alpha(\boldsymbol{x},\boldsymbol{k})} = \frac{1}{V_i\,\tilde{V}_i}. \tag{12}$$

But this is simply the inverse of the *phase space volume* $v_i = V_i\,\tilde{V}_i$. Therefore, the inverse of this Jacobian, $\partial P_\alpha(\boldsymbol{x},\boldsymbol{k})/\partial P^i_\alpha$, has the effect of limiting integrations in phase space, $\int \mathrm{d}^3x\,\mathrm{d}^3k/(2\pi)^3[\cdots]$, to the phase space volume of the bin i. Since this type of object will reappear later on, we employ the notation $\delta^i_{\boldsymbol{x},\boldsymbol{k}}$ to express the restriction of a phase space integral to a certain volume V_i, and we use the same notation to indicate restrictions in integrals over position space, $\delta^i_{\boldsymbol{x}}$, or Fourier space, $\delta^i_{\boldsymbol{k}}$. Hence, according to this notation:

$$\int \frac{\mathrm{d}^3x\,\mathrm{d}^3k}{(2\pi)^3}[\cdots]\times\delta^i_{\boldsymbol{x},\boldsymbol{k}} = \int_{v_i}\frac{\mathrm{d}^3x\,\mathrm{d}^3k}{(2\pi)^3}[\cdots]. \tag{13}$$

[1] In fact, all partial derivatives used in connection with the Fisher matrix should be replaced by functional derivatives in the continuum limit. It is only when we use bins (in real space and/or Fourier space) that these functional derivatives are converted to partial derivatives. Nevertheless, in order to keep the notation as simple as possible, we employ the same notation for both.

Moreover, for non-overlapping bins i and j it follows that:

$$\int \frac{\mathrm{d}^3x\,\mathrm{d}^3k}{(2\pi)^3}[\cdots]\times\delta^i_{\boldsymbol{x},\boldsymbol{k}}\times\delta^j_{\boldsymbol{x},\boldsymbol{k}} = \delta_{ij}\int_{v_i}\frac{\mathrm{d}^3x\,\mathrm{d}^3k}{(2\pi)^3}[\cdots]. \tag{14}$$

With these identities in mind, it is trivial to see that when using P^i_α as parameters, equation (7) reduces to

$$F^{ij}_\alpha = F[P^i_\alpha, P^j_\alpha] = \frac{\delta_{ij}}{(P^i_\alpha)^2}\int_{v_i}\frac{\mathrm{d}^3x\,\mathrm{d}^3k}{(2\pi)^3}\,\mathcal{F}_\alpha. \tag{15}$$

A more familiar form for this equation follows if we revert to the definition of averages over real- and Fourier-space bins:

$$F^{ij}_\alpha = \frac{\delta_{ij}\,\tilde{V}_i}{(P^i_\alpha)^2}\int_{V_i}\mathrm{d}^3x\;\langle\mathcal{F}_\alpha\rangle^{k_i}, \tag{16}$$

where the average is only over the Fourier-space bin (the bandpower). Up to a factor of 2, the integral over position space in the equation above defines the usual *effective volume* (Tegmark 1997; Tegmark et al. 1998). The uncertainty in the amplitude of the power spectrum at the bin i is therefore given by the covariance $\mathrm{Cov}[P^i_\alpha, P^j_\alpha] = (F^{ij}_\alpha)^{-1}$, which is diagonal in the Fourier modes – see, however, Abramo (2012). The relative uncertainty in the bandpowers of the power spectrum is then given by the well-known expression:

$$\left[\frac{\sigma(P^i_\alpha)}{P^i_\alpha}\right]^2 = \frac{1}{V_i\,\tilde{V}_i\,\mathcal{F}^i}. \tag{17}$$

When the number density of the tracer is very high, $\mathcal{P}_\alpha \gg 1$ and $\mathcal{F}_\alpha \to 1/2$, and if that is the case, then the survey is dominated by cosmic variance, $\sigma(P^i_\alpha)/P^i_\alpha \to \sqrt{2/v_i}$. The phase space volume v_i expresses the number of modes of the bandpower that fit in the physical volume V_i, and the factor of 2 comes from the fact that the density contrast is real.

2.3 Multitracer Fisher matrix

Galaxy surveys can detect a wide variety of objects: galaxies of different types, quasars, Ly α emmitters, Ly α absorbers, etc. In the future, all these data will coalesce into multilayer maps of the observable Universe, containing many different kinds of objects which can be regarded as tracers of the large-scale structure.

The multitracer Fisher information matrix describes how the contributions of cosmic variance and shot noise affect the signal-to-noise ratio (SNR) of the observables we are trying to measure – namely, the clustering properties of the tracers. While the nature of shot noise remains basically the same in the presence of multiple tracers, the effect of cosmic variance, which is shared among all tracers, mixes the different components.

The first authors to write a multitracer Fisher matrix (or, equivalently, a covariance matrix for the power spectra) were White, Song & Percival (2008), McDonald & Seljak (2009) and Hamaus et al. (2012). In Abramo (2012), we derived the multitracer Fisher directly from the covariance of the counts of the tracers (the 'pixel covariance'). The basic difference between the approaches of White et al. (2008) and McDonald & Seljak (2009) and ours is that those authors regard the cross-power spectra as independent parameters, while we implicitly assume that, for the purposes of estimating the power spectra from the data, the cross-spectra are determined by the auto-spectra – see, however, Swanson et al. (2008) and Bonoli & Pen (2008) for situations where this may not be true. The Fisher

matrix computed in equation (21) of Hamaus et al. (2012) also reduces to ours, if the cross-correlations are unaffected by shot noise – see also Smith (2009), Smith & Marian (2014) and Smith & Marian (2015).

We now show how to obtain the multitracer Fisher matrix from first principles. The generalization of equation (2) in the present context is

$$F^{ij}_{\mu\nu} = \frac{1}{2}\mathrm{Tr}\left[C^{-1}\frac{\partial C}{\partial P^i_\mu}C^{-1}\frac{\partial C}{\partial P^j_\nu}\right] = \frac{1}{2}\sum_{\alpha\beta\gamma\sigma}\int \mathrm{d}^3x\,\mathrm{d}^3y\,\mathrm{d}^3y'\,\mathrm{d}^3x' \times C^{-1}_{\alpha\beta}(\boldsymbol{x},\boldsymbol{y})\frac{\partial C_{\beta\gamma}(\boldsymbol{y},\boldsymbol{y}')}{\partial P^i_\mu}C^{-1}_{\gamma\sigma}(\boldsymbol{y}',\boldsymbol{x}')\frac{\partial C_{\sigma\alpha}(\boldsymbol{x}',\boldsymbol{x})}{\partial P^j_\nu}. \quad (18)$$

There are two main differences between equations (18) and (2): first, in equation (2) the data were assumed to be discretized, while here this is not the case anymore – the positions $\boldsymbol{x}$ of galaxies form a continuum, so the sums of equation (2) become the integrals of equation (18). Second, when we have N different types of galaxies (the tracers), there are $N(N+1)/2$ correlation functions, and we must sum over all those as well. Notice also that, in this notation, greek indices are always related to their respective spatial positions, to wit $\alpha \to \boldsymbol{x}$, $\beta \to \boldsymbol{y}$,$\gamma \to \boldsymbol{y}'$ and $\sigma \to \boldsymbol{x}'$. Hence, when going from equation (2) to equation (18) we replaced $\sum_x \to \sum_\alpha \int \mathrm{d}^3x$, which is the sum over all galaxy types α, and over all possible positions $\boldsymbol{x}$, and the same for $\sum_y \to \sum_\beta \int \mathrm{d}^3y$, etc. This expression reinforces a key aspect of the multitracer Fisher matrix: it couples the covariances of all the tracers involved – which is exactly the behaviour we aimed for.

In order to simplify the Fisher matrix of equation (18) we start by expressing the covariance of tracer counts as

$$C_{\alpha\beta}(\boldsymbol{x},\boldsymbol{x}') = \xi_{\alpha\beta}(\boldsymbol{x},\boldsymbol{x}') + \frac{\delta_{\alpha\beta}}{\bar{n}_\alpha(\boldsymbol{x})}\delta_\mathrm{D}(\boldsymbol{x}-\boldsymbol{x}') = \int \frac{\mathrm{d}^3k}{(2\pi)^3}\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{x}')} \times \left[B_\alpha(\boldsymbol{x},\boldsymbol{k})P_\mathrm{m}(\bar{x},\boldsymbol{k})B_\beta(\boldsymbol{x}',\boldsymbol{k}) + \frac{\delta_{\alpha\beta}}{\bar{n}_\alpha(\boldsymbol{x})}\right], \quad (19)$$

where $\bar{x}$ denotes the mean position (or redshift) in which the matter power spectrum is evaluated. A key difficulty with the covariance of counts is that, in any realistic situation, it cannot be inverted. However, if the effective biases and the power spectrum depend weakly on $\boldsymbol{k}$, then it is a fair approximation to integrate the complex exponential in equation (19) into a Dirac delta function, and to pull the rest of the integrand outside of the integral (Hamilton 2005a,b). This implies taking the approximation that

$$C_{\alpha\beta}(\boldsymbol{x},\boldsymbol{x}') \to \delta_\mathrm{D}(\boldsymbol{x}-\boldsymbol{x}') \times \left[\frac{\delta_{\alpha\beta}}{\bar{n}_\alpha} + B_\alpha\,P_\mathrm{m}B_\beta\right]. \quad (20)$$

This expression can now be easily inverted, as we will show next.

The inverse of the covariance should obey the property:

$$\sum_\beta \int \mathrm{d}^3y\, C^{-1}_{\alpha\beta}(\boldsymbol{x},\boldsymbol{y})\,C_{\beta\gamma}(\boldsymbol{y},\boldsymbol{x}') = \sum_\beta \int \mathrm{d}^3y\, C_{\alpha\beta}(\boldsymbol{x},\boldsymbol{y})\,C^{-1}_{\beta\gamma}(\boldsymbol{y},\boldsymbol{x}') = \delta_{\alpha\gamma}\,\delta_\mathrm{D}(\boldsymbol{x}-\boldsymbol{x}'). \quad (21)$$

Since $\int \mathrm{d}^3y\,\delta_\mathrm{D}(\boldsymbol{x}-\boldsymbol{y})\delta_\mathrm{D}(\boldsymbol{y}-\boldsymbol{x}') = \delta_\mathrm{D}(\boldsymbol{x}-\boldsymbol{x}')$, all we have to do is to invert the matrix inside the square brackets in equation (20).

But matrices of the type $M_{\alpha\beta} = \delta_{\alpha\beta} + v_\alpha v_\beta$ can be easily inverted, in fact $M^{-1}_{\alpha\beta} = \delta_{\alpha\beta} - v_\alpha v_\beta/(1+v^2)$, where $v^2 = \sum_\gamma v^2_\gamma$. A simple generalization of this simple case leads immediately to

$$C^{-1}_{\alpha\beta}(\boldsymbol{x},\boldsymbol{x}') \to \delta_\mathrm{D}(\boldsymbol{x}-\boldsymbol{x}') \times \left[\delta_{\alpha\beta}\,\bar{n}_\alpha - \bar{n}_\alpha\frac{B_\alpha\,P_\mathrm{m}B_\beta}{1+\mathcal{P}}\bar{n}_\beta\right], \quad (22)$$

where we define the *total clustering strength*, $\mathcal{P}$, as the sum of the *clustering strengths of the individual tracers*, $\mathcal{P}_\mu = \bar{n}_\mu\,B^2_\mu\,P_\mathrm{m}$:

$$\mathcal{P}(\boldsymbol{x},\boldsymbol{k}) = \sum_\mu \bar{n}_\mu(\boldsymbol{x})\,B^2_\mu(\boldsymbol{x},\boldsymbol{k})\,P_\mathrm{m}(\boldsymbol{x},\boldsymbol{k}) = \sum_\mu \mathcal{P}_\mu(\boldsymbol{x},\boldsymbol{k}). \quad (23)$$

The clustering strength of a tracer is simply the power spectrum of that tracer, expressed in units of the (Poissonian) shot noise of that same tracer.

The problem with equations (20) and (22) is that they refer to a scale $\boldsymbol{k}$ which does not exist in the original expression, equation (19). In fact, equations (20)–(22) treat the positions of the two-point function, $\boldsymbol{x}$ and $\boldsymbol{x}'$ as one and the same, due to the Dirac delta-function. Hence, one should think of the Fourier mode $\boldsymbol{k}$ which is implicit in equations (20)–(22), as the reciprocal of some typical physical distance between $\boldsymbol{x}$ and $\boldsymbol{x}'$, and in that sense, its role is to limit the scope of that distance in expressions involving these approximations. Notice that this issue appears already in the FKP and PVP methods.

Coming back to equation (18), we see that the last step before constructing the Fisher matrix is the computation of the term $\partial C_{\alpha\beta}(\boldsymbol{x},\boldsymbol{x}')/\partial P^i_\mu$. Once again, it is useful to employ the notion of functional derivatives and the results of the previous section. Using the second line of equation (19) and the fact that $\partial P_\alpha(\boldsymbol{x},\boldsymbol{k})/\partial P^i_\mu = \delta^i_{\boldsymbol{x},\boldsymbol{k}}$, we find, after some rearrangement, that:

$$\frac{\partial C_{\alpha\beta}(\boldsymbol{x},\boldsymbol{x}')}{\partial P^i_\mu} = \int \frac{\mathrm{d}^3k}{(2\pi)^3}\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{x}')}\left(\delta_{\alpha\mu}\delta^i_{\boldsymbol{x},\boldsymbol{k}} + \delta_{\beta\mu}\delta^i_{\boldsymbol{x}',\boldsymbol{k}}\right) \times \frac{B_\alpha(\boldsymbol{x},\boldsymbol{k})B_\beta(\boldsymbol{x}',\boldsymbol{k})}{2\,(B^i_\mu)^2}, \quad (24)$$

where B^i_μ is the effective bias of the species μ at the bin i. Equation (24) should be regarded as an *operator*: when it acts on functions of $\boldsymbol{k}$, it integrates the Fourier modes inside the volume of the bin i. Apart from a volume factor $\tilde{V}_i$ this integration is just an average over the bin $\boldsymbol{k}_i$.

We can now obtain the Fisher matrix by substituting equations (22) and (24) into equation (18). The result, after a bit of algebra, is that:

$$F^{ij}_{\mu\nu} = \frac{\delta_{ij}}{P^i_\mu P^i_\nu}\int_{v_i}\frac{\mathrm{d}^3x\,\mathrm{d}^3k}{(2\pi)^3}\mathcal{F}_{\mu\nu}, \quad (25)$$

where

$$\mathcal{F}_{\mu\nu}(\boldsymbol{x},\boldsymbol{k}) = \frac{1}{4}\frac{\delta_{\mu\nu}\mathcal{P}_\mu\mathcal{P}(1+\mathcal{P}) + \mathcal{P}_\mu\mathcal{P}_\nu(1-\mathcal{P})}{(1+\mathcal{P})^2}. \quad (26)$$

Equation (26) is in fact the Fisher information density per unit of phase space volume for $\log\mathcal{P}_\mu$ (Abramo 2012):

$$F[\log\mathcal{P}_\mu(\boldsymbol{x},\boldsymbol{k}),\log\mathcal{P}_\nu(\boldsymbol{x}',\boldsymbol{k}')] = (2\pi)^3\delta_\mathrm{D}(\boldsymbol{x}-\boldsymbol{x}')\delta_\mathrm{D}(\boldsymbol{k}-\boldsymbol{k}')\mathcal{F}_{\mu\nu}(\boldsymbol{x},\boldsymbol{k}), \quad (27)$$

or, equivalently, in bins of finite volume:

$$F[\log\mathcal{P}^i_\mu,\log\mathcal{P}^j_\nu] = \delta_{ij}\,\mathcal{F}^i_{\mu\nu}. \quad (28)$$

One can easily check that the multitracer Fisher matrix of equation (25) reduces to the FKP Fisher matrix, equation (16), when there is only one type of tracer.

3 THE OPTIMAL MULTITRACER QUADRATIC ESTIMATORS

Starting from the Fisher matrix of equation (25), and with the help of equations (22) and (24), we are in a position to implement the construction of the estimators which was presented in Section 2.1. For now we will not discuss the role of random maps, which help to subtract spurious fluctuations that could be generated by modulations on the mean number of tracers, $\bar{n}_\mu$. The calculations below are exactly the same with or without the random maps, so we come back to this issue at the end of this section, after we have shown how to construct the multitracer estimators.

