Reducing sample variance: halo biasing, non-linearity and stochasticity

Héctor Gil-Marín,[1,2,3]⋆ Christian Wagner,[2] Licia Verde,[2,4,5] Raul Jimenez[2,4,5] and Alan F. Heavens[3]

[1]*Institute of Space Sciences (IEEC-CSIC), Facultat de Ciències, Campus UAB, Bellaterra 08193, Spain*
[2]*Institut de Ciències del Cosmos(ICC), Universitat de Barcelona-IEEC, Martí Franquès 1, 08028 Barcelona, Spain*
[3]*Scottish Universities Physics Alliance (SUPA), Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ*
[4]*ICREA (Institució Catalana de Recerca i Estudis Avançats)*
[5]*Institute for the Physics and Mathematics of the Universe (IPMU), the University of Tokyo, Kashiwa, Chiba 277-8568, Japan*

ABSTRACT

Comparing clustering of differently biased tracers of the dark matter distribution offers the opportunity to reduce the sample or cosmic variance error in the measurement of certain cosmological parameters. We develop a formalism that includes bias non-linearities and stochasticity. Our formalism is general enough that it can be used to optimize survey design and tracers selection and optimally split (or combine) tracers to minimize the error on the cosmologically interesting quantities. Our approach generalizes the one presented by McDonald & Seljak of circumventing sample variance in the measurement of $f \equiv \mathrm{d}\ln D/\mathrm{d}\ln a$. We analyse how the bias, the noise, the non-linearity and stochasticity affect the measurements of Df and explore in which signal-to-noise regime it is significantly advantageous to split a galaxy sample in two differently biased tracers. We use N-body simulations to find realistic values for the parameters describing the bias properties of dark matter haloes of different masses and their number density. We find that, even if dark matter haloes could be used as tracers and selected in an idealized way, for realistic haloes, the sample variance limit can be reduced only by up to a factor $\sigma_{2\mathrm{tr}}/\sigma_{1\mathrm{tr}} \simeq 0.6$. This would still correspond to the gain from a three times larger survey volume if the two tracers were not to be split. Before any practical application one should bear in mind that these findings apply to dark matter haloes as tracers, while realistic surveys would select galaxies: the galaxy–host halo relation is likely to introduce extra stochasticity, which may reduce the gain further.

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

1 INTRODUCTION

One of the active topics of current research is the formation and growth of large-scale structure in the Universe. Knowledge of the physical origin of the growth of structure will allow us to know about the origin of dark matter and also provides a useful way to discriminate between different theories for the origin and evolution of dark energy. In particular, comparing and combining measurements of the Universe expansion history (as given by e.g. Baryon Acoustic Oscillations, cosmic chronometers and Supernovae) with measurements of the linear growth of structure can provide a tool to test whether dark energy is an extra component with negative pressure or a manifestation of the breakdown of general relativity on large scales. To this end, usually, the goal is to measure the $f(z)$ parameter defined as $f(z) \equiv \mathrm{d}\ln D(z)/\mathrm{d}\ln a(z)$, where $D(z)$ is the linear growth factor and $a(z)$ the scalefactor.

The two main approaches to measure the growth of structure are weak gravitational lensing (e.g. Hoekstra et al. 2002; Bacon et al. 2005) and galaxy clustering, which is the technique we consider here. Galaxy clustering is a relatively simple, high signal-to-noise ratio (S/N) measurement: the angular position of galaxies can be measured using photometry and the radial position using spectroscopy. With this information the three-dimensional power spectrum of galaxies can be computed as a function of redshift. However, at the two-point level, the galaxy field can only trace the dark matter field up to a bias factor b, which may depend on scale and redshift. Thus, only once the bias

⋆E-mail: hectorgil@icc.ub.edu

is known, the galaxy power spectrum P_{gg} can be related to that of the dark matter P_{mm} and thus yield the growth of structure. The main drawback associated to this technique is that the value of this bias cannot readily be predicted from theoretical models of galaxy formation.

There are several observational techniques to measure the bias parameter (Fry 1994; Feldman et al. 2001; Verde et al. 2002; Hawkins et al. 2003; Seljak et al. 2005; Guzzo et al. 2008). In this work we use the approach that takes advantage of the redshift-space distortions. Peculiar velocities of dark matter tracers are set by the gravitational field: using the measured redshift as a distance indicator distorts clustering, enhancing it along the line of sight and the redshift-space distortion parameter is $\beta \equiv f/b$ (Davis & Peebles 1983; Kaiser 1987; Hamilton 1998). Using measurements in different directions (different Fourier modes), one can compute β. In combination with the galaxy power spectrum measurement, this approach yields the divergence power spectrum $P_{\theta\theta} = f^2 P_{mm}$, which can be directly compared with theory predictions and encloses the desired dependence on the growth of structure.

There are two sources of errors in the measurement of the galaxy power spectrum: the shot noise and the sample variance (or cosmic variance). On one hand, the shot noise is due to the fact that we use a discrete set of objects to characterize the matter field. If this noise is Poisson, it is scale-independent and equal to the reciprocal of the number density of objects (Peebles 1980). On the other hand, the sample variance effect is due to the fact that the matter field has its origin in a random realization of the underlying cosmology. In a finite survey volume there are only a finite number of modes present, especially on large scales we only have a few modes to perform the averaging. Thus, the total error on the power spectrum P at a given scale k, is $\sigma_P/P = (2/N)^{1/2}(1 + \sigma_n/P)$. Here, N is the number of modes measured (at the scale given by k) and σ_n is the shot noise contribution. We see that just reducing the shot noise (increasing the number density of objects) does not help, as there is a natural limitation on our capacity to measure P [and consequently $f(z)$], due to sample variance.

In order to reduce this limitation, a multi-tracer technique has been advocated recently (Seljak 2009). It is based on the usage of two differently biased tracers of the dark matter field. With this method the sample variance limit can be reduced. The effectiveness of this method depends on a number of factors: the ratio of these different biases; the S/N regime and on the non-linearity of the biases. With the exception of gravitational lensing, one cannot see the dark matter directly nor the dark matter haloes, so in most practical applications, tracers need to be used such as galaxies, quasars, clusters or 21-cm emission.

The goal of this paper is to study the possibility of measuring the parameter $x(z) \equiv f^2(z)D^2(z)$ using the single- and the multi-tracer formalism and see whether the reduction of the sample variance is significant. Here, we present a new formalism of how to estimate the error on x using the multi-tracer formalism, taking into account that the bias may be scale-dependent, non-linear and stochastic. This formalism may be useful for galaxy surveys, because it has been observed that the galaxy biasing is significantly non-linear and stochastic. N-body simulations and theoretical models allow us to estimate which are the bias characteristics for dark matter tracers and therefore which precision can be reached with this model.

In Section 2, we begin by introducing the formalism of our method and analyse how the different parameters affect the reduction of the sample variance effect. In Section 3 we use both analytical approximations and N-body simulations to obtain physically motivated parameters for our model and compute realistic expected errors using the single- and multi-tracer formalism. In Section 4 we conclude with a summary and a discussion of the results.

2 METHOD

We start with the basic assumptions that the tracer (galaxies or haloes) number density is given by a Poisson sampling[1] of an underlying continuous field $n_g(\boldsymbol{x})$, with overdensity defined by $\delta_g(\boldsymbol{x}) \equiv n_g(\boldsymbol{x})/\bar{n}_g - 1$, and that the galaxy overdensity field is related to the mass overdensity field δ by a conditional probability distribution $P(\delta_g|\delta)$, including a stochastic element which we will describe later. In addition, in realistic surveys using the redshift as distance indicator, peculiar velocities distort clustering in a manner dependent on the angle with respect to the line of sight. In particular, clustering is enhanced – at least on large scales, in the linear regime – along the line of sight. It is the angular dependence of the effect that yields the signal to extract a measurement of the growth of structure. In this paper, our main goal is to estimate the error of the growth rate of perturbations, $f(z)$, generalizing the method proposed by McDonald & Seljak (2009) to measure $f(z)$, reducing the sample variance limit and using redshift-space distortions. The idea is to split the sample of objects into two subsamples with different biasing properties. In their work, they used a linear bias model and all stochasticity was due to shot noise. Here we generalize their work to the non-linear bias case and take into account the possibility of having off-diagonal noise terms, which can be introduced, for example, if there are objects in common in the two samples. In this regime, we compare the multi-tracer with the single-tracer approach, and we analyse how the noise and non-linearities affect the extent to which the sample variance limit can be improved. We also use results from simulations to set plausible values for these parameters to see how great the gains may be in practice.

2.1 Modelling of bias

The relation between clustering properties of the dark matter and those of the tracer (haloes or galaxies) goes under the name of 'bias'. The simplest, non-trivial bias model is linear bias, $\delta_g(\boldsymbol{x}) = b_1 \delta(\boldsymbol{x})$, with b_1 constant and independent of position and scale. This corresponds to a deterministic linear biasing, which has little physical motivation and is problematic if $b_1 > 1$ and the field is not very linear, since it allows an

---

[1] This model may be inadequate in details when dealing with real haloes (Smith, Scoccimarro & Sheth 2007; Seljak, Hamaus & Desjacques 2009). We will return to this point in Section 3.

unphysical $\delta_g < -1$ in voids. A more complex relation is almost certainly needed to properly describe galaxy clustering. Non-linear biasing with a bias which is no longer a constant but a function of $\delta(x)$ is a common way to improve the model.

Here we adopt the formalism proposed by Dekel & Lahav (1999). We assume that both $\delta(x)$ and $\delta_g(x)$ are random fields with one-point probability distributions functions, $P[\delta(x)]$ and $P[\delta_g(x)]$, with zero mean $\langle \delta(x) \rangle = \langle \delta_g(x) \rangle = 0$ and variances $\langle \delta^2(x) \rangle$ and $\langle \delta_g^2(x) \rangle$, respectively.