Since our data are the density contrasts of the tracers, and the points in space form a continuum, the proper generalization of the quadratic form in equation (3) is

$$\hat{Q}^i_\mu = \mathrm{Tr}\left[\delta\, E^i_\mu\, \delta\right] - \Delta Q^i_\mu = \sum_{\alpha\beta} \int \mathrm{d}^3x\, \mathrm{d}^3y\, \delta_\alpha(\boldsymbol{x})\, \left[E_{\alpha\beta}(\boldsymbol{x},\boldsymbol{y})\right]^i_\mu\, \delta_\beta(\boldsymbol{y}) - \Delta Q^i_\mu, \tag{29}$$

where ΔQ^i_μ is the *bias of the estimator*. According to the appropriate generalization of equation (4):

$$\left[E_{\alpha\beta}(\boldsymbol{x},\boldsymbol{y})\right]^i_\mu = \frac{1}{2}\sum_{\sigma\gamma}\int \mathrm{d}^3x'\,\mathrm{d}^3y' \times C^{-1}_{\alpha\sigma}(\boldsymbol{x},\boldsymbol{x}')\frac{\partial C_{\sigma\gamma}(\boldsymbol{x}',\boldsymbol{y}')}{\partial P^i_\mu} C^{-1}_{\gamma\beta}(\boldsymbol{y}',\boldsymbol{y}). \tag{30}$$

This matrix, when used in equation (29), combines the data from galaxies of type α at positions $\boldsymbol{x}$ with the data from galaxies of type β at positions $\boldsymbol{y}$, generating an estimator for the power spectrum of galaxies of type μ at some bandpower i.

Inserting equations (22) and (24) into equation (30), and then back on equation (29), leads to the following expression for the quadratic form:

$$\hat{Q}^i_\mu = \frac{1}{4\,(B^i_\mu)^2}\sum_{\sigma\gamma}\int \mathrm{d}^3x\,\mathrm{d}^3x'\int\frac{\mathrm{d}^3k}{(2\pi)^3}\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{x}')} \times f_\sigma(\boldsymbol{x},\boldsymbol{k})\left(\delta_{\sigma\mu}\delta^i_{\boldsymbol{x},\boldsymbol{k}} + \delta_{\gamma\mu}\delta^i_{\boldsymbol{x}',\boldsymbol{k}}\right) f_\gamma(\boldsymbol{x}',\boldsymbol{k}) - \Delta Q_{\mu,i}, \tag{31}$$

where

$$f_\sigma(\boldsymbol{x},\boldsymbol{k}) = \sum_\alpha w_{\sigma\alpha}(\boldsymbol{x},\boldsymbol{k})\,\delta_\alpha(\boldsymbol{x}) \tag{32}$$

are *weighted density contrasts* for the tracers. The weights are

$$w_{\sigma\alpha}(\boldsymbol{x},\boldsymbol{k}) = \left[\delta_{\sigma\alpha} - \frac{\mathcal{P}_\sigma(\boldsymbol{x},\boldsymbol{k})}{1+\mathcal{P}(\boldsymbol{x},\boldsymbol{k})}\right]\bar{n}_\alpha B_\alpha(\boldsymbol{x},\boldsymbol{k}). \tag{33}$$

These weights are the generalization of the FKP weights (Feldman et al. 1994) for the case of multiple tracers of large-scale structure. As we will prove in a moment, equation (33) defines the optimal weights for maps containing an arbitrary number of different types of tracers. In the case of a single species of tracer, the weights for the density contrast reduce to $w = \bar{n}B/(1+\bar{n}B^2P_\mathrm{m})$, which is precisely the FKP weight for the density contrast – except for a normalization, whose origin and purpose will become clearer soon.

Returning to equation (31), notice that the Kronecker delta functions are accompanied by their respective restrictions over the phase space volume of the bin where we are estimating the quantities of interest (in our case, the P^i_μ), hence:

$$\hat{Q}^i_\mu = \frac{1}{4\,(B^i_\mu)^2}\int_{\tilde{V}_i}\frac{\mathrm{d}^3k}{(2\pi)^3} \times \int_{V_i}\mathrm{d}^3x\;\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\,f_\mu(\boldsymbol{x})\int\mathrm{d}^3x'\,\mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}'}\,f(\boldsymbol{x}') + \mathrm{c.c.} - \Delta Q^i_\mu, \tag{34}$$

where *c.c.* stands for complex conjugate, and

$$f = \sum_\sigma f_\sigma = \frac{1}{1+\mathcal{P}}\sum_\sigma \bar{n}_\sigma B_\sigma \delta_\sigma. \tag{35}$$

Formally, the spatial integral over f_μ covers only the volume of the redshift slice V_i, while the spatial integral over f should be performed over the whole volume of the survey. Although this is a subtlety which is present already in the Fourier analysis of FKP, in practice we are always considering separately the data on different redshift slices, and then all integrations in real space are performed over the same volume. Nevertheless, a more rigorous treatment comparing different redshift slices would dictate that one of the Fourier integrations be performed over the whole available volume of the survey, while the other integration would be carried out over the volume of one particular redshift slice.

In what follows we will ignore this subtlety, and will consider that both integrations over spatial volume result in the Fourier transforms $\tilde{f}_\mu$ and $\tilde{f}$. We then obtain that

$$\hat{Q}^i_\mu = \frac{1}{4\,(B^i_\mu)^2}\int_{\tilde{V}_i}\frac{\mathrm{d}^3k}{(2\pi)^3}\left[\tilde{f}_\mu(\boldsymbol{k})\,\tilde{f}^*(\boldsymbol{k}) + \mathrm{c.c.}\right] - \Delta Q^i_\mu. \tag{36}$$

But the integration above is, up to the volume factor, simply the average over the Fourier bin, hence:

$$\hat{Q}^i_\mu = \frac{\tilde{V}_i}{4\,(B^i_\mu)^2}\left\langle \tilde{f}_\mu\,\tilde{f}^* + \mathrm{c.c.}\right\rangle^{k_i} - \Delta Q^i_\mu. \tag{37}$$

Notice that, in this expression and others like it, the factor of $(B^i_\mu)^{-2}$ is the *fiducial value* of the effective bias, whereas the weighted density contrasts f_μ must be computed directly from the data. However, since the weights of equation (33) are themselves also computed using the fiducial values of B_μ and P_m, the weighted fields f_μ are a combination of both theory and data. The situation is not different from the usual case of Fourier analysis of cosmological surveys employing the FKP or the PVP estimators. Evidently, these quadratic estimators are only truly optimal if the parameters take their fiducial values.

Starting either from equation (37), or more directly from equation (29), a long but straightforward calculation shows that the covariance of this quadratic form in fact results in $\mathrm{Cov}(\hat{Q}^i_\mu, \hat{Q}^j_\nu) = F^{ij}_{\mu\nu}$, where the Fisher matrix was given in equation (25).

Finally, we can construct the optimal quadratic estimators for the power spectra of any tracer species, by plugging the quadratic form above into the appropriate generalization of equation (5). The Fisher matrix that is relevant in this case was already given in equation (25). We have, therefore, that the optimal quadratic estimators, whose covariances are given by the inverse of the Fisher matrix, are given by

$$\hat{P}^i_\mu = \sum_\nu\sum_j \left[F^{ij}_{\mu\nu}\right]^{-1}\hat{Q}^j_\nu = \sum_\nu \left[F^{ii}_{\mu\nu}\right]^{-1}\hat{Q}^i_\nu, \tag{38}$$

where the second line follows from the fact that the Fisher matrix is diagonal in the bins (i, j).[2]

Now the origin of the normalizations of the weights, which appear both in the FKP and the PVP formulas, becomes clear: up to the prefactor in equation (37), those normalizations correspond to the inverse of the Fisher matrix in equation (38).

3.1 Subtracting the bias of the estimators

Although by definition $\mathrm{Cov}(\hat{P}^i_\mu, \hat{P}^j_\nu) = [F^{ij}_{\mu\nu}]^{-1}$, we still must ensure that the bias of estimator is properly subtracted. According to equation (6), those biases are

$$\Delta Q^i_\mu = \frac{1}{2}\sum_{\alpha\beta}\int \mathrm{d}^3x\,\mathrm{d}^3x'\,\frac{\partial C_{\alpha\beta}(\boldsymbol{x},\boldsymbol{x}')}{\partial P^i_\mu}\,C^{-1}_{\beta\alpha}(\boldsymbol{x}',\boldsymbol{x}) - \sum_\nu\sum_j F^{ij}_{\mu\nu}\,P^j_\nu. \tag{39}$$

This expression can be easily worked out, and the result is

$$\Delta Q^i_\mu = \frac{1}{2\,(B^i_\mu)^2}\int_{v_i}\frac{\mathrm{d}^3x\,\mathrm{d}^3k}{(2\pi)^3}\,\frac{\bar{n}_\mu B^2_\mu}{1+\mathcal{P}} - \sum_\nu F^{ii}_{\mu\nu}\,P^i_\nu. \tag{40}$$

It is also useful to compute the bias corrections for the power spectrum estimators, $\hat{P}^i_\mu$. For this calculation, we will employ the approximation that averages over the bins can be manipulated in such a way that $\langle A\,B\rangle^i \simeq \langle A\rangle^i\,\langle B\rangle^i$. This amounts to assuming that the bins are small compared with the coherence scale of the quantities of interest – e.g. if the power spectrum is almost constant inside a bin $\boldsymbol{k}_i$, then $\langle P(\boldsymbol{k})/[1+\bar{n}\,P(\boldsymbol{k})]\rangle^{\boldsymbol{k}_i} \simeq P(\boldsymbol{k}_i)/[1+\bar{n}\,P(\boldsymbol{k}_i)]$.

In order to go from $\hat{Q}^i_\mu$ to $\hat{P}^i_\mu$ we must first find the inverse of the Fisher matrix which was found in equation (25). But that is basically the inverse of the Fisher matrix for the $\log\mathcal{P}_\mu$ which was found in equation (26). This is a particular case of the same type of matrix inversion which we used in the case of the pixel covariance, and the result is that:

$$F^{-1}_{\mu\nu} = P_\mu\,P_\nu\,\mathcal{F}^{-1}_{\mu\nu}, \tag{41}$$

where

$$\mathcal{F}^{-1}_{\mu\nu} = \frac{4(1+\mathcal{P})}{\mathcal{P}}\left(\frac{\delta_{\mu\nu}}{\mathcal{P}_\mu} + \frac{\mathcal{P}-1}{2\mathcal{P}}\right). \tag{42}$$

Using this expression, and the approximation that bin averages can be freely rearranged, we obtain that the estimators of the power spectra of the tracers reduce to

$$\left\langle\left\langle\hat{P}^i_\mu\right\rangle\right\rangle \to \left\langle\left(1+\frac{1}{\mathcal{P}}\right)P_\mu\right\rangle^i - \Delta P^i_\mu. \tag{43}$$

In fact, one can also show directly from equation (40) that the bias of the estimator is

$$\Delta P^i_\mu \equiv \sum_\nu [F^{ii}_{\mu\nu}]^{-1}\Delta Q^i_\nu \to \left\langle\frac{P_\mu}{\mathcal{P}}\right\rangle^i, \tag{44}$$

which ensures that the expectation value of the estimator converges to the true (or, in this context, the fiducial) value, $\langle\langle\hat{P}^i_\mu\rangle\rangle \to P^i_\mu$. In the case of a single type of tracer, the bias of the estimator reduces to the (Poissonian) shot noise, $1/\bar{n}$.

[2] This is only true if the Fourier-space bins are sufficiently large, $\Delta k_i \gtrsim \pi/V^{1/3}$ – see, e.g. Abramo (2012).

3.2 The window functions

The expectation values of the power spectra obtained through the multitracer quadratic estimators are convolutions of the true power spectra with some window functions. These window functions can be obtained directly from the expectation value of the expression in equation (31), by taking $\delta_\alpha \to B_\alpha\,\delta_\mathrm{m}$ and neglecting the biases of the estimators:

$$\langle\langle\hat{Q}^i_\mu\rangle\rangle = \frac{1}{4\,(B^i_\mu)^2}\sum_{\sigma\alpha\beta}\int_{v_i}\frac{\mathrm{d}^3x\,\mathrm{d}^3k}{(2\pi)^3}\int\mathrm{d}^3x'\,\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{x}')} \times w_{\mu\alpha}(\boldsymbol{x},\boldsymbol{k})B_\alpha(\boldsymbol{x},\boldsymbol{k})\,w_{\sigma\beta}(\boldsymbol{x}',\boldsymbol{k})B_\beta(\boldsymbol{x}',\boldsymbol{k}) \times\left\langle\left\langle\delta_\mathrm{m}(\boldsymbol{x})\delta_\mathrm{m}(\boldsymbol{x}')\right\rangle\right\rangle + \mathrm{c.c.} \tag{45}$$

From the definition of the weight functions, equation (33), it is easy to derive that $\sum_\sigma w_{\sigma\alpha}B_\alpha = \bar{n}_\alpha B^2_\alpha/(1+\mathcal{P}) = \mathcal{P}_\alpha/P_\mathrm{m}(1+\mathcal{P})$, and expressing the matter two-point correlation function in terms of the matter power spectrum, we obtain:[3]

$$\langle\langle\hat{Q}^i_\mu\rangle\rangle = \frac{1}{4\,(B^i_\mu)^2}\int_{v_i}\frac{\mathrm{d}^3x\,\mathrm{d}^3k}{(2\pi)^3}\int\frac{\mathrm{d}^3x'\,\mathrm{d}^3k'}{(2\pi)^3}\,P_\mathrm{m}(\boldsymbol{k}') \times\mathrm{e}^{\mathrm{i}(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{x}}G_\mu(\boldsymbol{x},\boldsymbol{k})\,\mathrm{e}^{-\mathrm{i}(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{x}'}G(\boldsymbol{x}',\boldsymbol{k}) + \mathrm{c.c.}, \tag{46}$$

where

$$G_\mu(\boldsymbol{x},\boldsymbol{k}) = \frac{1}{P_\mathrm{m}}\frac{\mathcal{P}_\mu}{1+\mathcal{P}}, \tag{47}$$

and $G = \sum_\mu G_\mu = \mathcal{P}/P_\mathrm{m}(1+\mathcal{P})$. Notice that the integration over $\boldsymbol{k}'$ in equation (46) is *not* limited to the volume of the bandpower i – see also equation (54).

Hence, we define the window function:

$$W^{i(Q)}_\mu(\boldsymbol{k}') = \frac{1}{4(B^i_\mu)^2}\int_{\tilde{V}_i}\frac{\mathrm{d}^3k}{(2\pi)^3}\,\tilde{G}_\mu(\boldsymbol{k},\boldsymbol{k}')\,\tilde{G}^*(\boldsymbol{k},\boldsymbol{k}') + \mathrm{c.c.}, \tag{48}$$

where the Fourier transform of the kernels of equation (47) are

$$\tilde{G}_\mu(\boldsymbol{k},\boldsymbol{k}') = \int\mathrm{d}^3x\,\mathrm{e}^{\mathrm{i}(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{x}}\,G_\mu(\boldsymbol{x},\boldsymbol{k}), \tag{49}$$

$$\tilde{G}(\boldsymbol{k},\boldsymbol{k}') = \sum_\mu\tilde{G}_\mu(\boldsymbol{k},\boldsymbol{k}'). \tag{50}$$

Because the integral over d^3k is performed only over the Fourier bin $\tilde{V}_i$, it is often an accurate approximation to take $\boldsymbol{k}\to\boldsymbol{k}_i$ in the argument of the kernels of equations (49) and (50), and replace:

$$\tilde{G}_\mu(\boldsymbol{k},\boldsymbol{k}') \to \tilde{G}^i_\mu(\boldsymbol{k}-\boldsymbol{k}') = \int\mathrm{d}^3x\,\mathrm{e}^{\mathrm{i}(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{x}}G_\mu(\boldsymbol{x},\boldsymbol{k}_i), \tag{51}$$

$$\tilde{G}(\boldsymbol{k},\boldsymbol{k}') \to \tilde{G}^i(\boldsymbol{k}-\boldsymbol{k}') = \sum_\mu\tilde{G}^i_\mu(\boldsymbol{k}-\boldsymbol{k}'). \tag{52}$$

[3] In fact, a careful analysis shows that only one of the real-space integrals in equation (46) ought to be carried out only over the volume of the particular redshift slice under study, V_i, while the other should be in principle carried out over the whole volume of the survey (e.g. all redshift slices). In practice, it may be more conservative to treat each bin in position space as an entirely independent survey, and in that case the two integrals would be carried out over the volume of the redshift slice. In fact, it is only in this limit that the Fisher matrix of equation (25), or that of equation (28), are truly diagonal in the bins i and j (Abramo 2012), and therefore it is only in this sense that the optimal estimators satisfy the constraint that $\mathrm{Cov}(\hat{P}^i_\mu, \hat{P}^j_\nu) \to [F^{ij}_{\mu\nu}]^{-1}$.