We first define the *mean* biasing function, $b[\delta(x)]$, as the conditional mean between the galaxy and the matter field,

$$b[\delta(x)]\delta(x) \equiv \langle \delta_g(x)|\delta(x) \rangle = \int d\delta_g(x) P(\delta_g(x)|\delta(x))\delta_g(x). \tag{1}$$

This is the natural generalization of the deterministic linear biasing relation, where the function $b[\delta(x)]$ characterizes the non-linear bias behaviour. Note that $P(\delta_g|\delta)$ can have a width (i.e. a scatter) around the mean relation, $b(\delta)$, which is however not captured by the function $b(\delta)$. We characterize the function $b(\delta)$ by the first- and second-order moments, which are given by $\hat{b}(r)$ and $\tilde{b}(r)$ at zero lag (i.e. $r = 0$),

$$\hat{b}(r) \equiv \frac{\langle \delta(x+r)\delta(x)b[\delta(x)] \rangle}{\langle \delta(x)\delta(x+r) \rangle}, \tag{2}$$

$$\tilde{b}^2(r) \equiv \frac{\langle \delta(x+r)\delta(x)b[\delta(x)]b[\delta(x+r)] \rangle}{\langle \delta(x)\delta(x+r) \rangle}, \tag{3}$$

where $\langle \rangle$ represents the averaging over the volume of the survey or over different realizations. These two parameters take into account the non-linearity of the system as long as one is concerned with the two-point correlation function (or the power spectrum) and not higher order correlations. It is useful to define their ratio as

$$R(r) \equiv \frac{\hat{b}(r)}{\tilde{b}(r)}, \tag{4}$$

which is a useful parameter to measure the non-linearity of the bias. In the linear case, $\hat{b}(r) = \tilde{b}(r)$ and $R(r) = 1$, whereas for non-linear cases, $R(r) < 1$. Note that $\hat{b}(r)$ is the bias as it would appear in the tracer–dark matter cross-correlation, while $\tilde{b}^2(r)$ would appear in the tracer auto-correlation.

We next turn to stochasticity, by which we mean any physical or statistical process that produces a non-deterministic relation between the dark matter and the galaxy (or halo) field. This may arise from the discrete nature of galaxies, in which case it is called shot noise; if it is a Poisson process, its expression is inversely proportional to the mean density of objects, $1/\bar{n}_g$, but the formalism used here allows for other stochastic processes which are encoded in the width of $P(\delta_g|\delta)$.

In order to study the stochasticity of the bias, we define the random bias field $\epsilon(x)$ as the difference between the galaxy field and the dark matter field once biased by the mean bias relation $b[\delta(x)]$,

$$\epsilon(x) \equiv \delta_g(x) - b[\delta(x)]\delta(x). \tag{5}$$

If $P(\delta_g|\delta)$ is a univariate Gaussian then $b(\delta)$ and $\sigma_b^2(\delta)$, the variance of $P(\delta_g|\delta)$ at a given δ, completely specifies $P(\delta_g|\delta)$. The variance of the ϵ field is given by the average of $\sigma_b^2(\delta)$ over δ.

In general, once δ, δ_g and ϵ are defined, the corresponding correlation functions are

$$\xi_{mm}(r) \equiv \langle \delta(x)\delta(x+r) \rangle, \tag{6}$$

$$\xi_{gm}(r) \equiv \langle \delta_g(x)\delta(x+r) \rangle, \tag{7}$$

$$\xi_{gg}(r) \equiv \langle \delta_g(x)\delta_g(x+r) \rangle, \tag{8}$$

$$\xi_{\epsilon\epsilon}(r) \equiv \langle \epsilon(x)\epsilon(x+r) \rangle, \tag{9}$$

$$\xi_{\epsilon m}(r) \equiv \langle \epsilon(x)\delta(x+r) \rangle. \tag{10}$$

In what follows, we are only interested in the two-point correlation function or the power spectrum, thus we do not need to specify further moments of $P(\delta_g|\delta)$ or higher order correlations.

In order to give an illustrative example to the bias formalism, let us consider a simple non-linear bias model given by

$$b(\delta) = b_0 + b_1\delta + b_2\delta^2. \tag{11}$$

Let us also assume that this bias model is non-stochastic. Therefore according to equation (5), we have that the galaxy overdensity must be

$$\delta_g = b_0\delta + b_1\delta^2 + b_2\delta^3. \tag{12}$$

In order to deal with the simplest scenario, we assume that δ is a Gaussian random field. This means that the n-point correlation function, $\langle \delta^n \rangle$, can be expressed as a function of the two-point correlation function, $\langle \delta^2 \rangle$.

We have stated above that δ_g field has to satisfy $\langle \delta_g \rangle = 0$. Provided that $\langle \delta \rangle = \langle \delta^3 \rangle = 0$, we have that b_1 must be null:

$$b(\delta) = b_0 + b_2\delta^2. \tag{13}$$

Table 1. Different sets of parameters of the non-linear model given by equation (13) and the corresponding values of $\hat{b}$, $\tilde{b}$ and R.

Set	b_0	$b_2\langle\delta^2\rangle$	$\hat{b}$	$\tilde{b}$	R
Set 1	1	0	1	1	1
Set 2	2	0	2	2	1
Set 3	0.7	0	0.7	0.7	1
Set 4	1	0.5	2.50	2.78	0.90
Set 5	0.3	0.7	2.40	2.94	0.81
Set 6	2	−0.3	1.1	1.32	0.83
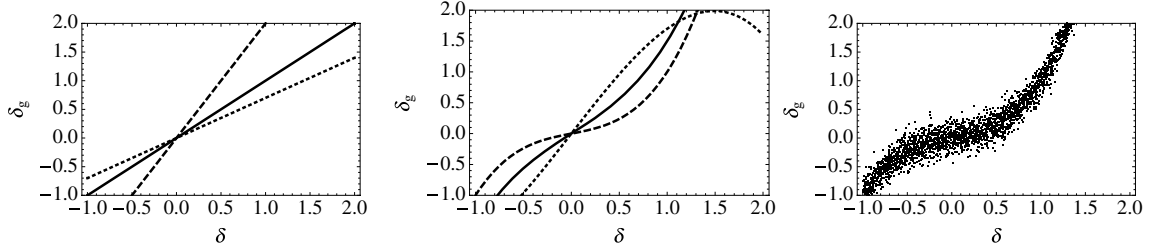
Figure 1. Different biasing models. Left-hand panel: Set 1 (solid line), Set 2 (dashed line) and Set 3 (dotted line). Middle panel: Set 4 (solid line), Set 5 (dashed line) and Set 6 (dotted line). Right-hand panel: Set 5 with stochasticity component. Sets are detailed in Table 1.

The biasing parameters given by equations (2) and (3) are

$$\hat{b} = b_0 + 3b_2\langle\delta^2\rangle, \tag{14}$$

$$\tilde{b}^2 = b_0^2 + 6b_0 b_2\langle\delta^2\rangle + 15b_2^2\langle\delta^2\rangle^2, \tag{15}$$

where we have used $\langle\delta^4\rangle = 3\langle\delta^2\rangle^2$ and $\langle\delta^6\rangle = 15\langle\delta^2\rangle^3$, if δ is Gaussian.

As an illustrative example, we consider this simple biasing model with different set of parameters, b_0 and b_2, listed in Table 1. Set 1, Set 2 and Set 3 are linear biasing models ($b_2 = 0$), and therefore $\hat{b} = \tilde{b}$ and $R = 1$. These models are plotted in Fig. 1 (left-hand panel) with solid, dashed and dotted lines, respectively. On the other hand, Set 4, Set 5 and Set 6 are non-linear biasing models ($b_2 \neq 0$) plotted in Fig. 1 (central panel) with solid, dashed and dotted lines, respectively. In these cases, it is clear that $\hat{b} \neq \tilde{b}$ and therefore $R < 1$. Note also that R is a good indicator of how non-linear the model is; the more deviate R is from 1 the more non-linear $b(\delta)$ is.

Finally, the stochasticity can easily be included in this formalism just adding a random field, namely ϵ, in equation (12). The result of doing this with Set 5 is shown in Fig. 1 (right-hand panel).

2.2 Redshift-space distortions of biased tracers

So far, we have defined the bias model in configuration space. However, to deal with redshift-space distortions, it is more convenient to work with the galaxy density function δ_g in k-space. The basic relation for redshift distortions is, assuming the distant observer approximation (Kaiser 1987),

$$\delta_g^s(\boldsymbol{k}) = \delta_g(\boldsymbol{k}) + f\mu^2(\hat{k})\delta(\boldsymbol{k}), \tag{16}$$

where $\delta_g(\boldsymbol{k})$ is the Fourier transform of the overdensity of galaxies, $\delta_g^s(\boldsymbol{k})$ its redshift-space counterpart and $\delta(\boldsymbol{k})$ the real-space transform of the overdensity of dark matter. $\mu(\hat{k})$ is the cosine of the angle between the line of sight and $\boldsymbol{k}$. Throughout this section, we consider a single snapshot at a fixed redshift, usually we take $z = 0$. Therefore we write, for instance, f instead of $f(z)$ and so on. If we consider that we have two dark matter tracers, then

$$\delta_{gi}^s(\boldsymbol{k}) = \delta_{gi}(\boldsymbol{k}) + f\mu^2(\hat{k})\delta(\boldsymbol{k}); \qquad i = 1, 2, \tag{17}$$

with a corresponding covariance matrix

$$C_{2\mathrm{tr}}(\boldsymbol{k}) = \begin{pmatrix} \langle\delta_{g1}^s(\boldsymbol{k})\delta_{g1}^{s\,*}(\boldsymbol{k})\rangle & \langle\delta_{g1}^s(\boldsymbol{k})\delta_{g2}^{s\,*}(\boldsymbol{k})\rangle \\ \langle\delta_{g1}^s(\boldsymbol{k})\delta_{g2}^{s\,*}(\boldsymbol{k})\rangle & \langle\delta_{g2}^s(\boldsymbol{k})\delta_{g2}^{s\,*}(\boldsymbol{k})\rangle \end{pmatrix}. \tag{18}$$

We define various power spectra as

$$\langle\delta_{gi}(\boldsymbol{k})\delta_{gj}^*(\boldsymbol{k})\rangle \equiv P_{gigj}(k) \tag{19}$$

$$\langle\delta(\boldsymbol{k})\delta_{gi}^*(\boldsymbol{k})\rangle \equiv P_{mgi}(k) \tag{20}$$

$$\langle\delta(\boldsymbol{k})\delta^*(\boldsymbol{k})\rangle \equiv P_{mm}(k). \tag{21}$$

Note that these quantities are related to equations (6–8) through their Fourier transforms,

$$P(k) = \int d^3\boldsymbol{r} \, \xi(r) \, e^{i\boldsymbol{k} \cdot r}. \tag{22}$$

The terms of the covariance matrix of equation (18) can be expressed as,[2]

$$\langle \delta_{gi}^s(\boldsymbol{k}) \delta_{gj}^s{}^*(\boldsymbol{k}) \rangle = P_{mm}(k) \left[\frac{P_{gigj}(k)}{P_{mm}(k)} + f\mu^2 \frac{P_{mgi}(k) + P_{mgj}(k)}{P_{mm}(k)} + f^2\mu^4 \right]. \tag{23}$$

In this paper, we will assume that the cross terms between the random field and the matter field are subdominant and can be assumed to be zero:

$$\langle \epsilon(\boldsymbol{x}) B[\delta(\boldsymbol{x} + r)] \rangle = 0, \tag{24}$$

where $B[\delta]$ is any function of δ. If we relate these quantities to the bias parameters defined in equations (2) and (3), we obtain

$$P_{mm}(k) = \int d^3\boldsymbol{r} \, \xi_{mm}(r) \, e^{-i\boldsymbol{k} \cdot r}, \tag{25}$$

$$P_{mgi}(k) = \int d^3\boldsymbol{r} \, \xi_{mgi}(r) \, e^{-i\boldsymbol{k} \cdot r} = \int d^3\boldsymbol{r} \, \hat{b}_i(r) \xi_{mm}(r) \, e^{-i\boldsymbol{k} \cdot r}, \tag{26}$$

$$P_{gigi}(k) = \int d^3\boldsymbol{r} \, \xi_{gigi}(r) \, e^{-i\boldsymbol{k} \cdot r} = \int d^3\boldsymbol{r} \, \tilde{b}_i^2(r) \xi_{mm}(r) \, e^{-i\boldsymbol{k} \cdot r} + \int d^3\boldsymbol{r} \, \xi_{\epsilon i \epsilon i}(r) \, e^{-i\boldsymbol{k} \cdot r}, \tag{27}$$

$$P_{g1g2}(k) = \int d^3\boldsymbol{r} \, \xi_{g1g2}(r) \, e^{-i\boldsymbol{k} \cdot r} = \int d^3\boldsymbol{r} \, \langle \delta(\boldsymbol{x}) b_1[\delta(\boldsymbol{x})] \delta(r + \boldsymbol{x}) b_2[\delta(r + \boldsymbol{x})] \rangle \, e^{-i\boldsymbol{k} \cdot r} + \int d^3\boldsymbol{r} \, \xi_{\epsilon_1 \epsilon_2}(r) \, e^{-i\boldsymbol{k} \cdot r}, \tag{28}$$

$$= \int d^3\boldsymbol{r} \, R_{12}(r) \tilde{b}_1(r) \tilde{b}_2(r) \xi_{mm}(r) \, e^{-i\boldsymbol{k} \cdot r} + \int d^3\boldsymbol{r} \, \xi_{\epsilon_1 \epsilon_2}(r) \, e^{-i\boldsymbol{k} \cdot r}, \tag{29}$$

where the parameter $R_{12}(r)$ is a new non-linearity parameter between tracers of type 1 and 2, which is defined as

$$R_{12}(r) \equiv \frac{\langle b_1[\delta(\boldsymbol{x})] \delta(\boldsymbol{x}) b_2[\delta(\boldsymbol{x} + r)] \delta(\boldsymbol{x} + r) \rangle}{\langle b_1[\delta(\boldsymbol{x})] \delta(\boldsymbol{x}) b_1[\delta(\boldsymbol{x} + r)] \delta(\boldsymbol{x} + r) \rangle^{1/2} \langle b_2[\delta(\boldsymbol{x})] \delta(\boldsymbol{x}) b_2[\delta(\boldsymbol{x} + r)] \delta(\boldsymbol{x} + r) \rangle^{1/2}}. \tag{30}$$

For convenience we define new bias parameters,

$$\hat{b}_i(k) \equiv \frac{\int d^3\boldsymbol{r} \, \hat{b}_i(r) \xi_{mm}(r) \, e^{-i\boldsymbol{k} \cdot r}}{\int d^3\boldsymbol{r} \, \xi_{mm}(r) \, e^{-i\boldsymbol{k} \cdot r}} = \frac{P_{mgi}(k)}{P_{mm}(k)} \tag{31}$$

$$\tilde{b}_i^2(k) \equiv \frac{\int d^3\boldsymbol{r} \, \tilde{b}_i^2(r) \xi_{mm}(r) \, e^{-i\boldsymbol{k} \cdot r}}{\int d^3\boldsymbol{r} \, \xi_{mm}(r) \, e^{-i\boldsymbol{k} \cdot r}} = \frac{P_{gigi}(k) - P_{\epsilon i \epsilon i}(k)}{P_{mm}(k)} \tag{32}$$

$$R_i(k) \equiv \frac{\hat{b}_i(k)}{\tilde{b}_i(k)} = \frac{P_{mgi}(k)}{\{P_{mm}(k)[P_{gigi}(k) - P_{\epsilon i \epsilon i}(k)]\}^{1/2}} \tag{33}$$

$$R_{12}(k) \equiv \frac{\int d^3\boldsymbol{r} \, R_{12}(r) \tilde{b}_1(r) \tilde{b}_2(r) \xi_{mm} \, e^{-i\boldsymbol{k} \cdot r}}{\left[\int d^3\boldsymbol{r} \, \tilde{b}_1^2(r) \xi_{mm}(r) \, e^{-i\boldsymbol{k} \cdot r} \right]^{1/2} \left[\int d^3\boldsymbol{r} \, \tilde{b}_2^2(r) \xi_{mm}(r) \, e^{-i\boldsymbol{k} \cdot r} \right]^{1/2}} = \frac{P_{g1g2}(k) - P_{\epsilon 1 \epsilon 2}(k)}{[P_{g1g1}(k) - P_{\epsilon 1 \epsilon 1}(k)]^{1/2} [P_{g2g2}(k) - P_{\epsilon 2 \epsilon 2}(k)]^{1/2}} \tag{34}$$

and the ϵ field power spectrum,

$$P_{\epsilon i \epsilon j}(k) \equiv \int d^3\boldsymbol{r} \, \xi_{\epsilon i \epsilon j}(r) \, e^{-i\boldsymbol{k} \cdot r}. \tag{35}$$

Thus the covariance matrix reads

$$C_{2tr}(k, \mu) = \begin{pmatrix} C_{11}(k, \mu) & C_{12}(k, \mu) \\ C_{12}(k, \mu) & C_{22}(k, \mu) \end{pmatrix} \tag{36}$$

with

$$C_{11}(k, \mu) = P_{mm}(k) \left[\tilde{b}_1^2(k) + 2f\mu^2 \tilde{b}_1(k) R_1(k) + f^2\mu^4 \right] + P_{\epsilon_1 \epsilon_1}(k), \tag{37}$$

$$C_{12}(k, \mu) = P_{mm}(k) \left\{ R_{12}(k) \tilde{b}_1(k) \tilde{b}_2(k) + f\mu^2 \left[\tilde{b}_1(k) R_1(k) + \tilde{b}_2(k) R_2(k) \right] + \mu^4 f^2 \right\} + P_{\epsilon_1 \epsilon_2}(k), \tag{38}$$

$$C_{22}(k, \mu) = P_{mm}(k) \left[\tilde{b}_2^2(k) + 2f\mu^2 \tilde{b}_2(k) R_2(k) + f^2\mu^4 \right] + P_{\epsilon_2 \epsilon_2}(k). \tag{39}$$

Because $\tilde{b}_i(k)$ are parameters which cannot be obtained readily from observations, it is preferable to work with the redshift-space distortion parameter $\tilde{\beta}_i(k) \equiv f/\tilde{b}_i(k)$. Also $P_{mm}(k, z)$ at a given z is not directly measurable. However, in the linear regime we can write it as

---

[2] In this expression the factor of 2 difference from that appearing in McDonald & Seljak (2009) is due to a different definition of $\delta_{\boldsymbol{k}}$; we consider that $\delta_{\boldsymbol{k}}$ is complex.

$P_{mm}(k, z) = D^2(z) P_{mm}^0(k)$, where $P_{mm}^0(k)$ is the fiducial power spectrum and $D(z)$ is the linear growth factor. Thus, defining $x(z) \equiv D^2(z) f^2(z)$ the covariance matrix elements are

$$C_{11}(k, \mu) = x P_{mm}^0(k) \left[\tilde{\beta}_1^{-2}(k) + 2\mu^2 \tilde{\beta}_1^{-1}(k) R_1(k) + \mu^4 \right] + P_{\epsilon_1 \epsilon_1}(k), \tag{40}$$

$$C_{12}(k, \mu) = x P_{mm}^0(k) \left\{ R_{12}(k) \tilde{\beta}_1^{-1}(k) \tilde{\beta}_2^{-1}(k) + \mu^2 \left[\tilde{\beta}_1^{-1}(k) R_1(k) + \tilde{\beta}_2^{-1}(k) R_2(k) \right] + \mu^4 \right\} + P_{\epsilon_1 \epsilon_2}(k), \tag{41}$$

$$C_{22}(k, \mu) = x P_{mm}^0(k) \left[\tilde{\beta}_2^{-2}(k) + 2\mu^2 \tilde{\beta}_2^{-1}(k) R_2(k) + \mu^4 \right] + P_{\epsilon_2 \epsilon_2}(k). \tag{42}$$

Note that the quantity x encompasses all the relevant cosmological information about the growth of structure. From equations (40–42) one can see that the covariance matrix can easily be separated in two parts: a signal part $\hat{S}$ and a noise contribution $\hat{N}$:

$$C_{2tr}(k, \mu) = \hat{S}(k, \mu) + \hat{N}(k), \tag{43}$$

where the noise matrix is

$$\hat{N}_{ij}(k) = P_{\epsilon_i \epsilon_j}(k). \tag{44}$$

If the two tracers are both a Poisson sample of the dark matter field and do not overlap then $\hat{N}_{ij}(k)$ is diagonal, scale-independent and its elements are $\hat{N}_{11} = 1/\bar{n}_1$; $\hat{N}_{22} = 1/\bar{n}_2$. This is the case considered by McDonald & Seljak (2009). Any other source of stochasticity would add to the discreteness effect and, in general, may yield non-zero off-diagonal contributions $\hat{N}_{12} \neq 0$.