Notice that the Fourier transform of the kernels with respect to their spatial dependence still remains, since we do not replace $\boldsymbol{k}$ by $\boldsymbol{k}_i$ in the exponentials.

With these definitions, we can write the effective window function of the quadratic form as

$$W_\mu^{i(Q)} = \frac{1}{4\,(B_\mu^i)^2}\int_{\tilde{V}_i}\frac{\mathrm{d}^3k}{(2\pi)^3}\left[\tilde{G}_\mu\tilde{G}^* + \text{c.c.}\right]$$
$$\simeq \frac{\tilde{V}_i}{4\,(B_\mu^i)^2}\left\langle\tilde{G}_\mu^i\tilde{G}^{i*}\right\rangle^i + \text{c.c.} \quad (53)$$

Hence, in terms of the window function we have

$$\langle\langle\hat{Q}_\mu^i\rangle\rangle \simeq \int\frac{\mathrm{d}^3k'}{(2\pi)^3}\,P_\mathrm{m}(\boldsymbol{k}')\,W_\mu^{i(Q)}(\boldsymbol{k}_i,\boldsymbol{k}'). \quad (54)$$

An interesting limiting case happens when we take all quantities to be constant inside the spatial bin, $\mathcal{P}_\mu(\boldsymbol{x},\boldsymbol{k}) \to \mathcal{P}_\mu(\boldsymbol{k})$, and then take the continuum limit. In that case the kernels in Fourier space become Dirac-delta functions, $\tilde{G}_\mu \to G_\mu^i\,(2\pi)^3\delta_\mathrm{D}(\boldsymbol{k}-\boldsymbol{k}')$, and the window function becomes:

$$W_\mu^{i(Q)} \to \frac{1}{4\,(B_\mu^i\,P_\mathrm{m}^i)^2}\frac{\mathcal{P}_\mu^i\mathcal{P}^i}{(1+\mathcal{P}^i)^2}\times(2\pi)^3\delta_\mathrm{D}(\boldsymbol{k}_i-\boldsymbol{k}'). \quad (55)$$

The most relevant window functions are, of course, not $W_\mu^{i(Q)}$, but those which apply for the estimators of the power spectra. Since $\langle\langle\hat{P}_\mu^i\rangle\rangle = \sum_\nu[F_{\mu\nu}^{ii}]^{-1}\langle\langle\hat{Q}_\nu^i\rangle\rangle$, we obtain that:

$$\langle\langle\hat{P}_\mu^i\rangle\rangle = \int\frac{\mathrm{d}^3k'}{(2\pi)^3}\,P_\mathrm{m}(\boldsymbol{k}')\,W_\mu^i, \quad (56)$$

where

$$W_\mu^i = \sum_\nu[F_{\mu\nu}^{ii}]^{-1}\frac{1}{4\,(B_\nu^i)^2}\int_{\tilde{V}_i}\frac{\mathrm{d}^3k}{(2\pi)^3}\left[\tilde{G}_\nu\tilde{G}^* + \text{c.c.}\right]. \quad (57)$$

Finally, in the same limit that was used to obtain equation (55), we can apply the identity:

$$\sum_\nu F_{\mu\nu}^{-1}\times\frac{1}{B_\mu^2}\,\frac{1}{P_\mathrm{m}}\,\frac{\mathcal{P}_\nu}{1+\mathcal{P}} = 2B_\mu^2\,P_\mathrm{m}\frac{1+\mathcal{P}}{\mathcal{P}}, \quad (58)$$

to show that $W_\mu^i \to (B_\mu^i)^2\times(2\pi)^3\delta_\mathrm{D}(\boldsymbol{k}_i-\boldsymbol{k}')$, as in fact it ought to be. This completes the demonstration that the estimators derived in this section satisfy all the desired criteria for optimal, unbiased estimators, with the correct continuum limits.

3.3 Random maps and the integral constraints

Up to now we have introduced the optimal multitracer quadratic estimators without mentioning the role of the random ('synthetic') maps. They help subtract the fluctuations that arise purely as a result of modulations in the mean number density of the tracers, $\bar{n}_\mu(\boldsymbol{x})$, and are caused by, e.g. angular- or redshift-dependent variations in the selection function of a survey (Feldman et al. 1994).

For each tracer species with mean number density $\bar{n}_\mu(\boldsymbol{x})$ we define a random (white noise) map with a mean number density with the same shape as that which is presumed for the data: $\bar{n}_\mu^r(\boldsymbol{x}) = \bar{n}_\mu(\boldsymbol{x})/\alpha_\mu$, where α_μ are (small) constants. The random data sets have no structure, in the sense that their pixel covariances are just given by the shot noises of each sample:

$$\langle\langle\delta_\mu^r(\boldsymbol{x})\delta_\nu^r(\boldsymbol{x}')\rangle\rangle = \frac{\delta_{\mu\nu}}{\bar{n}_\mu^r}\,\delta_\mathrm{D}(\boldsymbol{x}-\boldsymbol{x}') = \alpha_\mu\frac{\delta_{\mu\nu}}{\bar{n}_\mu}\,\delta_\mathrm{D}(\boldsymbol{x}-\boldsymbol{x}').$$

With the data and random sets we construct weighted density contrasts in a way similar to the definition of equation (32):

$$f_\mu(\boldsymbol{x},\boldsymbol{k}) = \sum_\nu w_{\mu\nu}(\boldsymbol{x},\boldsymbol{k})\,\frac{n_\nu(\boldsymbol{x}) - A_\nu\,n_\nu^r(\boldsymbol{x})}{\bar{n}_\nu},$$
$$= \sum_\nu w_{\mu\nu}(\boldsymbol{x},\boldsymbol{k})\,\left[\delta_\nu(\boldsymbol{x}) - \frac{A_\nu}{\alpha_\nu}\,\delta_\nu^r(\boldsymbol{x}) + 1 - \frac{A_\nu}{\alpha_\nu}\right], \quad (59)$$

where the weights were given in equation (33). The values of A_ν should be calibrated in such a way that the weighted fields f_μ have zero mean over the volume of the sample, thus ensuring the so-called *integral constraints*, $\langle\hat{P}_\mu(k=0)\rangle \to 0$ (Fisher et al. 1993). It is easy to check that the condition $\int\mathrm{d}^3x f_\mu = 0$ is satisfied by setting:

$$A_\mu = \sum_\nu R_{\mu\nu}^{-1}\,D_\nu, \quad (60)$$

where

$$D_\nu = \int\mathrm{d}^3x\sum_\sigma w_{\nu\sigma}(\boldsymbol{x},\boldsymbol{k})\,[1+\delta_\sigma(\boldsymbol{x})], \quad (61)$$

$$R_{\mu\nu} = \frac{1}{\alpha_\nu}\int\mathrm{d}^3x\;w_{\mu\nu}(\boldsymbol{x},\boldsymbol{k})\,\left[1+\delta_\nu^r(\boldsymbol{x})\right]. \quad (62)$$

Since D_ν and $R_{\mu\nu}$ are functions of $\boldsymbol{k}$, in principle the constants A_μ also depend on the wavenumber. In practice, we employ only a couple of putative values for P_m in all the weights, hence we compute A_μ only for those values.

Usually the mean density contrasts of the random catalogues are very close to zero, which means that $A_\mu \to \alpha_\mu$ to a very good approximation. Indeed, taking $\delta_\nu^r \to 0$ in equation (62) it follows that equation (60) can be recast as

$$\frac{A_\mu}{\alpha_\mu} \approx 1 + \sum_{\nu\sigma}\left[\int\mathrm{d}^3x\;w_{\mu\nu}(x)\right]^{-1}\int\mathrm{d}^3x'\;w_{\nu\sigma}(x')\delta_\sigma(x'). \quad (63)$$

The fractional difference between A_μ and α_μ is of the order of the average of the density contrast over the whole volume of the catalogue. This correction is negligible unless the galaxy catalogues are extremely sparse, hence it is often safe to take $A_\mu \to \alpha_\mu$. One can also improve this approximation by taking smaller values of α_μ, which makes equation (63) more accurate. However, if there are reasons (e.g. computational) to limit the size of the synthetic catalogues, such that α_μ cannot be too small, then A_μ may deviate from α_μ.

Using equation (59) instead of equation (32) in the estimators do not make much difference in our calculations, except for the biases of the estimators, which inherit the factors of A_μ and α_μ. Starting from equation (40) we obtain:

$$\Delta Q_\mu^i = \frac{1}{2\,(B_\mu^i)^2}\int_{v_i}\frac{\mathrm{d}^3x\,\mathrm{d}^3k}{(2\pi)^3}\,\frac{\bar{n}_\mu B_\mu^2}{(1+\mathcal{P})^2}$$
$$\times\left\{1 + \sum_\nu\frac{A_\nu^2}{\alpha_\nu}\left[\delta_{\mu\nu}(1+\mathcal{P}) - \mathcal{P}_\nu\right]\right\} + \delta Q_\mu^i, \quad (64)$$

where the extra term, δQ_μ^i, arises when $A_\mu \neq \alpha_\mu$, leading to the additional correction:

$$\delta Q_\mu^i = \frac{1}{2\,(B_\mu^i)^2}\int\mathrm{d}^3x\sum_\nu\left(\frac{A_\nu}{\alpha_\nu} - 1\right)w_{\mu\nu}$$
$$\times\sum_{\gamma\sigma}\left(\frac{A_\sigma}{\alpha_\sigma} - 1\right)w_{\gamma\sigma}. \quad (65)$$

This expression can be simplified with the help of the definitions $\Gamma_\mu = (A_\mu/\alpha_\mu - 1)\bar{n}_\mu B_\mu$, and $\Gamma = \sum_\mu \Gamma_\mu$, leading to

$$\delta Q^i_\mu = \frac{1}{2\,(B^i_\mu)^2} \int \mathrm{d}^3x \left[\Gamma_\mu - \frac{\Gamma\,\mathcal{P}_\mu}{1+\mathcal{P}}\right] \frac{\Gamma}{1+\mathcal{P}}. \tag{66}$$

As we discussed above, in most cases the tracers are sufficiently abundant to make $A_\mu \simeq \alpha_\mu$, so $\Gamma_\mu \to 0$, and the extra term of equation (66) can be neglected. The biases of the estimators are then given only by the first term of equation (64) – with the simplification that $A^2_\mu/\alpha_\mu \to \alpha_\mu$. If, in addition, we assume that the random maps are constructed such that the α_μ are all identical, $\alpha_\mu \to \alpha$, then the biases of the estimators become simply:

$$\Delta Q^i_\mu = \frac{1+\alpha}{2\,(B^i_\mu)^2} \int_{v_i} \frac{\mathrm{d}^3x\,\mathrm{d}^3k}{(2\pi)^3} \frac{\bar{n}_\mu B^2_\mu}{(1+\mathcal{P})^2}. \tag{67}$$

4 PROPERTIES AND RELATIONS OF THE MULTITRACER ESTIMATOR

The results of the previous section are closely related to other methods for the Fourier analysis of cosmological surveys, but they also extend their scope considerably.

The simplest limit is when we take all tracers to be a single species. In that case our formulas reduce to the ones by FKP. The weights of equation (33) reduce to $w = \bar{n}B/(1+\mathcal{P})$, which are the FKP weights after we make the identification $\mathcal{P} = \sum_\mu \mathcal{P}_\mu = \bar{n}B^2 P_\mathrm{m}$, where $\bar{n} = \sum_\mu \bar{n}_\mu$ and $B^2 = \bar{n}^{-1}\sum_\mu \bar{n}_\mu B^2_\mu$. Furthermore, the multitracer Fisher matrix of equation (26) also reduces to the FKP Fisher matrix once we sum over all the clustering strengths – i.e. when we combine all tracers into a single type. Changing variables in the Fisher matrix $\mathcal{F}_{\mu\nu} = F[\log\mathcal{P}_\mu, \log\mathcal{P}_\nu]$, from $\log\mathcal{P}_\mu$ to $\log\mathcal{P}$, introduces a constant Jacobian, $J_\mu = \mathbb{1}_\mu$. This can be seen by considering the inverse Jacobian, $J^{-1}_\mu = \partial\log\mathcal{P}/\partial\log\mathcal{P}_\mu = \mathcal{P}_\mu/\mathcal{P}$, which satisfies $\sum_\mu J^{-1}_\mu J_\mu = \sum_\mu J^{-1}_\mu = \sum_\mu \mathcal{P}_\mu/\mathcal{P} = 1$. Hence the multitracer Fisher matrix projected into the Fisher matrix for the total clustering strength becomes:

$$\begin{aligned} F[\log\mathcal{P}] &= \sum_{\mu\nu} J_\mu\,\mathcal{F}_{\mu\nu}\,J_\nu = \sum_{\mu\nu}\mathcal{F}_{\mu\nu} \\ &= \frac{1}{2}\left(\frac{\mathcal{P}}{1+\mathcal{P}}\right)^2, \end{aligned} \tag{68}$$

which is the FKP Fisher information density per unit of phase space volume.

We now discuss some of the main features of the multitracer technique, as well as its relations to other methods in the literature.

4.1 The PVP estimator

Suppose we fix all parameters B^i_μ, and try to estimate the matter power spectrum $P_\mathrm{m}(k)$ using data from all tracers. The optimal, unbiased estimator in that case was derived by Percival et al. (2004, PVP) – see also Smith & Marian (2015). The method used by PVP to construct their estimator was the same as that used by FKP – i.e. the weights which minimize the covariance $\mathrm{Cov}(P^i_\mathrm{m}, P^j_\mathrm{m})$ were obtained through a variational approach.

Here, instead, we built the optimal estimators directly on the basis of the pixel covariance, assuming Gaussianity of the data. We already showed that our estimators reduce to that of FKP in the case of a single species of tracer. Now we show that the PVP estimator is just one of many possible projections of the multitracer estimators.

If we fix the effective biases B_μ to their fiducial values (i.e. if the bias of each type of galaxy and the shape of the RSDs are set to their true values), then the remaining unknown is the matter power spectrum at the position- and Fourier-space bins, $P_{\mathrm{m},\,i}$. We may now ask what is the Fisher matrix for the matter power spectrum. This is easily derived from equation (25) through the change of variable:

$$\begin{aligned} F(P^i_\mathrm{m}, P^j_\mathrm{m}) &= \sum_{\mu\nu}\sum_{kl} \frac{\partial P^k_\mu}{\partial P^i_\mathrm{m}}\,F^{kl}_{\mu\nu}\,\frac{\partial P^l_\mu}{\partial P^j_\mathrm{m}} \\ &= \delta_{ij}\sum_{\mu\nu}(B^i_\mu\,B^j_\nu)^2\,F^{ij}_{\mu\nu} \\ &= \frac{\delta_{ij}}{(P^i_\mathrm{m})^2}\sum_{\mu\nu}\mathcal{F}^{ii}_{\mu\nu} \\ &= \frac{\delta_{ij}}{(P^i_\mathrm{m})^2}\int_{v_i}\frac{\mathrm{d}^3x\,\mathrm{d}^3k}{(2\pi)^3}\,\frac{1}{2}\left(\frac{\mathcal{P}}{1+\mathcal{P}}\right)^2, \end{aligned} \tag{69}$$

where we used that $\partial P^k_\mu/\partial P^i_\mathrm{m} = (B^i_\mu)^2\delta_{ki}$. Hence, the Fisher matrix for the matter power spectrum is simply a projection of the multitracer Fisher matrix, where we *sum* the Fisher information over all the tracers. Naturally, this result is also identical to what was found in equation (16) in the case of a single tracer – i.e. in that case the PVP estimator reduces to the FKP estimator.

Now, if one fixes the effective biases and wishes to estimate the matter power spectrum alone, then the generalization of equations (29) and (30) follow simply by replacing the functional derivative $\partial/\partial P^i_\mu \to \partial/\partial P^i_\mathrm{m}$, which is also equivalent to taking $\partial/\partial P^i_\mu \to \sum_\mu (B^i_\mu)^2\,\partial/\partial P^i_\mu$. The resulting quadratic form is basically a projection of equation (37):

$$\hat{Q}^{i(\mathrm{PVP})}_\mathrm{m} = \frac{\tilde{V}_i}{2}\left\langle|\tilde{f}|^2\right\rangle^{\boldsymbol{k}_i}, \tag{70}$$

where the weighted field f was defined in equation (35). Therefore, in the PVP estimator the density contrasts of all tracers are combined into a single weighted density contrast, at each point in space. The cross-correlations are all averaged out, in such a way that only the signal to noise of the matter power spectrum is optimized.

The optimal estimator for the matter power spectrum is then simply obtained by multiplying the quadratic form by the inverse of the Fisher matrix, i.e.:

$$\hat{P}^{i(\mathrm{PVP})}_\mathrm{m} = \frac{1}{N_i}\left\langle|\tilde{f}|^2\right\rangle^{\boldsymbol{k}_i}, \tag{71}$$

where the normalization is basically given by equation (69):

$$N_i = \frac{1}{\tilde{V}_i (P^i_\mathrm{m})^2}\int_{v_i}\frac{\mathrm{d}^3x\,\mathrm{d}^3k}{(2\pi)^3}\left(\frac{\mathcal{P}}{1+\mathcal{P}}\right)^2. \tag{72}$$

Noting that $\mathcal{P}/P_\mathrm{m} = \sum_\mu \bar{n}_\mu B^2_\mu$, we see that this estimator is precisely that of PVP.

One may ask also the converse question: what if we want to fix the matter power spectrum P_m, and estimate the effective biases B_μ? In that case, it is a simple exercise to show that this would lead right back to the optimal multitracer estimators, with the only difference that we would end up measuring $\hat{P}^i_\mu/P^i_\mathrm{m}$. However, in reality we can only measure the overall clustering of certain tracers of large-scale structure, which means to estimate the combined product of the matter power spectrum and the (square of the) effective bias. Any distinction between what belongs to the matter power spectrum, and what belongs to the bias, RSDs, NGs, etc. can only be made after some other type of prior knowledge is introduced – e.g. by constraining the normalization and shape of the spectrum

from CMB observations, by modelling the RSDs, or by introducing priors on the bias from gravitational lensing. Evidently, it would be an overuse of information to fix the power spectrum in order to measure the bias, and then employ that bias in order to estimate the power spectrum. Both are measured together in galaxy surveys, and this fundamental degeneracy can only be broken by introducing additional data into the problem.

4.2 The role of cross-correlations

Although our estimators only compute the power spectra of the individual tracers, $P_\mu = B_\mu^2 P_\mathrm{m}$, it is clear from equations (31)–(34) that the cross-correlations of the data, $\langle \delta_\alpha \delta_\beta \rangle$ (with $\alpha \neq \beta$), are also taken into account. In fact, the multitracer estimators express the optimal way to combine both the auto- and the cross-correlations in the computation of the physical parameters B_μ and P_m.

Depending on the total SNR, the power spectra of different tracers can have a positive or negative covariance. Since the SNR of a tracer is given by the amplitude of the power spectrum divided by shot noise, $\mathcal{P}_\mu = P_\mu/(\bar{n}_\mu)^{-1}$, the total SNR of a survey is expressed by the sum of the SNRs, $\mathcal{P} = \sum_\mu \mathcal{P}_\mu$. Hence, when $\mathcal{P} \gg 1$ the total SNR is high, and conversely, the total SNR is low if $\mathcal{P} \ll 1$.

When the total SNR is high, then from equations (41) and (42) we immediately see that the covariance between the clusterings of different types of tracers (the off-diagonal terms) is positive, in fact $C_{\mu\nu} = \mathrm{Cov}(P_\mu, P_\nu) \to 2P_\mu P_\nu$. In relative terms, the covariance in that limit is constant for all tracers, $C_{\mu\nu}/(P_\mu P_\nu) \to 2$. This is simply cosmic variance.

In the converse limit, of very low total SNR, the cross-covariance becomes negative, $C_{\mu\nu} \to -2P_\mu P_\nu/\mathcal{P}^2$ ($\mu \neq \nu$). In relative terms, the covariance in this limit is also independent of the tracer species, as it happens in the high-SNR limit, but now $C_{\mu\nu}/(P_\mu P_\nu) \to -2/\mathcal{P}^2$.

4.3 Tracers with low SNR

An obvious situation of interest arises when some tracer has low SNR. This can happen if the tracer is sparse ($\bar{n}_\mu \lesssim 10^{-5}$), or has a very small bias ($b_\mu \ll 1$), making its clustering strength $\mathcal{P}_\mu$ very low in some bin or bandpower.[4] The danger would be that the inverse of the Fisher matrix (the covariance matrix), which enters in the multitracer estimators through equation (38), could propagate this noise to the estimation of the spectra for the other tracers.

However, this is not the case, as can be seen from the expression for the covariance matrix in equations (41) and (42): because the reciprocals of the individual clustering strengths (i.e. the noises) only appear in the diagonal terms of the covariance matrix, if one of the tracers has a very high noise, this will only affect that same tracer. In particular, this means that our estimators are robust even when a galaxy survey includes tracers whose SNR are small.

This feature is very convenient if one would like to split a survey into several subsurveys, by dividing galaxies, quasars and other objects into different categories according to type, luminosity, colour, morphology, etc. – all of which may be indicators of the bias of those tracers. In doing that, even though the total SNR of the survey should remain approximately constant, the SNR of each individual tracer would decrease, leading us to wonder whether this could lead to a degradation of the information derived from that survey. However, the fact that a tracer with low SNR only affects its own estimator means that this strategy can be safely used even when some tracers have very low number densities.

[4] Notice that the values of the power spectra of the tracers, $P_\mu = B_\mu^2 P_\mathrm{m}$, should never all vanish. If they do, in some sense (e.g. on extremely large or small scales), then $\mathcal{P} \to 0$, making the entire Fisher matrix vanish for that bin – as it should indeed happen in that case.

4.4 Shot noise and the one-halo term

A fundamental assumption in our derivations has been that the covariance of the counts of the tracers is given by equation (19). However, this is often a simplification.

First, the statistics of counts in cells for galaxies in a redshift survey is only approximately Poissonian, so shot noise may be very different from the usual $1/\bar{n}_\mu$. Moreover, besides the two-halo term which usually dominates on large scales, there is an additional contribution to the power spectrum from the one-halo term (Cooray & Sheth 2002). In the $k \to 0$ limit the one-halo term is effectively an additional contribution to shot noise. In principle, any such corrections can be fixed simply by allowing for a more general form of shot noise for each tracer which, in the limit of negligible one-halo term and Poisson statistics, reduces to $\delta_{\mu\nu}/\bar{n}_\mu$.

A closely related problem arises when different types of tracers occupy the same dark matter haloes. Equation (19) states that the covariance between counts of different types of tracers do not have any shot noise. However, the Halo Model specifies that even for galaxies of different types there is a non-vanishing one-halo term, which is degenerate with shot noise in the $k \to 0$ limit. Usually this is a small contribution, subdominant to the shot noise of the individual tracers, but it ultimately means that the noise cannot be assumed to be diagonal in the tracers.

A third, and perhaps more serious problem, arises from that fact that different tracers are often found to inhabit haloes of very similar masses. Most galaxies (as well as quasars) are found in haloes of masses in the range $10^{13} h^{-1}\mathrm{M}_\odot \lesssim M_\mathrm{h} \lesssim 10^{15} h^{-1}\mathrm{M}_\odot$, with relatively small differences between the distributions of each type of object within haloes – the so-called HODs (Martinez & Saar 2001; Cooray & Sheth 2002). In particular, this means that the biases of those tracers are not entirely independent.

In other words, different tracers can be correlated by more than just the underlying dark matter field. These correlations arise through the one-halo terms of the power spectra, which contribute to the covariances of the counts of those tracers, as well as through additional contributions to the bispectrum and trispectrum. But the trispectrum also defines the covariance of the power spectra through $\langle\langle P_\mu(\boldsymbol{k}) P_\nu(\boldsymbol{k}')\rangle\rangle \sim T_{\mu\mu\nu\nu}(\boldsymbol{k}, -\boldsymbol{k}, \boldsymbol{k}', -\boldsymbol{k}')$, which means that it is not possible to assume that the trispectrum is given only by the connected pieces of the four-point function – i.e. it is not true anymore that $\langle\langle \delta_\mu \delta_\nu \delta_\alpha \delta_\beta \rangle\rangle = C_{\mu\nu} C_{\alpha\beta} + C_{\mu\alpha} C_{\nu\beta} + C_{\mu\beta} C_{\alpha\nu}$.

It is straightforward to incorporate the one-halo term systematically into the covariance in all our calculations (see Section 6). However, if there are significant correlations between the power spectra arising from the one-, two- and three-halo terms of the trispectrum, then the counts cannot be assumed to be nearly Gaussian. In that case it would be erroneous to assume that the tracers are truly independent, and a key assumption of our method would be undermined. Nevertheless, Smith & Marian (2015) were able to extend the PVP method (which does not rely on a direct construction based on the pixel covariance, but on variational methods) to incorporate these contributions from the Halo Model in formal expressions for the weights and for the Fisher matrix. However, recall that the PVP method, as well as its extension by Smith

& Marian (2015), only tackle the estimation of the matter power spectrum, after assuming that the bias, RSDs, NGs, etc. are known and fixed.

4.5 Degenerate tracers

While tracers with different biases can possess correlations beyond those associated with the large-scale structure of the Universe, it is not necessarily true that two tracers that have similar biases must be highly correlated. Two types of galaxies may have different HODs, but their biases could coincide. In those cases, if there is a significant contribution from the one-halo term, then it may still make sense to treat those species separately. It is only when two tracers have the same HOD (or, equivalently, the same bias, one-halo term, two-halo term, three-halo term, etc.), that they should be consolidated into a single species.

However, suppose we do not know whether or not two types of galaxies have the same HODs. If we use the multitracer approach and treat those two species as if they were different tracers, but they turn out to have the same HODs, would that initial assumption imply an overestimate of the information, or some distortion in the estimators?

The answer is no, and this follows from a very interesting property of the multitracer Fisher matrix. As shown in Abramo & Leonard (2013), the Fisher matrix can be diagonalized by changing variables, from the original power spectra $P_\mu = B_\mu^2 P_{\rm m}$ to a new set of parameters which correspond to the total clustering strength and certain ratios between the power spectra – the *relative clustering strengths*. In the case of two tracers with spectra P_1 and P_2, a choice of parameters which diagonalizes the Fisher matrix is $\log \mathcal{P} = \mathcal{P}_1 + \mathcal{P}_2$, and $\log \mathcal{R} = \log \mathcal{P}_1/\mathcal{P}_2$ (or, equivalently, $\log \mathcal{P}$ and $\log \mathcal{P}_2/\mathcal{P}_1 = -\log \mathcal{R}$). The Fisher information per unit of phase space for this new set of parameters is

$$\mathcal{F}[\log \mathcal{P}, \log \mathcal{R}] = \begin{pmatrix} \frac{1}{2}\frac{\mathcal{P}^2}{(1+\mathcal{P})^2} & 0 \\ 0 & \frac{1}{4}\frac{\mathcal{P}^2\,\mathcal{R}}{(1+\mathcal{P})(1+\mathcal{R})^2} \end{pmatrix}. \quad (73)$$

For an arbitrary number N of tracers, the change of variables that diagonalizes the Fisher matrix is identical to a change from Cartesian coordinates to spherical coordinates in N dimensions. Namely, if we regard the N clustering strengths as $\mathcal{P}_1 \to x_1^2$, $\mathcal{P}_2 \to x_2^2$, etc. then the variables that diagonalize the Fisher matrix are the radius, $\mathcal{P} \to r^2 = \sum_\mu x_\mu^2$, together with the $(N-1)$ angles $\tan^2\theta = (r^2 - x_N^2)/x_N^2$, $\cot^2\phi_1 = (r^2 - x_N^2 - x_{N-1}^2)/x_{N-1}^2$, $\cot^2\phi_2 = (r^2 - x_N^2 - x_{N-1}^2 - x_{N-2}^2)/x_{N-2}^2$, etc. Hence, the angle variables correspond to certain ratios between the tracers, (or *relative* clustering strengths), for which the matter power spectrum (the radius) cancels out. In particular, this means that the relative clustering strengths are immune to some statistical limitations that affect the matter power – namely, the relative clustering strengths can be measured to an accuracy which is not constrained by cosmic variance (Abramo & Leonard 2013).

Coming back to our example of the two tracers, if we now stipulate that they are in fact a single species, then $P_1 = P_2$, and $\mathcal{R} \to \bar{n}_1/\bar{n}_2$ is not a free parameter anymore, so $d\log\mathcal{R} \to 0$. This is equivalent to projecting the 2×2 Fisher matrix into a single component, thus eliminating the line and column corresponding to $\log\mathcal{R}$, and leaving $\log\mathcal{P}$ as the sole free parameter. Indeed, since $d\log\mathcal{R} \to 0$ in this case, we cannot constrain physical parameters such as RSDs or NGs on the basis of a measurement of $\mathcal{R}$.

Table 1. The three cases we use to illustrate the application of the multi-tracer method. In all cases tracer 1 has bias $b_1 = 1.0$, and tracer 2 has bias $b_2 = 1.2$. In case A, the two tracers have high number densities, so the signal to noise is high. In case B, tracer 1 is dense, but tracer 2 is sparse. In case C, both tracers are sparse, so the signal to noise is low.

Case	$\bar{n}_1$ (h^3 Mpc^{-3})	b_1	$\bar{n}_2$ (h^3 Mpc^{-3})	b_2
A	1×10^{-2}	1.0	$1. \times 10^{-2}$	1.2
B	1×10^{-2}	1.0	$1. \times 10^{-5}$	1.2
C	$1. \times 10^{-5}$	1.0	1×10^{-5}	1.2

Since the Fisher matrix is diagonal, the Fisher information for $\log\mathcal{P}$ is unchanged after this projection (or marginalization). In particular, the variance $\sigma^2(\log\mathcal{P}) = \sigma^2(\mathcal{P})/\mathcal{P}^2$ is untouched by a marginalization over $\mathcal{R}$, and it is still given by the inverse of the same Fisher matrix element in equation (73), so $\sigma^2(\mathcal{P}) = 2(1+\mathcal{P})^2$, which is nothing but the covariance (in units of phase space volume) for a single tracer species – see equation (17).

The argument above extends to any number of tracers: since the Fisher matrix is diagonal in the 'spherical coordinates' (the total and relative clustering strengths), projecting some of the tracers out by combining them into new species does nothing to the Fisher information of the total clustering strength, or to the relative clustering strengths of the remaining species. Therefore, in principle there is no difference between treating two identical tracer species (with the same HODs) separately, or joining them into a single type of tracer. Of course, one can always *destroy* information by treating two *different* tracer species as if they were just one, but there is no penalty for breaking a catalogue into as many subcatalogues as one wishes – even if some of the tracers turn out to be completely degenerate.

The argument is a bit more involved if we work with the power spectra as the parameters, but the conclusion is the same (see Appendix A).

5 TESTING THE ESTIMATORS

In Sections 2 and 3, we derived the optimal multitracer estimators. We also obtained the covariance of the estimators – which is simply the inverse of the multitracer Fisher matrix. In this Section we apply that formalism to simple simulated galaxy maps. The implementation of the estimators is quite straightforward, and should be familiar to anyone who has used the FKP or the PVP methods. Although we test the method in real space, the extension to redshift space is trivial: instead of bins in $|\boldsymbol{k}|$, one should have bins both in k and in μ_k^2.

For the generation of the galaxy maps we chose a simple method that is both efficient and computationally cheap enough that hundreds of realizations of a single fiducial matter power spectrum and galaxy model can be analysed. We implemented the multitracer estimators in a cubic grid with constant, uniform mean number density (or selection function), for the case of two different species of tracers, with biases $b_1 = 1.0$ and $b_2 = 1.2$. We checked that the estimators are as robust as the FKP or PVP methods against variations in the survey geometry.