If we can estimate the noise part and $P_{mm}^0(k)$ is given, e.g. by CMB observations, then the covariance matrix depends on six parametric functions: x, $\tilde{\beta}_1(k)$, $\tilde{\beta}_2(k)$, $R_1(k)$, $R_2(k)$ and $R_{12}(k)$.

Considering one dark matter tracer, the covariance matrix is simpler:

$$C_{1tr}(k, \mu) = x P_{mm}^0(k) \left[\tilde{\beta}^{-2}(k) + 2\mu^2 \tilde{\beta}^{-1}(k) R(k) + \mu^4 \right] + P_{\epsilon \epsilon}(k), \tag{45}$$

and depends on only x, $\beta(k)$ and $R(k)$.

An interesting point is the 'hidden' relation between the different variables. For instance, given $\beta_1(k)$, $\beta_2(k)$ and the relative number of objects of these two tracers, $\beta(k)$ is constrained. Also, in the two-tracer case, given $R_1(k)$ and $R_2(k)$, $R_{12}(k)$ is also constrained. We give these relations in Appendices A and B.

2.3 Forecasting errors

We use the Fisher matrix formalism (Fisher 1935) to estimate errors on x. When the means of the data are fixed (i.e. for a given fiducial model), the Fisher matrix is given by (Tegmark, Taylor & Heavens 1997)

$$F_{\lambda \lambda'} = \frac{1}{2} \text{Tr} \left[C_{,\lambda} C^{-1} C_{,\lambda'} C^{-1} \right], \tag{46}$$

where $C_{,\lambda} \equiv dC/d\lambda$, C is the covariance matrix and λ the parameters of the model. The marginalized variance of parameter λ is given by

$$\sigma_\lambda^2 = (F^{-1})_{\lambda \lambda}. \tag{47}$$

Following e.g. Feldman, Kaiser & Peacock (1994) for a survey volume V_u, in the continuum approximation,

$$F_{\lambda \lambda'}^V = \frac{V_u}{(2\pi)^3} \int_{k_{min}}^{k_{max}} F_{\lambda \lambda'}(\boldsymbol{k}) d^3 \boldsymbol{k} \tag{48}$$

which we evaluate by a discrete sum

$$F_{\lambda \lambda'}^V \simeq \frac{V_u}{(2\pi)^2} \sum_{\mu=-1}^{+1} \Delta \mu \sum_{k=k_{min}}^{k_{max}} \Delta k F_{\lambda \lambda'}(k, \mu) k^2. \tag{49}$$

In this paper, we assume that the fiducial power spectrum, the noise matrix and the R parameters are known and we fix them to their fiducial values. We will explore the dependence of the results on the assumed fiducial values in the following sections. The lambda parameters are $\tilde{b}_1$, $\tilde{b}_2$ and x or, equivalently, $\tilde{\beta}_1$, $\tilde{\beta}_2$ and x for the two-tracers case; when reporting errors on x, we marginalize over $\tilde{\beta}_1$ and $\tilde{\beta}_2$. For the one-tracer case the parameters are $\tilde{b}$ and x and the reported errors on x are marginalized over $\tilde{b}$.

The specific values of k_{min} and k_{max} depend on the features of the survey, k_{min} being set by the survey volume and k_{max} is usually set by the onset of non-linearities. In our case, we set $k_{min} = 2\pi/V_u^{1/3}$ and conservatively set $k_{max} = 0.1 \, \text{Mpc}^{-1} \, h$ for $z = 0$.

In the next section, we will compare the errors of x, obtained using the one- and the two-tracer approaches. To produce the figures, we assume we have a single snapshot at $z = 0$, which corresponds to $f = 0.483$ in a standard Λ cold dark matter (CDM) universe, the power spectrum is given by CAMB (Code for Anisotropies in the Microwave Background: Lewis, Challinor & Lasenby 2000), the sampling volume is set to be $V_u = 1 \, (\text{Gpc} \, h^{-1})^3$ and all biases and all R coefficients are taken to be scale-independent. The relative number of tracers is $Y \equiv \bar{n}_1/\bar{n}_2$, and the S/N is $\text{S/N} \equiv \tilde{b}^2 P^0(k = 0.1 \, h \, \text{Mpc}^{-1}) \bar{n}$. Note that the S/N is defined relative to the *combined* sample of tracers. The relation between $\tilde{b}$ and $\tilde{b}_1$, $\tilde{b}_2$ is given in equation (A8). Since the fractional cosmic-variance error on the power spectrum (in a shell in Fourier space) is constant with redshift, the quantities reported below are valid at different z provided that the S/N and the various bias parameters are defined at the redshift of interest. Note however that the fiducial value of x (and that
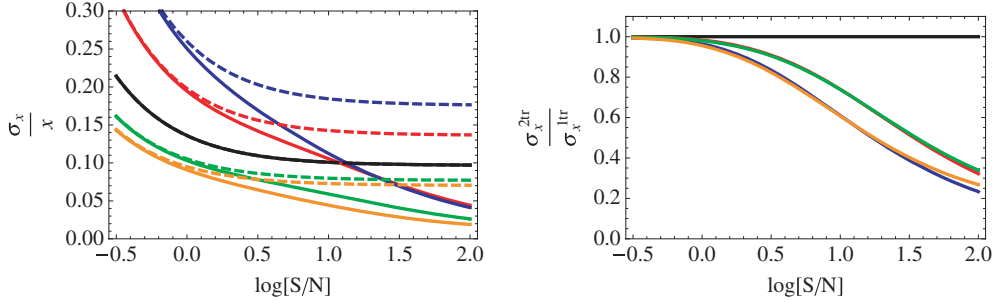
Figure 2. Left-hand panel: σ_x/x versus S/N. Right-hand panel: $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}}$ versus S/N. In both panels $\tilde{b}_1 = 1$ and (from bottom to top) $\tilde{b}_2 = 1/3$ (orange line), 1/2 (green line), 1 (black line), 2 (red line) and 3 (blue line). For the combined sample according to equation (A8) the total bias factors are $\tilde{b} = 2/3$ (orange line), $\tilde{b} = 3/4$ (green line), $\tilde{b} = 1$ (black line), $\tilde{b} = 3/2$ (red line) and $\tilde{b} = 2$ (blue line). In the left-hand panel the solid lines represent the errors for two-tracer case, the dashed lines for one tracer.

of f) change with redshift: x increases with redshift up $z = 0.5$ and decreases for larger z, while f increases with redshift tending to 1 asymptotically. We find that the dependence of the fractional error, σ_x/x, and of the ratio of errors, $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}}$, on the value of x is weak. More importantly, the k_{max} at which non-linearities become important is expected to depend on redshift and to increase roughly as $(1 + z)$. The number of independent modes N in a given volume grows roughly like k_{max}^3 and the variance scales like $1/N$.

2.4 Dependence on bias

Given that the forecasted error on x depends on many variables, we will start by considering the effects of one variable at a time. The first important effect to be analysed is how the bias of the tracers (absolute and relative) affects the measurements of x (and therefore of f). For simplicity, we assume linear biasing (i.e. all R parameters equal to 1) and the same number of objects for two distinct populations ($Y = 1$) with diagonal Poisson-like noise.

In Fig. 2, we show σ_x/x (left-hand panel) and $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}}$ (right-hand panel) versus S/N. The two-tracer case indicates that the two tracers are treated separately yielding a covariance matrix as in equations (36–42). The one-tracer case indicates that both tracers are included in the same sample and the covariance matrix is given by equation (45). In both panels, we have set $\tilde{b}_1 = 1$ (i.e. tracer 1 is unbiased) and $\tilde{b}_2$ ranges from 0.33 to 3: $\tilde{b}_2 = 1/3$ (orange line), 1/2 (green line), 1 (black line), 2 (red line), 3 (blue line). The bias factors for the combined sample according to equation (A8) are $\tilde{b} = 2/3$ (orange line), 3/4 (green line), 1 (black line), 3/2 (red line) and 2 (blue line). In the left-hand panel, the solid lines correspond to the two-tracer case, the dashed lines to the one-tracer case.

From Fig. 2 (left-hand panel), we see that for the single-tracer case (dashed lines), the lower the combined bias $\tilde{b}$, the lower the error σ_x. This can be understood if we recall that to measure f we are using the redshift-space distortion effect. Since this effect is proportional to $1/\tilde{b}$, the lower the bias, the larger the redshift-space distortions. As noted in McDonald & Seljak (2009), this means that, slightly counter-intuitively, low-bias objects may act as useful tracers, even if they are used as a single population. On the other hand, for the two-tracer case (solid lines) we can see that (i) the improvement (compared to the one-tracer case) is significant only when the signal dominates ($\log[\mathrm{S/N}] \gtrsim 0.7$) and (ii) the improvement increases as the difference in the biases of the two populations increases, also as noted by McDonald & Seljak (2009). Note that McDonald & Seljak (2009) measure the normalization of $P_{\theta\theta}$ (with the shape given by external observables) which is a different quantity from x considered here. However the fractional forecasted errors in the two quantity are the same, $\sigma_x/x = \sigma_{P_{\theta\theta}}/P_{\theta\theta}$ as well as the relative two-tracer versus one-tracer improvement $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}} = \sigma_{P_{\theta\theta}}^{2\mathrm{tr}}/\sigma_{P_{\theta\theta}}^{2\mathrm{tr}}$.

Fig. 2 (right-hand panel) further quantifies the effect. The improvement between the two cases depends on the *ratio of biases* and is only significant if the S/N is large enough. In particular, for a bias ratio of 3 (blue and orange lines), the improvement is significant ($\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}} \leq 0.5$) for $\log[\mathrm{S/N}] \geq 1.3$. However, for ratios of 2 (red and green lines), the improvement starts to be significant only when $\log[\mathrm{S/N}] \geq 1.5$. Note also that for the special case $\tilde{b}_2 = \tilde{b}_1$ (black line), as expected, there is no improvement. When comparing these results – especially Fig. 2 (left-hand panel) – with those of McDonald & Seljak (2009), one should keep in mind that they define the S/N at $k = 0.4\,h\,\mathrm{Mpc}^{-1}$ while we use $k = 0.1\,h\,\mathrm{Mpc}^{-1}$, and that $P(k = 0.1\,h\,\mathrm{Mpc}^{-1}) \simeq 12\,P(k = 0.4\,h\,\mathrm{Mpc}^{-1})$. With this in mind, we reproduce their results.