In order to test the performance of the estimators in situations of high or low signal to noise, we consider three different cases, as shown in Table 1. In each case we generate 1000 galaxy maps (each map consisting of two catalogues, one for each tracer), and estimate the spectra using the methods described in Section 3.

5.1 Lognormal maps

Our mocks follow the same procedure used in, e.g. PVP. A detailed description of the generation of lognormal maps can be found in Coles & Jones (1991). The basic idea is that a Gaussian density contrast $\delta^{(\mathrm{G})}(\boldsymbol{x})$ is not bounded from below, which implies that negative values for the density are possible in any finite-volume realization of such a Gaussian field. Lognormal fields, on the other hand, are positive definite, so we map the Gaussian field into a lognormal field.

A lognormal field obeys the condition $\delta^{(\mathrm{L})}(\boldsymbol{x}) \geq -1$ and approximately describes the non-linear density field at low redshifts. We can obtain a lognormal density field in terms of a Gaussian density field through the definition $1 + \delta^{(\mathrm{L})}(\boldsymbol{x}) = \exp[\delta^{(\mathrm{G})}(\boldsymbol{x}) - \sigma_{\mathrm{G}}^2/2]$, where σ_{G}^2 is the variance of the Gaussian field inside a cell. The Gaussian correlation function is related to the physical (assumed lognormal) correlation function by $\xi^{(\mathrm{G})}(x) = \ln[1 + \xi^{(\mathrm{ph})}(x)]$. Given a fiducial cosmology, we obtain the $z = 0$ matter power spectrum $P_{\mathrm{m}}(k)$ from the Boltzmann code CAMB[5] (Lewis, Challinor & Lasenby 2000), and inverse-Fourier transform it to get the physical correlation function $\xi^{(\mathrm{ph})}(x)$. We then convert the physical (assumed lognormal) correlation function to the correlation function of the corresponding Gaussian field, and Fourier-transform that correlation function into a power spectrum for the Gaussian field. This is the power spectrum which is employed to generate the Gaussian random modes for the density contrast.

The next step is the generation of biased lognormal maps for each galaxy type. We *define* the lognormal maps as $1 + \delta_{\mu}^{(\mathrm{L})}(\boldsymbol{x}) = \exp[b_{\mu}\,\delta^{(\mathrm{G})}(\boldsymbol{x}) - b_{\mu}^2\sigma_{\mathrm{G}}^2/2]$.[6] Finally, we create the galaxy maps as independent Poisson realizations over the lognormal fields. Each tracer has its own spatial number density $\bar{n}_{\mu}(\boldsymbol{x})$ and bias b_{μ}, so that the maps for each tracer are given by integer numbers for each cell of volume $\mathrm{d}V$ in our cube through a Poisson sampling, $N_{\mu}(\boldsymbol{x}) \leftarrow \mathbb{P}\{\bar{n}_{\mu}(\boldsymbol{x})[1 + \delta_{\mu}^{(\mathrm{L})}(\boldsymbol{x})]\mathrm{d}V\}$, where $\mathbb{P}\{\lambda\}$ is a Poisson distribution with mean λ.

In the three cases detailed above, we considered cubic 256^3 grids with a fiducial cosmology characterized by a flat Λ cold dark matter (ΛCDM) model with $\Omega_{\mathrm{b}}h^2 = 0.0226$, $\Omega_{\mathrm{CDM}}h^2 = 0.112$ and $h = 0.72$. Each cube has a physical (comoving) volume of $(1280\,h^{-1}\mathrm{Mpc})^3$. It is important to note that lognormal maps created this way do not show the usual effect of suppression in power at small scales when a smoothing algorithm is applied to convert from a continuous distribution to a discrete grid, such as nearest grid point. In any case, the formalism is general enough to accommodate this necessity. Furthermore, since the grid used is cubic, it is unnecessary to deconvolve the estimated spectra from the window function. Even though any discretization scheme could be used, the square grid is required in order to employ an implementation in terms of a fast Fourier transform (FFt), which is, as a matter of fact, the only practical way to perform a Fourier analysis of large data sets.

[5] http://CAMB.info

[6] Notice that, for a lognormal map with bias b, the correlation function used in the generation of the Gaussian random modes should be defined as $\xi^{(\mathrm{G})}(x) = b^{-2}\ln[1 + b^2\,\xi^{(\mathrm{ph})}(x)]$. Therefore, strictly speaking, this prescription only is self-consistent when there is a single type of galaxy, with one bias. However, using the same correlation function for tracers of different biases introduces only a small spectral distortion on small scales, which we corrected for in our simulations.

5.2 The data analysis algorithm

With the galaxy maps $n_{\mu}(\boldsymbol{x})$ as input, along with an initial guess for the biases b_{μ}, we can start to deploy the machinery developed in Sections 2 and 3. A previous step, in case we had not explicitly generated maps with constant, uniform number densities, would be to estimate $\bar{n}_{\mu}(\boldsymbol{x})$.

We start by constructing random maps, $n_{\mu}^{r}(\boldsymbol{x})$, for each tracer as a Poisson process, in each cell of the grid, with the same shape for the mean number density as the data (i.e. the real maps), but with a larger number of particles, $\bar{n}_{\mu}^{r} = \bar{n}_{\mu}/\alpha_{\mu}$, where α_{μ} are small constants. We then construct the density contrasts according to equation (59): $\delta_{\mu}(\boldsymbol{x}) = (n_{\mu}^{d} - A_{\mu}\,n_{\mu}^{r})/\bar{n}_{\mu}$ – where recall that A_{μ} are constants found according to the discussion in Section 3.3.

With an initial guess for the biases and for the amplitude of the power spectrum, we can construct $\mathcal{P}_{\mu}$ and $\mathcal{P} = \sum_{\mu}\mathcal{P}_{\mu}$, plug them into the weights (equation 33), and calculate the weighted density contrasts of equation (59). We then perform an FFt over $f(\boldsymbol{x})$ and $f_{\mu}(\boldsymbol{x})$, in order to obtain the integrand of equation (36). Taking proper care of the volume factors (in real and in Fourier space), this step should be analogous to the average over modes in equation (2.4.5) of FKP.

The next step is to subtract the biases of the estimators – the ΔQ_{μ}^{i} in equation (36) or, equivalently, equation (67). Assuming that averages over bins are such that $\langle A\,B\rangle^{i} \approx \langle A\rangle^{i}\,\langle B\rangle^{i}$, and taking a single value for all the $\alpha_{\mu} \to \alpha$, equation (40) can be rearranged to yield:

$$\Delta Q_{\mu}^{i} = \frac{1+\alpha}{2}\int_{v_i}\frac{\mathrm{d}^3x\,\mathrm{d}^3k}{(2\pi)^3}\frac{\bar{n}_{\mu}}{(1+\mathcal{P})^2}. \tag{74}$$

With our choice of $\alpha = 10^{-6}$, we find that $A_{\mu} \to \alpha$ to an excellent approximation, which means that the biases of the estimators are given only by equation (74) – see Section 3.3. Finally, the estimated power spectra are computed with the help of equation (38).

We present our results for the estimated spectra of two types of tracers in three cases, A, B and C – see Table 1 and Fig. 1. Case A represents a low-redshift survey which is highly complete, so both tracers are dense. Case B represents a low- or intermediate-redshift survey, with one dense species of tracer (type 1 – say, red galaxies) and one sparse species of tracer (type 2 – say, quasars). Case C represents a high-redshift survey, with two sparse types of tracers.

Our estimates were evaluated in evenly separated bandpowers with $\Delta k = 0.005\,h\,\mathrm{Mpc}^{-1}$. We show the estimated spectra in Fig. 1, only up to $k = 0.2\,h\,\mathrm{Mpc}^{-1}$ – slightly into the non-linear regime but still below the Nyquist frequency, such that our results are not affected by discretization effects. When estimating the spectra we adopted a commonly used simplification, which is to fix the value of the matter power spectrum that is used in the weights, equation (33) – in our case, we found that fixing $P_{\mathrm{m}} \to 10^4\,h^{-3}\,\mathrm{Mpc}^3$ in the weights was a suitable choice. Our results did not change significantly over the dynamical range of interest when that value was multiplied by 2 or by 1/2.

5.3 Empirical versus theoretical covariances

We now check whether the theoretical covariance matrix (the inverse of the multitracer Fisher matrix) is a good approximation to the true (i.e. empirical) covariance matrix. If the theory is accurate, then the method is validated; if it is not, then the multitracer estimators are suboptimal.

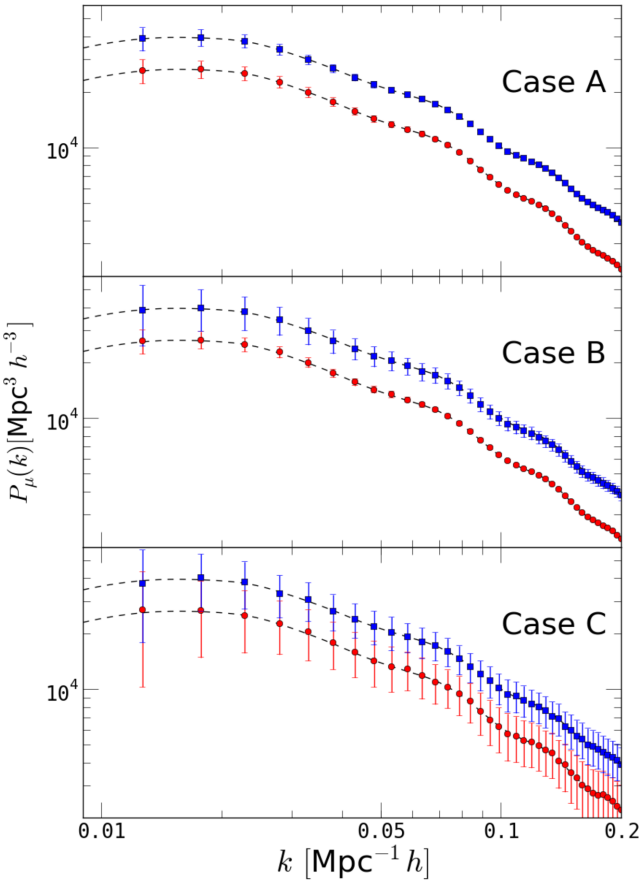

Figure 1. Estimated auto-spectra *v.* real auto-spectra. Filled (red) circles correspond to the power spectrum of the tracer 1, with $b_1 = 1.0$, and filled (blue) squares correspond to the estimated power spectrum of the tracer 2, with bias $b_2 = 1.2$. The symbols and error bars correspond to the mean and to the variance, respectively, of 1000 realizations. The dashed lines are the input (theoretical) spectra of the tracers, given their biases and our fiducial cosmology. The upper, middle and lower panels correspond to cases A, B and C, respectively (see Table 1). The error bars are the theoretical ones – i.e. the inverse of the Fisher matrix, equation (25).

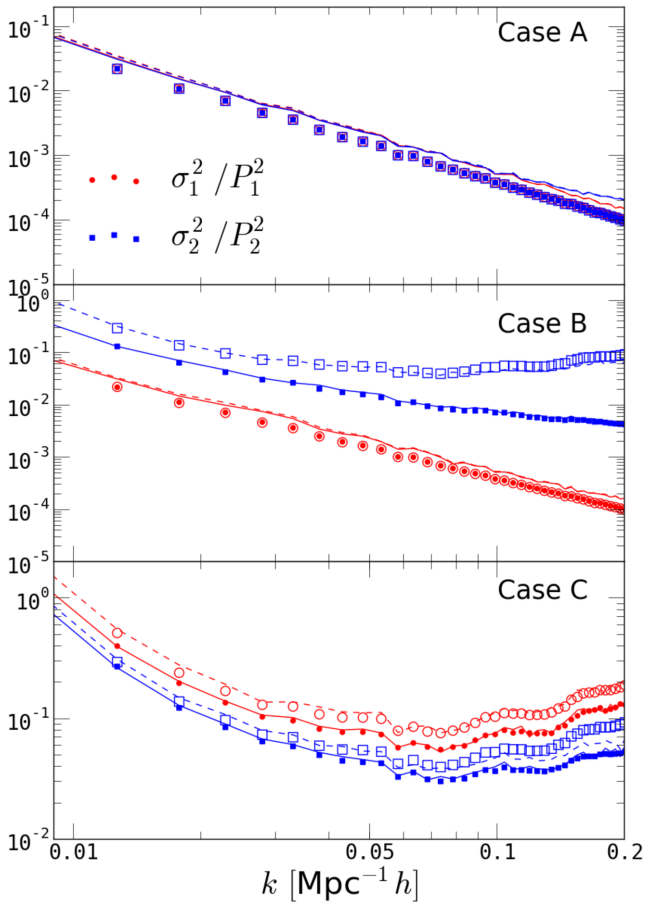

Figure 2. Theoretical versus empirical relative covariances of the auto-spectra, $\mathrm{Cov}(P_\mu^i, P_\mu^i)/(P_\mu^i)^2$. The upper, middle and lower panels correspond to cases A, B and C, respectively (see Table 1). Red circles and blue squares correspond to the theoretical covariances of the tracers 1 and 2, respectively. The lines of the same colours are the standard deviation of our 1000 lognormal mocks. Solid symbols and lines correspond to multitracer estimates, while open symbols and dashed lines correspond to FKP estimates. In case A (upper panel), since the two tracers have high signal to noise (both $\mathcal{P}_1 \gg 1$ and $\mathcal{P}_2 \gg 1$ in this range of scales), both the multitracer and the FKP formulas for the auto-covariances reduce to $\mathrm{Cov}(P_\mu^i, P_\mu^i)/(P_\mu^i)^2 \simeq 2/v_i = 2/(V_i\,\bar{V}_i) \sim k_i^{-2}$ (see equations (41) and (42)). Hence, in this case most symbols and lines overlap. In most cases, the empirical covariances are slightly higher than the theoretical ones – as expected. In case B (middle panel), the covariance of spectrum of the sparse tracer species is significantly higher in the FKP method: in this case, the multitracer method reduces the uncertainty in the spectrum by a large factor.

The empirical result was obtained from 1000 realizations. This was compared with the theoretical covariance – i.e. the inverse of the binned Fisher matrix of equation (25):

$$\mathrm{Cov}(P_\mu^i,\, P_\nu^j) = \delta_{ij}\left[\frac{1}{P_\mu^i\, P_\nu^i}\int_{v_i}\frac{\mathrm{d}^3x\,\mathrm{d}^3k}{(2\pi)^3}\mathcal{F}_{\mu\nu}\right]^{-1}, \qquad (75)$$

where $\mathcal{F}_{\mu\nu}$ was defined in equation (26).

In Fig. 2, we present the comparison between the theoretical and empirical covariances for the auto-spectra of the two species, obtained, respectively, from equation (75) and from taking the standard deviation of 1000 lognormal realizations. We find that our theoretical expression properly reproduces the behaviour of the statistical fluctuations in all cases, matching more closely the variances when compared with the FKP method. The theoretical variances sometime underestimate slightly the empirical variance, which is consistent with the notion that the inverse of the Fisher matrix is an underestimate of the true covariance – which is consistent with what is usually found in implementations of the FKP method. In cases B and C the multitracer estimator performs better (much better, in case B) than the FKP estimator on all scales. The reason this improvement is that, in the multitracer analysis, the estimation of the power spectrum of any one type of tracer factors in the data from all tracers – in particular, the estimation of the power spectrum of a sparse tracer includes data coming also from the more abundant tracer populations.

In Fig. 3, we compare the theoretical and empirical cross-covariances for the spectra of the two tracers (green triangles), as well as the covariance for the ratios of the two spectra, P_1/P_2 (black diamonds). Since the FKP method cannot predict theoretical covariances in these two cases, we only show the multitracer theoretical variances. The theoretical variance for the ratio P_1/P_2 follows from the multitracer Fisher information matrix, equation (26), which can be diagonalized by a change of variables (Abramo

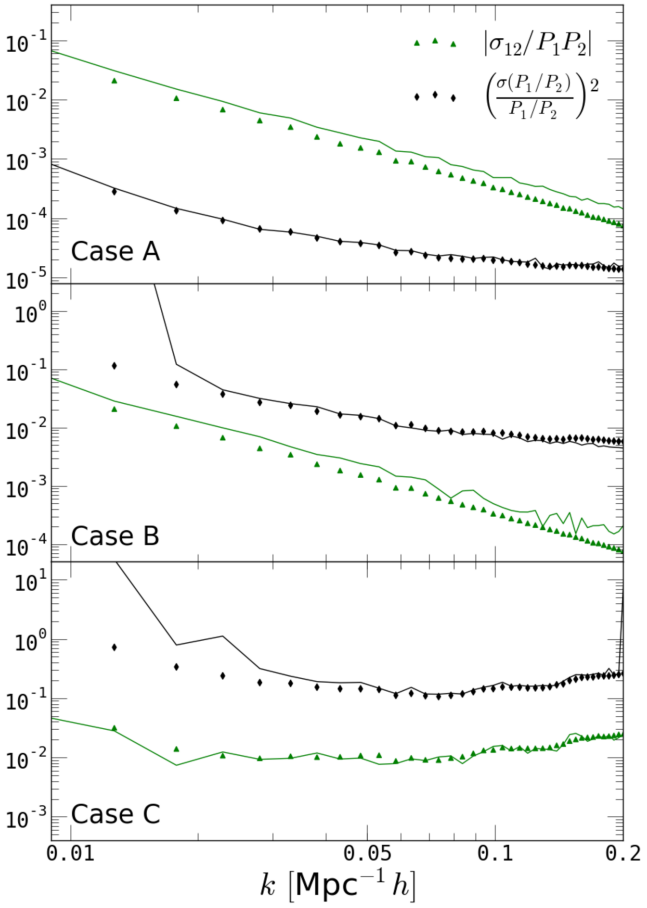

Figure 3. Theoretical versus empirical covariances between the different auto-spectra, $\sigma_{12} = \mathrm{Cov}(P_1^i, P_2^i)$, and for the ratio between the spectra, $\mathrm{Cov}(P_1^i/P_2^i, P_1^i/P_2^i)$ (the relative covariance is identical for P_2/P_1). The upper, middle and lower panels correspond to cases A, B and C, respectively (see Table 1). Diamonds (black) correspond to the theoretical relative cross-covariances of the two spectra, while triangles (green) correspond to the theoretical covariance for the ratios between the spectra (see text). The solid lines correspond to the empirical covariances, using the multitracer estimators (we do not show the results using the FKP estimator in these plots because it performs significantly worse compared with the multitracer estimators, and in any case the FKP method does not predict these covariances). Notice that in case C (lower panel) the cross-covariance is negative, since $\mathcal{P} < 1$ – see equations (41) and (42). Notice also that in case A the ratio between the spectra has a much lower relative uncertainty than the cross-correlation between the two spectra (for an explanation, see the text).

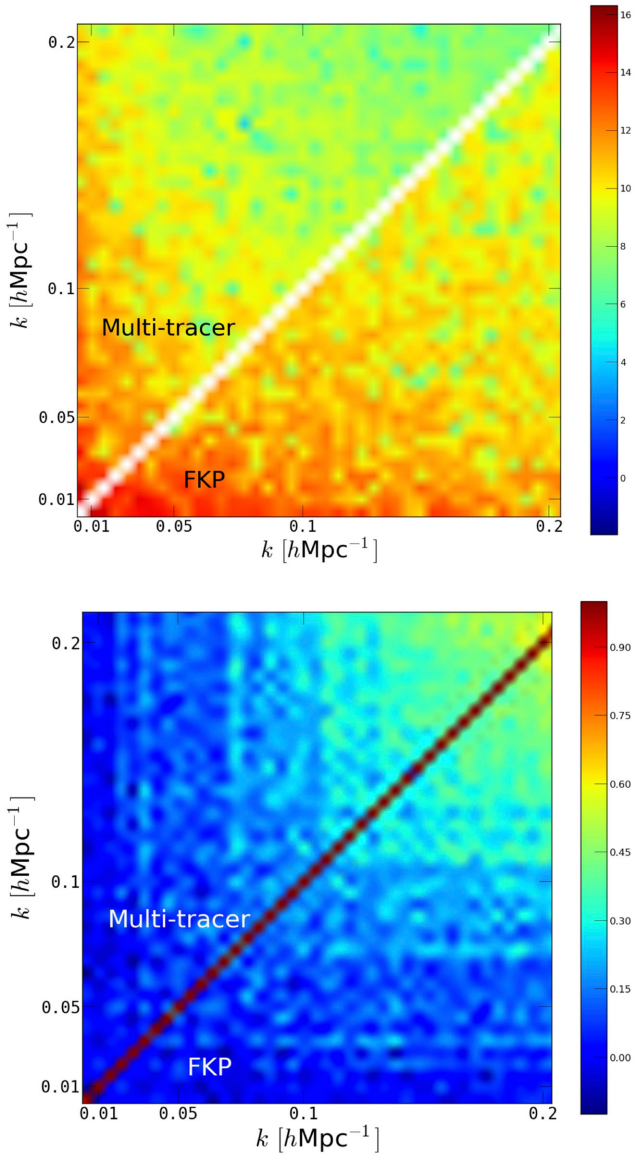

Figure 4. Upper panel: covariance matrix for tracer 2 in case B. The upper triangle is the result using the multitracer estimator, and the lower triangle results from using the FKP estimator. The multitracer technique performs significantly better on all scales. Lower panel: correlation matrix for tracer 2 in case A. Both the multitracer and the FKP estimators perform similarly regarding the correlations between Fourier bins. We checked that the estimators result in similar correlation matrices for both tracers, in the three different cases we analysed.

& Leonard 2013), where the new parameters (the 'eigenvectors' of the Fisher matrix) are not the individual clustering strengths $\mathcal{P}_\mu$, but the total clustering strength, $\mathcal{P}$, and certain ratios between the clustering strengths. In particular, a diagonal Fisher matrix means that the degrees of freedom are independent – there are no cross-covariances. For two types of tracers, the variables which diagonalize the 2×2 Fisher matrix are $\mathcal{P} = \mathcal{P}_1 + \mathcal{P}_2$, and $\mathcal{P}_1/\mathcal{P}_2$ (or, equivalently, $\mathcal{P}$ and $\mathcal{P}_2/\mathcal{P}_1$). As shown in Abramo & Leonard (2013), the Fisher matrix per unit of phase space volume for $\log(P_1/P_2)$ is $F_{\rm ratio} = \mathcal{P}_1\,\mathcal{P}_2/4(1 + \mathcal{P}_1 + \mathcal{P}_2)$, from which follows that the relative covariance of that ratio is $[\int \mathrm{d}^3x\,\mathrm{d}^3k/(2\pi)^3 F_{\rm ratio}]^{-1}$. This figure demonstrates the power of the multitracer technique to measure $P_1/P_2 = B_1^2(z, k, \mu_k)/B_2^2(z, k, \mu_k)$, something that can be used to place stronger constraints not only the biases of the two species, but also on RSDs, NGs, etc.

The upper panel of Fig. 4 shows the covariance matrix for tracer 2 ($b_2 = 1.2$) in case B – i.e. $\mathrm{Cov}_{22}^{(B)}(k_i, k_j)$. We exploited the symmetry of the covariance matrix under $k_i \leftrightarrow k_j$ in order to compare the multitracer and FKP estimators directly. In the lower panel of this figure we show the correlation matrix, defined as $\mathrm{Corr}_{ij} = \mathrm{Cov}_{ij}/\sqrt{\mathrm{Cov}_{ii}\,\mathrm{Cov}_{jj}}$. We find that both the multitracer and the FKP estimators yield roughly similar correlation matrices, with weakly correlated bins up to scales $k \lesssim 0.1\,h\,\mathrm{Mpc}^{-1}$.

The upper panel of Fig. 4, together with the middle panels of Figs 2 and 3, shows that in case B the multitracer estimator performs significantly better than the FKP estimator at all scales, with uncertainties up to one order of magnitude smaller for the spectrum of the sparse tracer. The multitracer technique is also clearly

superior in estimating the auto-spectra in case C, when both tracers are sparse – see the lower panel of Fig. 2.

6 INCLUDING THE ONE-HALO TERM

The fundamental object in this paper, which was used to derive the Fisher information matrix, as well as the optimal weights, is the pixel covariance. In the limit where bias and RSDs depend weakly on $\boldsymbol{k}$, the covariance can be approximated by equation (20). However, this is not a complete description: in addition to the 'signal', $P_\alpha = B_\alpha^2 P_{\rm m}$, and the shot noise, $\delta_{\alpha\beta}/\bar{n}_\alpha$, there is another source of correlations between the density contrasts of different species of tracers at different points in space: the one-halo term of the power spectrum. According to the Halo Model (Cooray & Sheth 2002), dark matter haloes are the genuine tracers of the underlying matter density, while galaxies only trace the haloes. In particular, this means that many galaxies may be hosted by the same halo, in which case they would be tracing the same features of the underlying fluctuations of the matter density.

This additional covariance between galaxy counts is expressed by the one-halo term:

$$P^{\rm 1h}_{\alpha\beta}(k) = \frac{1}{\bar{n}_\alpha\,\bar{n}_\beta}\int {\rm d}\ln M\,\frac{{\rm d}\,\bar{n}_h}{{\rm d}\ln M}\,u^2(k|M)\,\langle\langle N_\alpha N_\beta\rangle\rangle_M, \tag{76}$$

where ${\rm d}\,\bar{n}_h/{\rm d}\ln M$ is the mass function for haloes of mass M, N_α is the number of galaxies of type α, and $u(k|M)$ is the Fourier transform of the halo profile (Cooray & Sheth 2002). The expectation value is over the probability distribution function for the numbers of galaxies (the HOD) at a given halo mass. For the species of tracers which are typically used in cosmological surveys the one-halo term is only relevant on small scales ($k \gtrsim 1\,h\,{\rm Mpc}^{-1}$) – although, since $u(k \to 0) = 1$, it still contributes a constant factor on large scales.

Inclusion of the one-halo term would lead the approximated pixel covariance of equation (20) to assume the expression:

$$C_{\alpha\beta}(\boldsymbol{x},\boldsymbol{x}') \to \delta_{\rm D}(\boldsymbol{x},\boldsymbol{x}')\times\left[\frac{\delta_{\alpha\beta}}{\bar{n}_\alpha} + P^{\rm 2h}_{\alpha\beta} + P^{\rm 1h}_{\alpha\beta}\right], \tag{77}$$

where we write the two-halo term $P^{\rm 2h}_{\alpha\beta} = B_\alpha B_\beta P_{\rm m}$. In principle, if we are only interested on the properties of the clustering on large scales, this term can be included systematically, in every step of the calculations – see also Hamaus et al. (2010), in a similar context. These are straightforward computations, but for a general form of $P^{\rm 1h}_{\alpha\beta}$ there is no closed-form expression for the inverse of the pixel covariance matrix, which means that we cannot give explicit formulas for the Fisher matrix, the weights, the window functions, etc.

6.1 Fisher matrix of the two-halo term for separable one-halo terms

In some cases the populations of tracers are such that the one-halo term is approximately separable, i.e. it can be expressed as a direct product of two terms, $P^{\rm 1h}_{\alpha\beta} \sim H_\alpha H_\beta$ – just as happens with the two-halo term. We have checked that, for a class of HODs that is commonly used to describe red and blue galaxies (Zheng et al. 2005), all entries of the correlation matrix $P^{\rm 1h}_{\alpha\beta}/\sqrt{P^{\rm 1h}_{\alpha\alpha}P^{\rm 1h}_{\beta\beta}}$ are very close to unity, which justifies this approximation. However, we only verified this feature of the one-halo term while ignoring the distinction between central and satellite galaxies, since it is not clear how to generalize $\langle N_\alpha N_\beta\rangle$ in that case. It would be interesting to find out whether this property holds for more realistic HODs.

If the one-halo term is separable, it turns out that we *can* invert the covariance matrix. This result follows from the exquisite properties of matrices that can be written as $M_{\alpha\beta} = \delta_{\alpha\beta} + v_\alpha v_\beta + u_\alpha u_\beta$. This type of matrix appeared already in Section 2, where we showed that the inverse of $M_{v,\,\alpha\beta} = \delta_{\alpha\beta} + v_\alpha v_\beta$ is given by $M^{-1}_{v,\alpha\beta} = \delta_{\alpha\beta} - v_\alpha v_\beta/(1+v^2)$, where $v^2 = \sum_\mu v_\mu^2$.

As shown in Appendix B, the inverse of the matrix $M_{\alpha\beta} = \delta_{\alpha\beta} + v_\alpha v_\beta + u_\alpha u_\beta$ is

$$M^{-1}_{\alpha\beta} = \sum_{\mu\nu} M^{-1/2}_{v,\alpha\mu}\,M^{-1}_{u',\mu\nu}\,M^{-1/2}_{v,\nu\beta}, \tag{78}$$

where

$$M^{-1/2}_{v,\alpha\beta} = \delta_{\alpha\beta} - v_\alpha v_\beta/(1+v^2+\sqrt{1+v^2}), \quad \text{and}$$

$$u'_\alpha = \textstyle\sum_\mu M^{-1/2}_{v,\alpha\mu}u_\mu.$$

After some algebra, using equation (78) we can express the inverse of the covariance, equation (77), as

$$C^{-1}_{\alpha\beta}(\boldsymbol{x},\boldsymbol{x}') \to \delta_{\rm D}(\boldsymbol{x},\boldsymbol{x}') \times\left[\delta_{\alpha\beta}\bar{n}_\alpha - \bar{n}_\alpha\frac{P^{\rm 2h}_{\alpha\beta}+P^{\rm 1h}_{\alpha\beta}+Y_{\alpha\beta}}{1+\mathcal{T}}\bar{n}_\beta\right], \tag{79}$$

where the cross-term is

$$Y_{\alpha\beta} = \sum_\mu \bar{n}_\mu\Big(P^{\rm 2h}_{\alpha\mu}\,P^{\rm 1h}_{\mu\beta} + P^{\rm 1h}_{\alpha\mu}\,P^{\rm 2h}_{\mu\beta} - P^{\rm 2h}_{\alpha\beta}\,P^{\rm 1h}_{\mu\mu} - P^{\rm 1h}_{\alpha\beta}\,P^{\rm 2h}_{\mu\mu}\Big), \tag{80}$$

and the term appearing of the denominator in equation (79) is

$$\mathcal{T} = \sum_\mu \bar{n}_\mu\,(P^{\rm 2h}_{\mu\mu} + P^{\rm 1h}_{\mu\mu}) + \sum_{\mu\nu}\bar{n}_\mu\bar{n}_\nu(P^{\rm 2h}_{\mu\mu}\,P^{\rm 1h}_{\nu\nu} - P^{\rm 2h}_{\mu\nu}\,P^{\rm 1h}_{\mu\nu}). \tag{81}$$

Compare this result with equation (22). We detect some familiar expressions, in particular:

$$\mathcal{P} = \sum_\mu \bar{n}_\mu B_\mu^2 P_{\rm m} \equiv \sum_\mu \mathcal{P}_\mu = \sum_\mu \bar{n}_\mu P^{\rm 2h}_{\mu\mu}. \tag{82}$$

It is now useful to rename the clustering strength of the two-halo term as $\mathcal{P}_\mu \to \mathcal{P}^{\rm 2h}_\mu$, $\mathcal{P} \to \mathcal{P}^{\rm 2h}$, and to define the one-halo clustering strength as $\mathcal{P}^{\rm 1h} = \sum_\mu \bar{n}_\mu P^{\rm 1h}_{\mu\mu} = \sum_\mu \mathcal{P}^{\rm 1h}_\mu$. The cross-terms mixing the one-halo and the two-halo terms appear in the combinations:

$$P^c_{\alpha\beta} \equiv \sum_\mu P^{\rm 2h}_{\alpha\mu}\,\bar{n}_\mu\,P^{\rm 1h}_{\mu\beta},$$

$$\mathcal{P}^c_{\alpha\beta} \equiv \bar{n}_\alpha\bar{n}_\beta P^{\rm 2h}_{\alpha\beta}P^{\rm 1h}_{\alpha\beta}. \tag{83}$$

Once again, we find it useful to define the dimensionless clustering strengths of these cross-terms, as was done for the two-halo and the one-halo terms. They are

$$\mathcal{P}^c_\alpha \equiv \bar{n}_\alpha P^c_{\alpha\alpha} = \sum_\beta \mathcal{P}^c_{\alpha\beta},$$

$$\mathcal{P}^c \equiv \sum_\alpha \bar{n}_\alpha P^c_{\alpha\alpha} = \sum_{\alpha\beta}\mathcal{P}^c_{\alpha\beta}. \tag{84}$$

With these definitions we find that

$$\mathcal{T} = \mathcal{P}^{\rm 2h} + \mathcal{P}^{\rm 1h} + \mathcal{P}^{\rm 2h}\mathcal{P}^{\rm 1h} - \mathcal{P}^c. \tag{85}$$

Similarly, we get:

$$Y_{\alpha\beta} = P^c_{\alpha\beta} + P^c_{\beta\alpha} - P^{\rm 2h}_{\alpha\beta}\,\mathcal{P}^{\rm 1h} - P^{\rm 1h}_{\alpha\beta}\,\mathcal{P}^{\rm 2h}. \tag{86}$$

The Fisher matrix was defined in a generic sense in equation (18). That definition, as well as the construction of the optimal quadratic estimators, are valid for any Gaussian variables (Tegmark et al. 1998). In a related result, Smith & Marian (2015) recently derived an optimal estimator for the matter power spectrum, as well as the Fisher matrix for the power spectrum, including not only the one-halo term, but also the two- and three-halo contributions to the trispectrum – most of which are, strictly speaking, non-Gaussian contributions. We did include the one-halo term in the pixel covariance, as well as in the trispectrum, but only through the assumption of Gaussianity of the four-point function. Due to the non-Gaussian terms that will appear in the trispectrum, our estimators are not exactly optimal. Nevertheless, in some sense our result are more general than those of Smith & Marian (2015), since the multitracer estimators can be employed not only in the computation of the matter power spectrum, but also for the biases and the RSDs.