We also conclude that for surveys with a low S/N ($\log[\mathrm{S/N}] \lesssim 0.7$), it is better to have a low-bias tracer; splitting the sample and using two tracers will not yield significant improvement. For example, for a bias of $\tilde{b} = 0.75$ (green dashed line in left-hand panel of Fig. 2), we obtain fractional errors of $\sigma_x/x \simeq 0.08$ for a $\log[\mathrm{S/N}] = 0.7$. On the other hand, for surveys with higher S/N ($\log[\mathrm{S/N}] \gtrsim 0.7$), it is better to use the two-tracer case, choosing two tracers with the highest possible bias ratio. In this case, for biases of $\tilde{b}_1 = 1$ and $\tilde{b}_2 = 0.5$ (green solid line in left-hand panel of Fig. 2), we reach $\sigma_x/x \simeq 0.04$ for $\log[\mathrm{S/N}] = 1.5$. In practice, of course, the choice of which tracers to use is complicated by their number density; a high bias may be desirable, but one probably pays a penalty through low density and high shot noise.

2.5 Effect of bias non-linearities

The second interesting issue is to consider how non-linearities in the bias (i.e. the R parameters) affect σ_x. In this case, we fix $\tilde{b}_1 = 1$ and $\tilde{b}_2 = 2$ with the same number of objects for each tracer ($Y = 1$) and Poisson noise. In Fig. 3 and 4 (left-hand panels), we show how $\sigma_x^{1\mathrm{tr}}/x$
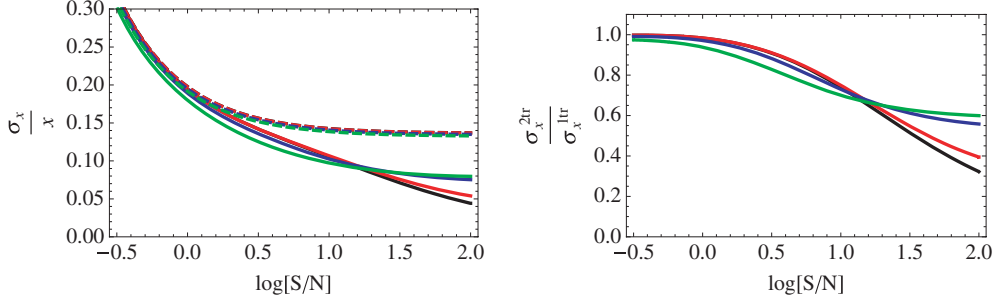
Figure 3. Left-hand panel: fractional error σ_x/x versus S/N. Dashed lines represent the single-tracer case and solid lines the two-tracer case. Right-hand panel: $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}}$ versus S/N. In both panels the colours show different non-linear cases: $R_1 = 1.00$ (black solid line), $R_1 = 0.99$ (red solid line), $R_1 = 0.90$ (blue solid line) and $R_1 = 0.80$ (green solid line); $R_2 = 1$ and $R_{12} = R_1$. For dashed lines, R is the corresponding value for the full sample given the above values for R_2 and R_{12} as in equation (A4): $R = 1.000$ (black dashed line), $R = 0.999$ (red dashed line), $R = 0.989$ (blue dashed line) and $R = 0.978$ (green dashed line).
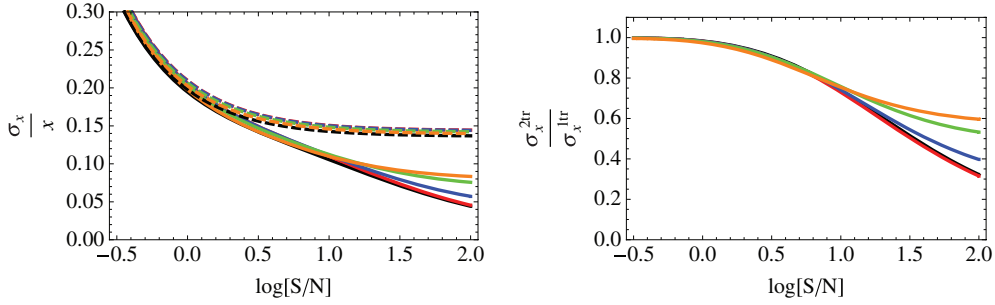

Figure 4. Left-hand panel: fractional error σ_x/x versus S/N. Dashed lines represent the single-tracer case and solid lines the two-tracer case. Right-hand panel: $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}}$ versus S/N. In both panels, the colours show different non-linear cases: $R_{12} = 1.00$ (red solid line), $R_{12} = 0.99$ (blue solid line), $R_{12} = 0.95$ (green solid line) and $R_{12} = 0.90$ (orange solid line); $R_1 = R_2 = 0.9$. The solid black line is the prefect linear case. For dashed lines, R is the corresponding value for the full sample given the above values for R_1, R_2 and R_{12} as in equation (A4): $R = 1.000$ (black dashed line), $R = 0.900$ (red dashed line), $R = 0.902$ (blue dashed line), $R = 0.910$ (green dashed line) and $R = 0.921$ (orange dashed line).

(dashed lines) and $\sigma_x^{2\mathrm{tr}}/x$ (solid lines) vary with S/N. In both cases, the black line is for the perfect linear bias case $R_1 = R_2 = R_{12} = 1$ and colour lines show different non-linear cases (see Figs 3 and 4 captions for details). In Figs 3 and 4 (right-hand panels), we show how the ratio $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}}$ varies with S/N.

In general, we see from Figs 3 and 4 (left-hand panels) that the two-tracer case is more sensitive to non-linear bias effects than the one-tracer case for high S/N ($\log[\mathrm{S/N}] \geq 1.2$). In particular, in Fig. 3, for the single-tracer case a deviation from unity of R_1 produces a slight reduction of the error which is the same for all the S/N range explored. This is due to a reduction of the combined bias $\tilde{b}$: as we reduce R_{12} (because we set $R_1 = R_{12}$), $\tilde{b}$ is reduced (see equation A7) and therefore the total error is also reduced as we have seen in Section 2.4. On the other hand, for the two-tracer case, there are two opposite behaviours depending on the value of S/N: for $\log[\mathrm{S/N}] \gtrsim 1.2$, we observe that non-linearities produce an increase in the error, whereas for $\log[\mathrm{S/N}] \lesssim 1.2$, they produce a reduction. In the high S/N regime, reducing R_1 reduces the μ^2 coefficient in equation (40), and thus reduces the angular dependence. In the low S/N regime, this is compensated by the fact that the off-diagonal terms of the covariance matrix are reduced by non-linearities (recall that the noise off-diagonal terms are set to zero here). While this may not be clear at first sight from equation (41), we have verified it numerically.

From Fig. 3 (right-hand panel), we see that the non-linear bias increases the improvement between the two approaches for $\log[\mathrm{S/N}] \lesssim 1.2$, and limits it for $\log[\mathrm{S/N}] \gtrsim 1.2$. In particular, non-linear bias with $R_1 = 0.9$ and $R_2 = 1$ gives $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}} \simeq 0.6$ for $\log[\mathrm{S/N}] = 1.5$, whereas for the perfect linear bias case it is $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}} \simeq 0.5$. At lower S/N, namely $\log[\mathrm{S/N}] = 0.7$, we obtain for the same non-linear bias case, $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}} \simeq 0.85$, and for the linear bias case, $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}} \simeq 0.9$. Therefore, non-linearities affect the two-tracer approach considerably more than the single-tracer approach. Non-linearities in the mean bias relation slightly increase the precision of the x measurement for low S/N regime ($\log[\mathrm{S/N}] \lesssim 1.2$), but they limit the effectiveness of the two-tracer approach for high S/N ($\log[\mathrm{S/N}] \gtrsim 1.2$).

On the other hand, from Fig. 4 we observe a very similar behaviour. In this case, we have set $R_1 = R_2 = 0.9$ and we change the value of R_{12}. First of all, we observe that the one-tracer case is not very sensitive to non-linearities in this range of Rs. The small changes for one tracer are mainly due to the change of the combined bias as we have noted above for Fig. 3. The second point is that little deviations from $R_{12} = 1.0$ produce an increase on the fractional error for the two-tracer case for $\log[\mathrm{S/N}] \geq 1.2$, as it can be seen in Fig. 4 (left-hand panel). Also in Fig. 4 (right-hand panel), we observe that the ratio of errors increases quickly for $\log[\mathrm{S/N}] \geq 1.2$, as we leave $R_{12} = 1.00$

In summary, it is important to note (Figs 2, 3 and 4) that the two-tracer method yields a substantial improvement compared to the one-tracer approach only in the high S/N regime and if the non-linearity parameters R are close to unity. Even if R_1 is 0.9, with $R_2 = 1$ or $R_1 = R_2 = 0.9$ with $R_{12} = 0.95$, the gain saturates at about a factor of two. In Section 3.2, we will address the issue of whether dark matter
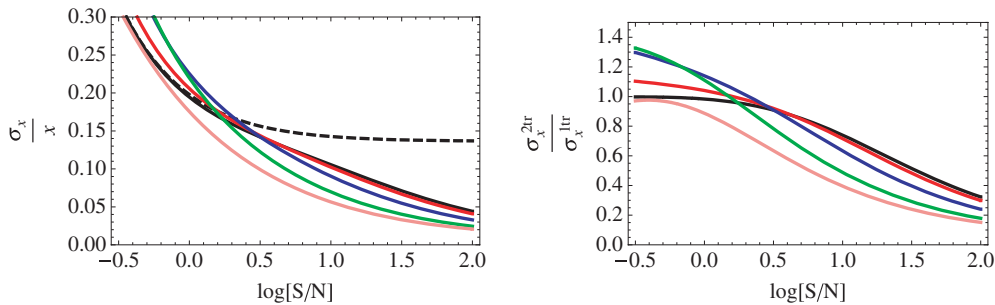
Figure 5. Left-hand panel: fractional error σ_x/x versus S/N. Right-hand panel: $\sigma_x^{2tr}/\sigma_x^{1tr}$ versus S/N. For both panels, $N_{12}/N_{11} = 0$ (black line), $N_{12}/N_{11} = 0.4$ (red line), $N_{12}/N_{11} = 0.8$ (blue line), $N_{12}/N_{11} = 0.9$ (green line) and $N_{12}/N_{11} = 1.0$ (pink line). The dashed line on the left-hand panel corresponds to the one-tracer case which is not affected by N_{12}.
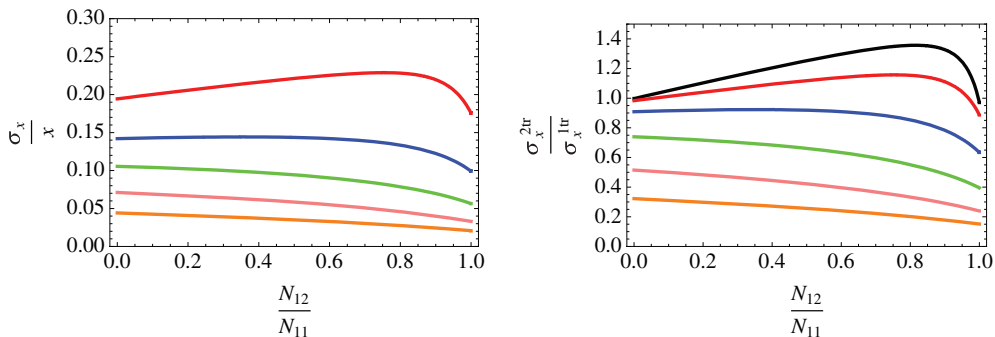