Since we keep the assumption of Gaussianity, all we have to do is work out the algebra with the covariance of equation (77), and its inverse, given by equation (79). After a lengthy calculation, we find that the Fisher matrix which generalizes the expression in the integrand of equation (25) can be expressed as

$$\begin{aligned}\mathcal{F}^{2\mathrm{h}}_{\mu\nu} = \frac{1}{4(1+\mathcal{T})^2}\Big\{ &\left[(1+\mathcal{P}^{1\mathrm{h}})\mathcal{P}^{2\mathrm{h}} - \mathcal{P}^{c}\right] \\ &\times \left[\delta_{\mu\nu}\mathcal{P}^{2\mathrm{h}}_{\mu}(1+\mathcal{T}) - (1+\mathcal{P}^{1\mathrm{h}})\mathcal{P}^{2\mathrm{h}}_{\mu}\mathcal{P}^{2\mathrm{h}}_{\nu}\right. \\ &\left. + (1+\mathcal{P}^{2\mathrm{h}})\mathcal{P}^{c}_{\mu\nu} + \mathcal{P}^{2\mathrm{h}}_{\mu}\mathcal{P}^{c}_{\nu} + \mathcal{P}^{2\mathrm{h}}_{\nu}\mathcal{P}^{c}_{\mu}\right] \\ &+ (1+\mathcal{P}^{1\mathrm{h}})^2\,\mathcal{P}^{2\mathrm{h}}_{\mu}\mathcal{P}^{2\mathrm{h}}_{\nu} - (\mathcal{P}^{c})^2\mathcal{P}^{c}_{\mu\nu} \\ &- (1+\mathcal{P}^{1\mathrm{h}})\left(\mathcal{P}^{2\mathrm{h}}_{\mu}\mathcal{P}^{c}_{\nu} + \mathcal{P}^{2\mathrm{h}}_{\nu}\mathcal{P}^{c}_{\mu}\right)\Big\}\,. \end{aligned} \tag{87}$$

It can be easily verified that taking $P^{1\mathrm{h}}_{\mu\nu} \to 0$ implies $\mathcal{P}^{c}_{\mu\nu} \to 0$, $\mathcal{T} \to \mathcal{P}^{2\mathrm{h}}$, etc. and this expression reduces to the matrix $\mathcal{F}_{\mu\nu}$ which is inside the integral in equation (25). The matrix is also manifestly symmetric.

It can also be shown that this Fisher matrix is positive definite, with positive diagonal terms and a positive determinant. This guarantees that the covariance of the two-halo power spectrum is also positive definite.

We should stress once again that the expression above is only valid in the approximation that the one-halo term is separable, i.e. $P^{1\mathrm{h}}_{\alpha\beta}/\sqrt{P^{1\mathrm{h}}_{\alpha\alpha}P^{1\mathrm{h}}_{\beta\beta}} \simeq \delta_{\alpha\beta}$. For a general form of the one-halo term, the pixel covariance matrix cannot be inverted analytically, which means that there are no closed-form expressions for the Fisher matrix or for the optimal weights. One could still go ahead and compute them numerically, without any difficulty.

6.2 Fisher matrix and optimal weights for the one-halo term

Now, suppose that what we are in fact interested in measuring the one-halo term. The one-halo term is now the 'signal', while the two-halo term, as well as shot noise, become the 'noise'. This should in fact be the case for very small scales ($k \gg 1\,h\,\mathrm{Mpc}^{-1}$), where the one-halo term dominates over the two-halo term (Cooray & Sheth 2002).

The pixel covariance is still the same, as in equation (77), and, as long as the one-halo term is separable, the inverse covariance is also unaltered – see equation (79). The basic difference is that now instead of writing $P^{2\mathrm{h}}_{\alpha\beta} = B_\alpha(z,k,\mu_k)B_\beta(z,k,\mu_k)P_{\mathrm{m}}(k)$, we assume that the one-halo term can be written effectively as something like $P^{1\mathrm{h}}_{\alpha\beta} = H_\alpha(z,k,\mu_k)H_\beta(z,k,\mu_k)U(k)$, where $U(k)$ contains information about the shape of the mean halo profile. This is a strong assumption: it means that, for N species of tracers, the one-halo term would have only N degrees of freedom (the $P^{1\mathrm{h}}_{\alpha} = H^2_\alpha\,U$), while the full expression in fact has $N(N-1)/2$ degrees of freedom.

Keeping the hypothesis that the one-halo term is separable, all we need to do now, in order to find its Fisher matrix, is to exchange all the two-halo terms by the one-halo terms in equation (87). This procedure can also be used to define the optimal weights that ought to be used when extracting information about the one-halo term from galaxy surveys.

Since many of the objects defined above are already symmetric under the exchange $P^{2\mathrm{h}}_{\mu\nu} \leftrightarrow P^{1\mathrm{h}}_{\mu\nu}$ (this includes $\mathcal{T}$, $Y_{\mu\nu}$ and $\mathcal{P}^{c}_{\mu\nu}$), the one-halo Fisher matrix can be immediately written as

$$\begin{aligned}\mathcal{F}^{1\mathrm{h}}_{\mu\nu} = \frac{1}{4(1+\mathcal{T})^2}\Big\{ &\left[(1+\mathcal{P}^{2\mathrm{h}})\mathcal{P}^{1\mathrm{h}} - \mathcal{P}^{c}\right] \\ &\times \left[\delta_{\mu\nu}\mathcal{P}^{1\mathrm{h}}_{\mu}(1+\mathcal{T}) - (1+\mathcal{P}^{2\mathrm{h}})\mathcal{P}^{1\mathrm{h}}_{\mu}\mathcal{P}^{1\mathrm{h}}_{\nu}\right. \\ &\left. + (1+\mathcal{P}^{1\mathrm{h}})\mathcal{P}^{c}_{\mu\nu} + \mathcal{P}^{1\mathrm{h}}_{\mu}\mathcal{P}^{c}_{\nu} + \mathcal{P}^{1\mathrm{h}}_{\nu}\mathcal{P}^{c}_{\mu}\right] \\ &+ (1+\mathcal{P}^{2\mathrm{h}})^2\,\mathcal{P}^{1\mathrm{h}}_{\mu}\mathcal{P}^{1\mathrm{h}}_{\nu} - (\mathcal{P}^{c})^2\mathcal{P}^{c}_{\mu\nu} \\ &- (1+\mathcal{P}^{2\mathrm{h}})\left(\mathcal{P}^{1\mathrm{h}}_{\mu}\mathcal{P}^{c}_{\nu} + \mathcal{P}^{1\mathrm{h}}_{\nu}\mathcal{P}^{c}_{\mu}\right)\Big\}\,. \end{aligned} \tag{88}$$

For very small scales the two-halo term can be neglected, and we are left just with the one-halo terms.

The Fisher matrix in bins of k is just as in equation (25):

$$F^{ij(1\mathrm{h})}_{\mu\nu} = \frac{\delta_{ij}}{[P^{i(1\mathrm{h})}_{\mu\nu}]^2}\int_{v_i}\frac{\mathrm{d}^3x\,\mathrm{d}^3k}{(2\pi)^3}\mathcal{F}^{1\mathrm{h}}_{\mu\nu}. \tag{89}$$

The optimal weights follow in a straightforward manner from this expression, just as was done for the two-halo term.

6.3 Joint Fisher matrix for the two-halo and one-halo terms

The next obvious question is: what if we wish to estimate *both* the two-halo and the one-halo terms in a multitracer cosmological survey, simultaneously? The two contributions are clearly correlated, so their information contents are not independent. Evidently, on either very large or very small scales the correlations between the two are small, and one can treat the signal ($P^{2\mathrm{h}}$ on large scales; $P^{1\mathrm{h}}$ on small scales) as effectively independent of the noise. However, on intermediate scales (around $k \sim 1\,h\,\mathrm{Mpc}^{-1}$) the one-halo and the two-halo terms may have significant correlations. Furthermore, the approximation of separable one-halo term becomes more accurate on those intermediate scales.

The pixel covariance is still given by equation (77), and its inverse also remains unchanged, but we would now be considering our 'signal' as the sum $P_\mu = P^{2\mathrm{h}}_{\mu} + P^{1\mathrm{h}}_{\mu}$. The main difference is that the derivatives of the pixel covariance, which in the case when we neglected the one-halo term were given by equation (24), should now be computed with respect to the signal including the one-halo term:

$$\begin{aligned}\frac{\partial C_{\alpha\beta}(\boldsymbol{x},\boldsymbol{x}')}{\partial P^{i}_{\mu}} &= \frac{\partial P^{i(2\mathrm{h})}_{\mu}}{\partial P^{i}_{\mu}}\frac{\partial C_{\alpha\beta}(\boldsymbol{x},\boldsymbol{x}')}{\partial P^{i(2\mathrm{h})}_{\mu}} + \frac{\partial P^{i(1\mathrm{h})}_{\mu}}{\partial P^{i}_{\mu}}\frac{\partial C_{\alpha\beta}(\boldsymbol{x},\boldsymbol{x}')}{\partial P^{i(1\mathrm{h})}_{\mu}} \\ &= \int\frac{\mathrm{d}^3k}{(2\pi)^3}\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{x}')}\left(\delta_{\alpha\mu}\delta^{i}_{\boldsymbol{x},\boldsymbol{k}} + \delta_{\beta\mu}\delta^{i}_{\boldsymbol{x}',\boldsymbol{k}}\right) \\ &\quad\times\left[\frac{B_\alpha(\boldsymbol{x},\boldsymbol{k})B_\beta(\boldsymbol{x}',\boldsymbol{k})}{2\,(B^{i}_{\mu})^2}\right. \\ &\quad\left. + \frac{H_\alpha(\boldsymbol{x},\boldsymbol{k})H_\beta(\boldsymbol{x}',\boldsymbol{k})}{2\,(H^{i}_{\mu})^2}\right]. \end{aligned} \tag{90}$$

Substitution of this expression, together with the inverse of the pixel covariance, into equation (18), leads to the full Fisher matrix for the power spectrum. This can be written as

$$F^{ij}_{\mu\nu} = \delta_{ij} \int_{v_i} \frac{\mathrm{d}^3x\,\mathrm{d}^3k}{(2\pi)^3} \left\{ \frac{\mathcal{F}^{2\mathrm{h}}_{\mu\nu}}{\left[P^{i(2\mathrm{h})}_{\mu\nu}\right]^2} + \frac{\mathcal{F}^{1\mathrm{h}}_{\mu\nu}}{\left[P^{i(1\mathrm{h})}_{\mu\nu}\right]^2} + \mathcal{F}^{c}_{\mu\nu} \left[\frac{1}{P^{i(2\mathrm{h})}_{\mu\mu} P^{i(1\mathrm{h})}_{\nu\nu}} + \frac{1}{P^{i(2\mathrm{h})}_{\nu\nu} P^{i(1\mathrm{h})}_{\mu\mu}} \right] \right\}, \quad (91)$$

where $\mathcal{F}^{c}_{\alpha\beta}$ contains the cross-terms between the one-halo and the two-halo terms which follow from equation (90). Expressing the inverse covariance of equation (79) in terms of $C^{-1}_{\alpha\beta}(\boldsymbol{x}, \boldsymbol{x}') = \delta_{\mathrm{D}}(\boldsymbol{x}, \boldsymbol{x}')D_{\alpha\beta}$, the information mixing between the one-halo and two-halo terms is given by

$$\mathcal{F}^{c}_{\mu\nu} = \frac{1}{8} \sum_{\alpha\beta} \left(P^{1\mathrm{h}}_{\alpha\beta} D_{\alpha\beta}\, P^{2\mathrm{h}}_{\mu\nu} D_{\mu\nu} + P^{2\mathrm{h}}_{\alpha\beta} D_{\alpha\beta}\, P^{1\mathrm{h}}_{\mu\nu} D_{\mu\nu} + P^{2\mathrm{h}}_{\nu\alpha} D_{\alpha\mu} P^{1\mathrm{h}}_{\mu\beta} D_{\beta\nu} + P^{1\mathrm{h}}_{\nu\alpha} D_{\alpha\mu} P^{2\mathrm{h}}_{\mu\beta} D_{\beta\nu} \right). \quad (92)$$

It is trivial to obtain the full expression, although it turns out to be rather long. It can be significantly simplified if we employ two additional auxiliary definitions:

$$Z^{(1,2)\mathrm{h}}_{\mu\nu} \equiv \sum_{\alpha} P^{(1,2)\mathrm{h}}_{\mu\alpha} D_{\alpha\nu}, \quad (93)$$

$$W^{(1,2)\mathrm{h}}_{\mu\nu} \equiv P^{(1,2)\mathrm{h}}_{\mu\nu} D_{\mu\nu}. \quad (94)$$

In terms of these variables we have, e.g.:

$$Z^{2\mathrm{h}}_{\mu\nu} = \frac{\bar{n}_\nu}{1+\mathcal{T}} \left[\left(1 + 2\mathcal{P}^{2\mathrm{h}} + \mathcal{P}^{1\mathrm{h}}\right) P^{2\mathrm{h}}_{\mu\nu} + P^{c}_{\mu\nu} \right], \quad (95)$$

and

$$W^{2\mathrm{h}}_{\mu\nu} = \delta_{\mu\nu}\mathcal{P}^{2\mathrm{h}}_{\mu} + \frac{\left(1 - \mathcal{P}^{1\mathrm{h}}\right)\mathcal{P}^{2\mathrm{h}}_{\mu}\mathcal{P}^{2\mathrm{h}}_{\nu}}{1+\mathcal{T}} + \frac{\mathcal{P}^{2\mathrm{h}}_{\mu}\mathcal{P}^{c}_{\nu} + \mathcal{P}^{2\mathrm{h}}_{\nu}\mathcal{P}^{c}_{\mu} + \left(1 - \mathcal{P}^{2\mathrm{h}}\right)\mathcal{P}^{c}_{\mu\nu}}{1+\mathcal{T}}, \quad (96)$$

as well as the analogous expressions obtained by exchanging $1h \leftrightarrow 2h$. In terms of these definitions we have:

$$\mathcal{F}^{c}_{\mu\nu} = \frac{1}{8} \left[Z^{1\mathrm{h}} W^{2\mathrm{h}}_{\mu\nu} + Z^{2\mathrm{h}} W^{1\mathrm{h}}_{\mu\nu} + \sum_{\alpha} \left(Z^{2\mathrm{h}}_{\mu\alpha} Z^{1\mathrm{h}}_{\alpha\nu} + Z^{1\mathrm{h}}_{\mu\alpha} Z^{2\mathrm{h}}_{\alpha\nu} \right) \right],$$

where $Z = \sum_\mu Z_{\mu\mu} = \sum_{\mu\nu} W_{\mu\nu}$. In fact, these definitions are also helpful when computing $\mathcal{F}^{2\mathrm{h}}_{\mu\nu}$ (taking all $Z \to Z^{2\mathrm{h}}$ and $W \to W^{2\mathrm{h}}$) and $\mathcal{F}^{1\mathrm{h}}_{\mu\nu}$ (taking all $Z \to Z^{1\mathrm{h}}$ and $W \to W^{1\mathrm{h}}$).

7 CONCLUSIONS

We have obtained optimal estimators for the Fourier analysis of multitracer cosmological surveys. The formulas were derived in Section 3, and a practical algorithm for the Fourier analysis of multitracer surveys was summarized in Section 5.2. Those are the main results of this paper.

The multitracer technique estimates the individual redshift-space power spectra for each tracer, $P_\alpha(z, k, \mu_k)$, taking into account the covariance between the tracers which is induced by the large-scale structure. In contrast to the estimators obtained by Percival et al. (2004) or Smith & Marian (2015), which are suited for estimating the underlying matter power spectrum *after* fixing the biases and the RSDs, our optimal estimators can be used to measure both the power spectrum, the biases, the shape of RSDs, etc. In particular, our estimators facilitate measurements of RSDs, scale-dependent bias and NGs from cosmological surveys of multiple tracers, helping realize the potential for determining those physical parameters to an accuracy which is not limited by cosmic variance (Seljak 2009; McDonald & Seljak 2009; Gil-Marín et al. 2010; Hamaus et al. 2011; Abramo & Leonard 2013).

We also included the contribution from the one-halo term in our calculations (Section 6). Although on very large scales ($k \ll 1\,h\,\mathrm{Mpc}^{-1}$) the two-halo term is dominant, the one-halo term gives a nearly constant contribution in that limit, adding to shot noise – and, unlike shot noise, it does affect the cross-correlations.

It is important to stress that our formulas are relatively simple generalizations of those by FKP (Feldman et al. 1994) and PVP (Percival et al. 2004), so readers familiar with these standard methods should have no trouble implementing the multitracer technique. We tested the estimators (see Section 5) in a wide variety of situations, and they performed quite robustly – in many instances, significantly better than the FKP method. It should now be straightforward to combine cosmological surveys targeting different types of galaxies, quasars and other tracers of large-scale structure.

ACKNOWLEDGEMENTS

We would like to thank Miguel Quartín for useful comments. We thank FAPESP, CNPq, and the University of São Paulo's *NAP LabCosmos* for financial support.

REFERENCES

Abbott T. et al., 2005, preprint (arXiv:astro-ph/0510346)
Abell P. et al., 2009, preprint (arXiv:0912.0201)
Abramo L. R., 2012, MNRAS, 420, 3
Abramo L. R., Leonard K. E., 2013, MNRAS, 432, 318
Abramo L. R. et al., 2012, MNRAS, 423, 335
Adelman-McCarthy J. K. et al., 2008, ApJS, 175, 297
Anderson L. et al., 2012, MNRAS, 427, 3435
Anderson L. et al., 2014, MNRAS, 439, 83
Bandura K. et al., 2014, Proc. SPIE, 9145, 914522
Bartolo N., Komatsu E., Matarrese S., Riotto A., 2004, Phys. Rep., 402, 103
Battye R. et al., 2013, MNRAS, 434, 1239
Benítez N. et al., 2009, ApJ, 691, 241
Benítez N. et al., 2014, preprint (arXiv:1403.5237)
Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000, MNRAS, 311, 793
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Blake C., Glazebrook K., 2003 ApJ, 594, 665
Blake C. et al., 2011, MNRAS, 415, 2876
Blake C. et al., 2013, MNRAS, 436, 3089
Bonoli S., Pen U.-L., 2008, MNRAS, 396, 1610
Bull P., Ferreira P., Patel P., Santos M. G., 2015, ApJ, 803, 21
Cai Y.-C., Bernstein G. M., 2011a, MNRAS, 416, 3009
Cai Y.-C., Bernstein G. M., 2011b, MNRAS, 422, 1045
Cai Y.-C., Bernstein G. M., Sheth R. K., 2011, MNRAS, 412, 995
Cole S. et al., 2005, MNRAS, 362, 505
Coles P., Jones B., 1991, MNRAS, 248, 1
Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Croom S. et al., 2005, MNRAS, 356, 415
da Ângela J. et al., 2005, MNRAS, 360, 1040
Dalal N., Doré O., Huterer D., Shirokov A., 2008, Phys. Rev. D, 77, 123514

Dawson K. et al., 2012, AJ, 145, 10
Dekel A., Lahav O., 1999, ApJ, 520, 24
Eisenstein D. J., Hu W., Tegmark M., 1999, ApJ, 518, 2
Ellis R. et al., 2012, preprint (arXiv:1206.0737)
Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23
Ferramacho L. D., Santos M. G., Jarvis M. J., Camera S., 2014, MNRAS, 442, 2511
Fisher K. B., Davis M., Strauss M. A., Yahil A., Huchra J. P., 1993, ApJ, 402, 42
Gil-Marín H., Wagner C., Verde L., Jimenez R., Heavens A., 2010, MNRAS, 407, 772
Guzzo L., Strauss M. A., Fisher K. B., Giovanelli R., Haynes M. P., 1997, ApJ, 489, 37
Hamaus N., Seljak U., Desjacques V., Smith R., Baldauf T., 2010, Phys. Rev. D, 82, 043515
Hamaus N., Seljak U., Desjacques V., 2011, Phys. Rev. D, 84, 083509
Hamaus N., Seljak U., Desjacques V., 2012, Phys. Rev. D, 86, 103513
Hamilton A. J. S., 2005a, Lect. Notes Phys., 665, 415, preprint (astro-ph/0503603)
Hamilton A. J. S., 2005b, Lect. Notes Phys., 665, 433, preprint (astro-ph/0503604)
Hill G. J. et al., 2008, in Kodama T., Yamada T., Aoki K., ASP Conf. Ser. Vol. 399, Panoramic Views of the Universe. Astron. Soc. Pac., San Francisco, p.115
Kaiser N., 1987, MNRAS, 227, 1
Leistedt B., Peiris H. V., 2014, MNRAS, 444, 2
Leistedt B., Peiris H. V., Mortlock D. J., Benoit-Lévy A., Pontzen A., 2013, MNRAS, 435, 1857
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Linder E., 2005, Phys. Rev. D, 72, 043529
McDonald P., Seljak U., 2009, J. Cosmol. Astropart. Phys., 910, 007
Martinez V., Saar E., 2001, Statistics of the Galaxy Distribution. CRC Press, Boca Raton, FL
Percival W. J., Verde L., Peacock J. A., 2004, MNRAS, 347, 645
Raccanelli A. et al., 2012, MNRAS, 436, 89
Ross A. et al., 2009, ApJ, 697, 1634
Sawangwit A. et al., 2012, MNRAS, 420, 1916
Schlegel D. J. et al., 2009, preprint (arXiv:0904.0468)
Scoville N. et al., 2007, ApJS, 172, 1
Sefusatti E., Komatsu E., 2007, Phys. Rev. D, 76, 083004
Seljak U., 2009, Phys. Rev. Lett., 102, 021302
Seljak U., Hamaus N., Desjacques V., 2009, Phys. Rev. Lett., 103, 091303
Seo H.-J., Eisenstein D. J., 2003, ApJ, 598, 720
Shen Y. et al., 2007, AJ, 133, 2222
Slosar A., 2009, J. Cosmol. Astropart. Phys., 003, 004
Slosar A. et al., 2013, J. Cosmol. Astropart. Phys., 004, 026
Smith R. E., 2009, MNRAS, 400, 851
Smith R. E., Marian L., 2014, preprint (arXiv:1406.1800)
Smith R. E., Marian L., 2015, MNRAS, 454, 1266
Smith R. E., Scoccimarro R., Sheth R. K., 2007, Phys. Rev. D, 75, 063512
Smith R. E., Desjacques V., Marian L., 2011, Phys. Rev. D, 83, 043526
Swanson M., Tegmark M., Blanton M., Zehavi I., 2008, MNRAS, 385, 1635
Tegmark M., 1997, Phys. Rev. Lett., 79, 3806
Tegmark M., Taylor A. N., Heavens A. F., 1997, ApJ, 480, 22
Tegmark M., Hamilton A. J. S., Strauss M. A., Vogeley M. S., Szalay A. S., 1998, ApJ, 499, 555
Tonry J. L. et al., 2012, ApJ, 750, 99
Verde L., Wang L., Heavens A. F., Kamionkowski M., 2000, MNRAS, 313, 141
White M., Song Y.-S., Percival W. J., 2008, MNRAS, 397, 1348
Woodbury M. A., 1950, Memorandum Rept. 42, Statistical Research Group. Princeton Univ., Princeton, NJ
Yahata K. et al., 2005, PASJ, 57, 529
York D. G. et al., 2000, AJ, 120, 1579
Zheng Z. et al., 2005, ApJ, 633, 791

APPENDIX A: DEGENERATE TRACERS IN THE BASIS OF THE AUTO-POWER SPECTRA

Suppose we have N types of tracers, but we would like to combine the last two of those tracers into a *single* species, for a new total of $N-1$ tracers. For simplicity, let us regard our original parameters as $\mathcal{P}_\mu$ ($\mu = 1 \ldots N$). We would like to change variables to $\mathcal{P}'_a$ ($a = 1 \ldots N-1$), where $\mathcal{P}'_a = \mathcal{P}_a$ for $a = 1, \ldots N-2$, and the new tracer species is constructed by combining the last two tracers, $\mathcal{P}'_{N-1} = \mathcal{P}_{N-1} + \mathcal{P}_N$. When the biases of the two species which are combined into one are identical, $B_{N-1} = B_{N-2} = B'_{N-1}$, this linear combination ensures that $\mathcal{P}'_{N-1} = (\bar{n}_{N-1} + \bar{n}_N) B'^2_{N-1} P_\mathrm{m}$, so clearly $\bar{n}'_{N-1} = \bar{n}_{N-1} + \bar{n}_N$ – i.e. the total number of galaxies of the new species is the sum of the number of galaxies of the two original species.

The Jacobian for the transformation $\mathcal{P}'_a \to \mathcal{P}_\mu$ is $\partial \mathcal{P}'_a / \partial \mathcal{P}_\mu$, and this Jacobian equals $\delta_{a\mu}$ when $\mu < N-1$, it vanishes if $a = N-1$ and $\mu < N-1$, and it is equal to 1 if $a = N-1$ and $\mu \geq N-1$.

However, what we need for the new Fisher matrix is the Jacobian for the inverse transformation,[7] $\mathcal{P}_\mu \to \mathcal{P}'_a$, i.e. $J_{\mu a} = (\partial \mathcal{P}_\mu / \partial \mathcal{P}'_a)^{-1}$. But this turns out to be a very simple matrix: $J_{\mu a} = \delta_{\mu a}$ when $\mu < N-1$, it vanishes if $a = N-1$ and $\mu < N-1$, and when $a = N-1$ and $\mu \geq N-1$ the Jacobian is equal to $\mathcal{P}_\mu / \mathcal{P}'_{N-1} = \mathcal{P}_\mu / (\mathcal{P}_{N-1} + \mathcal{P}_{N-1})$ – see also the discussion at the beginning of Section 4.

Hence, in the new variables the Fisher matrix (or, more precisely, the Fisher information density per unit of phase space volume) is:

$$F'_{ab} = F[\mathcal{P}'_a, \mathcal{P}'_b] = \sum_{\mu\nu}^{N} J_{\mu a} \frac{\mathcal{F}_{\mu\nu}}{\mathcal{P}_\mu \mathcal{P}_\nu} J_{\nu b}, \tag{A1}$$

where $\mathcal{F}_{\mu\nu} = F[\log \mathcal{P}_\mu, \log \mathcal{P}_\nu]$ – see equation (26). This turns out to be given by

$$F'_{ab} = \left(\begin{array}{c|c} \frac{1}{\mathcal{P}'_a \mathcal{P}'_b} \mathcal{F}_{ab} & \frac{1}{\mathcal{P}'_a \mathcal{P}'_{N-1}} \sum_{\nu=N-1}^{N} \mathcal{F}_{a\nu} \\ \hline \mathrm{Sym} & \frac{1}{(\mathcal{P}'_{N-1})^2} \sum_{\mu,\nu=N-1}^{N} \mathcal{F}_{\mu\nu} \end{array} \right), \tag{A2}$$

where the upper left block is an $(N-2) \times (N-2)$ matrix, the right block is an $1 \times (N-2)$ column, the lower left block is an $(N-2) \times 1$ row, and the lower right block is a single entry. Hence, the resulting $(N-1)$-dimensional Fisher matrix is given simply by summing the lines and columns corresponding to the two tracers which were combined into a single type.

Now, it can be easily verified from equation (26) that summing any two lines and columns of the fisher matrix $\mathcal{F}_{\mu\nu}$ yields precisely the Fisher matrix where the new entries correspond to the Fisher information for the *sum* of the clustering strengths of the two species that were combined. In other words, if we take equation (26) and use $\mathcal{P}'_a$ to compute $\mathcal{F}'_{ab} = F[\log \mathcal{P}'_a, \log \mathcal{P}'_b]$, then the Fisher matrix $F'_{ab} = \mathcal{F}'_{ab} / \mathcal{P}'_a \mathcal{P}'_b$ is identical to equation (A2).

This argument can be iterated to show that combining any number of tracers into a single species corresponds to adding their clustering strengths, and this operation results in a simple sum of the Fisher information of those tracers.

[7] In fact, since this Jacobian is not a square matrix, it only has a pseudo-inverse. However, in this case the pseudo-inverse is exact.

APPENDIX B: INVERSION OF THE COVARIANCE MATRIX

Consider a matrix of the form $M_v = \mathbb{1} + v \otimes v$ – i.e. $M_{v,\mu\nu} = \delta_{\mu\nu} + v_\mu v_\nu$. As discussed in Section 2, it can be shown that

$$M_v^{-1} = \mathbb{1} - \frac{v \otimes v}{1 + v^2}, \tag{B1}$$

where $v^2 = \mathrm{Tr}(v \otimes v) = \sum_\alpha v_\alpha v_\alpha$. This is in fact a special case of the Sherman–Morrison–Woodbury formula (Woodbury 1950).

The matrix M_v also has a simple 'square root', as well as an 'inverse square root', given by

$$M_v^{1/2} = \mathbb{1} + \frac{v \otimes v}{1 + \sqrt{1 + v^2}} \tag{B2}$$

$$M_v^{-1/2} = \mathbb{1} - \frac{v \otimes v}{1 + v^2 + \sqrt{1 + v^2}}, \tag{B3}$$

where $M_v^{-1/2} \cdot M_v^{-1/2} = M_v^{-1}$ and $M_v^{-1/2} \cdot M_v = M_v^{1/2}$, from which follows that

$$M_v^{-1/2} \cdot M_v \cdot M_v^{-1/2} = \mathbb{1}. \tag{B4}$$

Now, take a matrix $M = M_v + u \otimes u$. The first piece of that matrix can be diagonalized following the procedure outlined above, so we have that

$$\begin{aligned} M_v^{-1/2} \cdot M \cdot M_v^{-1/2} &= \mathbb{1} + (M_v^{-1/2} \cdot u) \otimes (u \cdot M_v^{-1/2}) \\ &= \mathbb{1} + u' \otimes u', \end{aligned} \tag{B5}$$

where $u' = M_v^{-1/2} \cdot u$ (i.e. $u'_\alpha = \sum_\mu M_{v,\alpha\mu}^{-1/2} u_\mu$). But the matrix of equation (B5) can now be inverted using the equivalent of equation (B1), and moreover it has an inverse square root $M_{u'}^{-1/2}$, as in equation (B3). Therefore, we have that

$$M_{u'}^{-1/2} \cdot M_v^{-1/2} \cdot M \cdot M_v^{-1/2} \cdot M_{u'}^{-1/2} = \mathbb{1}.$$

Therefore, the inverse of the matrix M is given by

$$\begin{aligned} M^{-1} &= M_v^{-1/2} \cdot M_{u'}^{-1/2} \cdot M_{u'}^{-1/2} \cdot M_v^{-1/2} \\ &= M_v^{-1/2} \cdot M_{u'}^{-1} \cdot M_v^{-1/2}. \end{aligned} \tag{B6}$$

Of course, one could equally write this inverse as

$$M^{-1} = M_u^{-1/2} \cdot M_{v'}^{-1} \cdot M_u^{-1/2},$$

where $v' = M_u^{-1/2} \cdot v$.

This paper has been typeset from a TEX/LATEX file prepared by the author.