Figure 6. Left-hand panel: fractional error σ_x/x versus N_{12}/N_{11}. Right-hand panel: $\sigma_x^{2tr}/\sigma_x^{1tr}$ versus N_{12}/N_{11}. For both panels, $\log S/N = -0.5$ (black line), $\log S/N = 0.0$ (red line), $\log S/N = 0.5$ (blue line), $\log S/N = 1.0$ (green line), $\log S/N = 1.5$ (pink line) and $\log S/N = 2.0$ (orange line).

haloes, as seen in N-body simulations, trace the underlying dark matter with a bias that is linear enough for two tracers to be significantly advantageous compared to one.

2.6 Effect of off-diagonal noise terms

So far we have assumed that the noise matrix in equation (43) is diagonal; but any source of stochasticity in addition to Poisson sampling of two disjunct set of objects will add extra contributions to the noise matrix, which are non necessarily diagonal. Here we explore how σ_x/x and $\sigma_x^{2tr}/\sigma_x^{1tr}$ change for a non-zero off-diagonal noise term. For simplicity we discuss the linear bias case with $\tilde{b}_1 = 1$ and $\tilde{b}_2 = 2$ and the same number of objects for each tracer. Here, for direct comparison with previous examples, we will still set the diagonal elements of the noise matrix by the number density of the tracers (as if it was Poisson), and since $Y = 1, N_{11} = N_{22}$, but we allow $N_{12} \neq 0$. Note however that in any realistic application, a process that adds non-zero off-diagonal noise terms will also increase the diagonal matrix elements. This should be kept in mind during the following discussion.

In Fig. 5, we show how the error σ_x/x (left-hand panel) and the ratio of errors $\sigma_x^{2tr}/\sigma_x^{1tr}$ (right-hand panel) change with S/N for different values of N_{12}: from the reference case $N_{12} = 0$ (black-solid line) to $N_{12} = N_{11}$ (pink-solid line).[3] The black-dashed line represents the one-tracer case, which is not affected by changes in N_{12}.

We see a different behaviour in the high S/N regime ($\log[S/N] \gg 0$) and in the low S/N regime ($\log[S/N] \lesssim 0$). In the high S/N regime, the higher the off-diagonal noise, the lower the error σ_x. This can be understood if we imagine the off-diagonal noise term as a correlation term between the noise terms. The more correlated the noise of the two tracers, the less the total noise; for high values of N_{12}, knowing N_{11} means knowing N_{22}. In the low S/N regime, we observe the opposite behaviour: the higher the N_{12}, the higher the error. We also observe that when the value of N_{12} is very close to N_{11}, then σ_x decreases abruptly. In the low-signal case, adding a non-perfect correlation between noise terms just means adding more noise, and only in the case this correlation between noise terms is nearly perfect ($N_{12} \simeq N_{11}$) means an improvement in the measure.

In Fig. 6, we show the fractional error of x (left-hand panel) and the ratio of errors between two- and one-tracer case (right-hand panel) versus N_{12}/N_{11} for different S/N regimes: from $\log(S/N) = -0.5$ (black line) to $\log(S/N) = 2.0$ (orange line). Here, the same effect is observed. For high S/N (orange, pink and green lines), increasing N_{12} decreases the error, whereas for low S/N (black, red and blue lines) the error increases. Here, the effect on the error when $N_{12} \rightarrow N_{11}$ can be seen more clearly.

---

[3] Note that, the maximum value for N_{12} is $\sqrt{N_{22}N_{11}}$, as can be deduced from the Cauchy–Schwarz inequality: $|\langle \epsilon_1 \epsilon_2 \rangle|^2 \leq \langle \epsilon_1^2 \rangle \langle \epsilon_2^2 \rangle$.
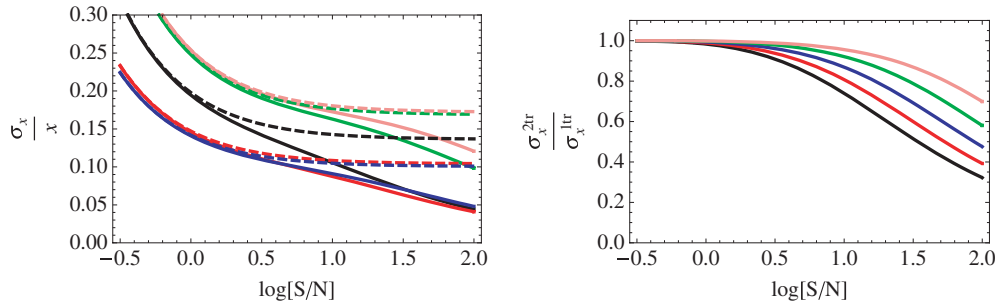
Figure 7. Left-hand panel: σ_x/x versus S/N. Right-hand panel: $\sigma_x^{2\text{tr}}/\sigma_x^{1\text{tr}}$ versus S/N. In the left-hand panel, the dashed-lines are the errors for the one-tracer model and the solid-lines for the two-tracer model. For both panels $Y = 1/20$ (pink line), $Y = 1/10$ (green line), $Y = 1$ (black line), $Y = 10$ (red line) and $Y = 20$ (blue line). $\tilde{b}_1 = 1$ and $\tilde{b}_2 = 2$ are assumed.

On the right-hand panels of Figs 5 and 6, for low S/N and for non-zero values of N_{12}, we have that $\sigma_x^{2\text{tr}}/\sigma_x^{1\text{tr}} > 1$. This effect is due to the fact that we are using equation (A12) to relate the noise elements of the two-tracer case with the single-tracer case. However equation (A12) only provides a correct relation among the diagonal noise matrix elements when the off-diagonal noise terms are zero, which is no longer the case. In fact, the noise for a single tracer built out of two tracers for which the noise matrix is strongly non-diagonal is not strictly Poisson. Therefore it cannot be fully described by the Poisson noise that it would have if the two tracers had a diagonal noise matrix (as equation A12 assumes). However, we have no other way to model it, and we thus stick to equation (A12). This effect is important only in the low S/N regime and where N_{12} is comparable to the diagonal terms ($N_{12} \lesssim N_{11}$)

2.7 Dependence on the relative number density

Finally it is also interesting to see how the relative number of tracers can affect the error of x. We vary the ratio between the number densities of tracers, $\bar{n}_1$ and $\bar{n}_2$, namely $Y \equiv \bar{n}_1/\bar{n}_2$, keeping the total number of tracers, $\bar{n}$, fixed. Again, we assume linear bias, Poisson noise and that the biases are $\tilde{b}_1 = 1$ and $\tilde{b}_2 = 2$.

In Fig. 7 (left-hand panel), we show the error of x versus S/N for $Y = 1/20$ (pink line), $1/10$ (green line), 1 (black line), 10 (red line) and 20 (blue line), for the one-tracer model (dashed lines) and for the two-tracer one (solid lines). In the right-hand panel, we show the ratio $\sigma_x^{2\text{tr}}/\sigma_x^{1\text{tr}}$ versus S/N for different values of Y using the same colour scheme.

Note that $Y > 1$ means that the highly biased tracer has the lower number density; for $Y < 1$, the highly biased tracer has the higher number density. From Fig. 7 (left-hand panel), we see that the error on the one-tracer model is as expected from Fig. 2: an increase in Y causes a reduction in σ_x/x, because it is equivalent to reducing the effective bias $\tilde{b}$ (recall that $\tilde{b}_1 < \tilde{b}_2$). On the other hand, for the two-tracer model, we observe that the fractional error is lower if the low bias tracer is more abundant than the high-bias one. This is also expected because as we have seen in Fig. 2: the lower the bias, the smaller is the error of x. In Fig. 7 (right-hand panel), we see that the maximal improvement for the two-tracer approach compared to the one-tracer approach is realized when the two tracers number densities are equal, independently of the S/N. For unequal number densities, the two-tracer approach gives a better improvement over the one-tracer approach, if the number density of the highly biased tracer is lower than that of the tracer with lower bias.

3 EXPECTED VALUES FOR PARAMETERS DESCRIBING BIAS AND STOCHASTICITY

As we have seen, the improvements achievable by using two tracers, depend on various features of the tracers population, such as the S/N, the bias parameters and the amount of bias non-linearity. In the next two sections, we explore what are plausible and realistic values if dark matter haloes are taken to be the tracers. We first use analytical arguments, and in the next section, N-body simulations.

3.1 Extended Press–Schechter approach

In this section, we identify dark matter haloes with the peaks of an initially Gaussian field, and compute their number densities and biases. We assume that the tracers (haloes) are linearly biased, as the non-linear corrections to halo bias derived in this frameworks are very small. The volume effect that may arise in this formalism is due to the z-dependence of the parameters, in particular, the bias. A narrow–deep survey has a strong z-dependence in the bias and a wide–shallow has a very weak one. It is because of this, that in this section we assume a volume-limited cubic survey of comoving side 1 Gpc h^{-1} for all z, and we perform the analysis at different values of z. This way it is easier to understand how the z-dependence of the bias affects the errors of the one- and two-tracer model. We set $k_{\min} = 2\pi/V_u^{1/3}$ and $k_{\max} = 0.1 D(0)/D(z)$ Mpc h^{-1}. $P^0(k)$ is given by CAMB for a standard Λ CDM universe. We consider a range of redshift between 0 and 4, and we parametrize the redshift dependence of f as, $f(z) = \Omega_m(z)^\gamma$, with $\gamma = 0.56$. We choose the two tracers to be haloes of masses 10^{12} M$_\odot$ $h^{-1} < M < 10^{13}$ M$_\odot$ h^{-1} for tracer 1 and 10^{13} M$_\odot$ $h^{-1} < M < 10^{14}$ M$_\odot$ h^{-1} for tracer 2.

The number density of these haloes is related to the halo mass function, given by

$$n(M, z) = \frac{2\rho_{\rm m}}{M\,\sigma(M)} f(\nu)\nu \left|\frac{{\rm d}\sigma(M)}{{\rm d}M}\right|, \tag{50}$$

where $\sigma(M)$ is the rms of the power spectrum linearly extrapolated at $z = 0$ filtered with a top-hat sphere of mass M, $\rho_{\rm m}$ is the mean density of the Universe (we set it at $\rho_{\rm m} = 7 \times 10^{10}\,{\rm M}_\odot\,h^{-1}\,{\rm Mpc}^{-3}$) and $\nu \equiv \delta_{\rm sc}^2(z)/\sigma^2(M)$. Here, $\delta_{\rm sc}(z)$ denotes the critical threshold for collapse and is given by $\delta_{\rm sc}(z) \simeq 1.686/D(z)$. We use the Sheth & Tormen (1999) mass function:

$$\nu f(\nu) = A(p)[1 + (q\nu)^{-p}]\left(\frac{q\nu}{2\pi}\right)^{1/2}\exp(-q\nu/2), \tag{51}$$

where $p = 0.3, q = 0.75$ and the normalization factor $A(p) = [1 + 2^{-p}\Gamma(1/2 - p)/\Gamma(1/2)]^{-1}$. Thus, the number density of objects is

$$\bar{n}(M_1, M_2, z) \equiv \int_{M_1}^{M_2} n(M, z)\,{\rm d}M. \tag{52}$$

Assuming Poisson noise, we can directly relate this to the noise matrix elements,

$$N_{ii}(z) = 1/\bar{n}(M_{1i}, M_{2i}, z). \tag{53}$$

As long as we are considering Poisson noise, the off-diagonal terms of the noise matrix are 0.

For the bias dependence we assume a linear bias ($\tilde{b} = \hat{b} \equiv b$). The bias of an object of mass M at redshift z is given by (Kaiser 1984; Mo, Jing & White 1997; Scoccimarro et al. 2001)

$$b(M, z) = 1 + \frac{1}{D(z)}\left[q\frac{\delta_{\rm sc}(z)}{\sigma^2(M)} - \frac{1}{\delta_{\rm sc}(z)}\right]. \tag{54}$$

The bias of a set of objects with masses between M_1 and M_2 is given by

$$\bar{b}(M_1, M_2, z) \equiv \frac{\int_{M_1}^{M_2} {\rm d}M\, n(M, z)b(M, z)}{\bar{n}(M_1, M_2, z)}. \tag{55}$$

In Fig. 8 (left-hand panel), we show how the bias of tracer 1 ($10^{12}\,{\rm M}_\odot\,h^{-1} < M < 10^{13}\,{\rm M}_\odot\,h^{-1}$ red-solid line), tracer 2 ($10^{13}\,{\rm M}_\odot\,h^{-1} < M < 10^{14}\,{\rm M}_\odot\,h^{-1}$ blue-solid line) and the combined sample (black-dashed line) change as a function of redshift. Bias increases with redshift roughly as $1/D^2(z)$. As there are many more lower mass haloes, the overall bias is dominated by this population. In Fig. 8 (right-hand panel), we show the S/N as a function of redshift. As before, we define ${\rm S/N}(z) \equiv \bar{b}^2(z)P^0(k = 0.1)D^2(z)[\bar{n}_1(z) + \bar{n}_2(z)]$, where $\bar{b}(z)$ corresponds to the bias of the whole sample. Note that $b^2(z)P^0(k = 0.1)D^2(z)$ goes roughly $\propto 1/D(z)^2$, but the number density of objects (the mass function) drops exponentially rapidly: the S/N decreases with increasing redshift; the maximum value of S/N is at $z = 0$, where ${\rm S/N} \simeq 15$.

In Fig. 9, we show the errors σ_x/x corresponding to the one- and two-tracer case (left-hand panel) and its ratio (right-hand panel) as a function of the redshift. In Fig. 9 (left-hand panel), we see that there is a minimum in the value of σ_x/x at $z \simeq 1.5$, both for the one- and for
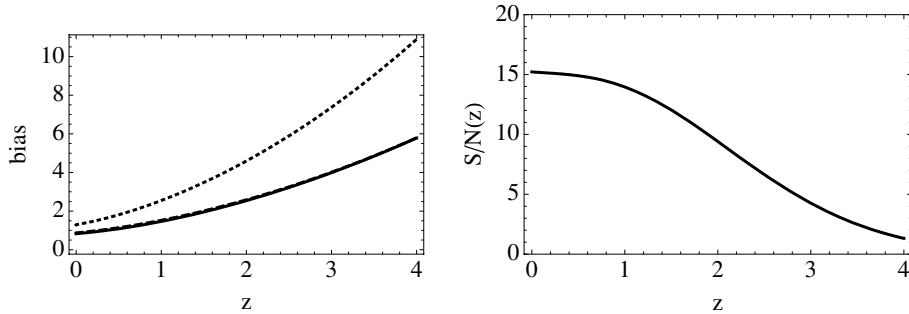

Figure 8. Left-hand panel: the bias as a function of redshift for tracer 1 ($10^{12}\,{\rm M}_\odot\,h^{-1} < M < 10^{13}\,{\rm M}_\odot\,h^{-1}$ solid line), tracer 2 ($10^{13}\,{\rm M}_\odot\,h^{-1} < M < 10^{14}\,{\rm M}_\odot\,h^{-1}$, dotted line) and for the whole sample (black-dashed line). Right-hand panel: S/N for the full sample (see text for definition) as a function of redshift.
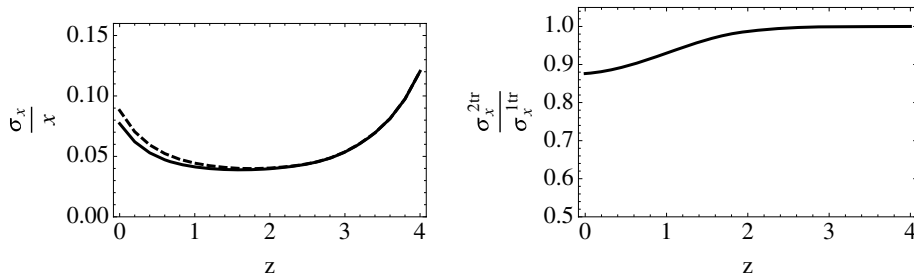

Figure 9. Left-hand panel: σ_x/x versus z for the one-tracer case (dashed line) and for the two-tracer case (solid line). Right-hand panel: $\sigma_x^{\rm 2tr}/\sigma_x^{\rm 1tr}$ versus z.

the two-tracer cases. This seems paradoxical (at least for the one-tracer model) because we have seen in Section 2.4 that as we increase the bias, the error on x increases. Furthermore, shot noise increases with redshift. The explanation of this seemingly paradoxical effect has to do with the fact that not only the noise and the bias change with z, but also $k_{max}(z)$ and $x(z)$. Recall that $x(z)$ is our signal and that the fractional error per interval of k goes like the square root of the number of modes, i.e. $k_{max}^3 - k_{min}^3$. It turns out that, as we increase the available range of k of equation (49), the error σ_x/x is reduced, scaling as these considerations indicate.

The last effect dominate at low z. At higher z, the noise and the bias effect dominate, and therefore σ_x/x increases with z as we have seen in Section 2.4. In Fig. 9 (right-hand panel), we see that only at low redshifts ($z < 1$), where the S/N is high, the improvement between the two cases is significant, reaching the minimum value of $\sigma_x^{2tr}/\sigma_x^{1tr} \simeq 0.88$ at $z = 0$. This result may seem worse than that we have shown in the last section for S/N $\simeq 15$. However, taking into account that now the ratio of biases is not 2, but $\simeq 1.5$ and also recalling that now $Y \neq 1$, the fractional error increases from $\simeq 0.7$ to $\simeq 0.88$.

From Fig. 9, it is clear that if the survey is dominated by low-redshift objects (wide-shallow surveys), then splitting the sample may reduce the error of x. However, the higher precision in measuring x is reached with surveys whose volume configuration results in most objects having $z \simeq 1.5$)

A mass selection of haloes as here may provide a modelling applicable to Sunyaev–Zel'dovich-selected clusters or, in an idealized way, to luminous red galaxies. But other type of surveys may select tracers in a radically different way; for example, emission-line-selected blue galaxies have a bias that evolves with redshift much more slowly than considered above. We therefore also consider a complementary, yet still highly idealized, way to select haloes and explore whether in that case the improvement in splitting the sample can be much larger. We select tracers by their peak height, i.e. their ν of equation (50), keeping the maximum and minimum ν (rather than the mass) of each tracer sample constant in redshift. We have explored different cuts and found the following. (i) The improvement in splitting the sample is maximized when the two samples have comparable number of objects. When, to maximize the bias difference between the two samples, the highly biased tracer include objects that are very rare, the shot noise for that sample become important and the gain in splitting the sample decreases. (ii) By suitably choosing the ν cuts, we have sampled the parameter space and managed to achieve $\sigma_x^{2tr}/\sigma_x^{1tr} \simeq 0.6$, but we have not been able to improve the gain further. For instance, choosing as tracer 1 structures with ν between 0.9 and 1.5 and as tracer 2 structures with ν between 1.5 and 20, we reach a gain of 0.6 at $z \simeq 2$.

Before any practical application of these findings, one should bear in mind that we have assumed a volume-limited sample (i.e. that all haloes in the required range are detected). In addition, we have considered a fixed survey volume seen at different redshift: for a given sky coverage, the volume per unit redshift increases with redshift for $z \lesssim 2.5$. Finally, we have selected haloes in a very idealized way. In practice, the selection will be likely applied on galaxies, which halo occupation distribution is not straightforward. We will discuss this further in Section 4.

3.2 Simulations

As we have seen in Figs 3 and 4, the gain of splitting the sample is dependent on the non-linearity parameters, specially for high S/N. In this case, Rs have to be very close to unity to exhibit substantial gain. In addition we have so far relied on the Poisson sampling assumption, which may not hold in details. For tracers that can be identified with dark matter haloes, these issues can be addressed by looking at N-body simulations.

We choose a flat ΛCDM cosmology with cosmological parameters consistent with current observational data. More specifically, the cosmological parameters of the simulation are $\Omega_m = 0.27$, $\Omega_\Lambda = 0.73$, $h = 0.7$, $\Omega_b h^2 = 0.023$, $n_s = 0.95$ and $\sigma_8 = 0.8$. Our cosmological simulation consists of 1024^3 particles in a volume of $(1\,\text{Gpc}\,h^{-1})^3$. This results in a particle mass of about $7 \times 10^{10}\,\text{M}_\odot\,h^{-1}$. The initial conditions of the simulation were generated at redshift $z = 65.67$, by displacing the particles according to the Zel'dovich approximation from their initial grid points. The initial power spectrum of the density fluctuations was computed by CAMB (Lewis et al. 2000).

Taking only the gravitational interaction into account, the simulation was performed with GADGET-2 (Springel 2005) using a softening length of comoving 30 kpc h^{-1} and a particle mesh grid size of 2048^3. The chosen mass resolution and force resolution enable us to resolve haloes with masses above $\simeq 10^{12}\,\text{M}_\odot\,h^{-1}$, i.e. each halo contains at least 15 particles. We identify haloes at redshift $z = 0$ by the friends-of-friends algorithm with a linking length of 0.2 times the mean interparticle separation. We split the haloes in two mass bins: $10^{12}\,\text{M}_\odot\,h^{-1} < M < 10^{13}\,\text{M}_\odot\,h^{-1}$ (M12) and $M > 10^{13}\,\text{M}_\odot\,h^{-1}$ (M13). The mass bins M12 and M13 consist of about 2.1 and 0.4 million haloes, respectively.

In order to derive the mean conditional bias (see equation 1), $b(\delta)$, for the two different tracers, we first compute the halo overdensity, δ_h, and matter overdensity, δ, by assigning the haloes and particles, respectively, on a 512^3 grid using the cloud in cell scheme. The overdensities are then further smoothed by a Gaussian filter, $\exp(-k^2 l_s^2/2)$, where we choose the smoothing length to be $l_s = 2.5, 5, 10$ and 20 Mpc h^{-1}. One expects that any stochasticity, non-linear or non-local effects on the bias relation should decrease as the field is smoothed with increasing smoothing lengths. The biased density as a function of the matter overdensity, $b(\delta)\delta$, is then computed by averaging the δ_h in the corresponding δ bin (see equation 1). Using the mean bias relation so obtained, we can compute the noise field ϵ on the grid by applying equation (5). After Fourier transforming the different fields using the same 512^3 grid, we can compute the power spectra and cross power spectra of the different quantities by spherical averaging the product of their Fourier modes, i.e. for example $P_{\epsilon\epsilon}(k) = \langle \epsilon(k)\epsilon^*(k) \rangle$ and $P_{hm}(k) = \langle \delta_h(k)\delta^*(k) \rangle$. Note that, in what follows, in evaluating the bias and R parameters from the simulations we have used explicitly equations (31–35), the noise terms are not assumed to be Poissonian but are computed directly from the ϵ field.
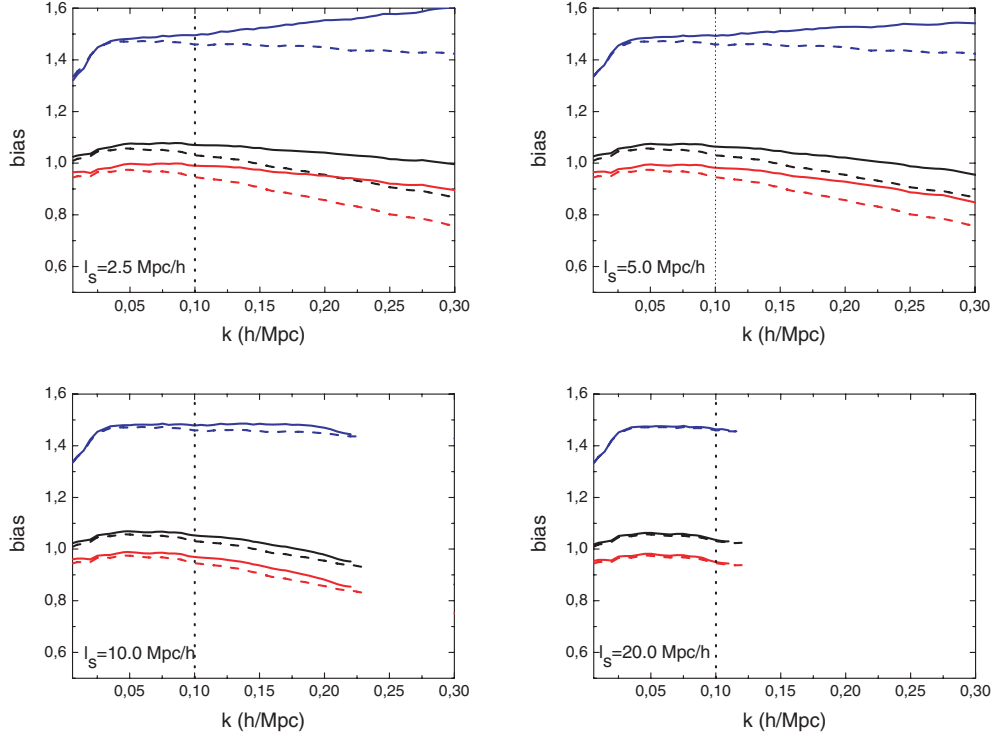
Figure 10. Bias parameters as a function of the scale k obtained from simulations. $\hat{b}$ are the dashed lines and $\tilde{b}$ the solid lines. The black lines correspond to the whole sample of haloes, whereas red lines corresponds to sample M12 ($10^{12}\,M_\odot < M < 10^{13}\,M_\odot$) and blue lines to sample M13 ($M > 10^{13}\,M_\odot$). The smoothing length is 2.5 for the top-left panel, 5.0 for the top-right panel, 10.0 for the bottom-left panel and 20.0 Mpc h^{-1} for the bottom-right panel.

In Fig. 10, we show the bias parameters $\hat{b}$ (dashed lines) and $\tilde{b}$ (solid lines) versus k, for the whole sample (black lines), for M12 (red lines) and for M13 (blue lines) for different smoothing lengths.

The vertical dotted line marks $k = 0.1\,h\,\mathrm{Mpc}^{-1}$, which is the typical scale at $z = 0$ where non-linearities appear. In this paper, we always work in the linear regime, where $k < 0.1\,D(0)/D(z)\,h\,\mathrm{Mpc}^{-1}$. From these plots, we can see that $\hat{b}$ is robust to changes in l_s and also is approximately scale-independent. $\tilde{b}$ is also very close to be scale invariant and varies little with the smoothing scale. In order to have a numerical reference, $\tilde{b}$ changes by about 7 per cent and $\hat{b}$ by about 5 per cent compared with their values for $l_s = 2.5\,\mathrm{Mpc}\,h^{-1}$, over a range from $k_{\min} = 0.01\,h\,\mathrm{Mpc}^{-1}$ to $k_{\max} = 0.1\,h\,\mathrm{Mpc}^{-1}$. For $l_s = 20\,\mathrm{Mpc}\,h^{-1}$, both parameters change by about 7 per cent in the same k range. On the other hand, for $k \simeq 0.1\,h\,\mathrm{Mpc}^{-1}$, $\tilde{b}$ changes by about 3 per cent as l_s goes from 2.5 to 20.0 Mpc h^{-1}; $\hat{b}$ changes less than 1 per cent in the same range.

In Fig. 11, we show the values of the non-linearity parameters versus k obtained from simulations: $R(k)$ (black line), $R_1(k)$ (red line), $R_2(k)$ (blue line) and $R_{12}(k)$ (green line). Different panels correspond to different smoothing scales l_s, as in Fig. 10. All non-linearity parameters decrease as the scale decreases, but this trend disappears as we increase the smoothing length. When the smoothing length is large enough ($\simeq 20\,\mathrm{Mpc}\,h^{-1}$), all R parameters are approximately scale-invariant and are very close to 1.

In Fig. 12 (left-hand panel), we show the different noise components obtained from simulations versus k, for different smoothing lengths: 2.5 (red lines), 5.0 (blue lines), 10.0 (green lines) and 20.0 Mpc h^{-1} (orange lines). The solid lines are $\langle \epsilon_1(\boldsymbol{k})\epsilon_1^*(\boldsymbol{k}) \rangle$ (M12), the dashed lines $\langle \epsilon_2(\boldsymbol{k})\epsilon_2^*(\boldsymbol{k}) \rangle$ (M13) and the dotted lines the cross terms between the two samples, $\langle \epsilon_1(\boldsymbol{k})\epsilon_2^*(\boldsymbol{k}) \rangle$. The noise terms for the whole sample are not shown for clarity. In order to get a better comparison between all these noise terms, we have removed the effects of the smoothing, dividing each noise by the filter squared.

The black lines are the noise predictions assuming a Poisson-like noise ($\langle \epsilon_i(\boldsymbol{k})\epsilon_i^*(\boldsymbol{k}) \rangle = 1/\bar{n}_i$). The solid line is for the M12 sample and the dashed line for M13. The cross term is relatively small (black dotted line). We observe that the M13 noise is sub-Poisson, whereas the one for M12 is super-Poisson. This is in agreement with the findings of Seljak et al. (2009). It has been noted before (Smith et al. 2007) that for massive haloes the noise could be sub-Poisson. At scales smaller than the ones of interest here, Smith et al. (2007) ascribe this to halo-exclusion effects. Noise above the Poisson level is expected if other sources of stochasticity affect halo formation. The formation and evolution of dark matter haloes is a highly complicated process: dark matter haloes grow through a mixture of smooth accretion, violent encounters and fragmentation. In the classical extended Press–Schechter/excursion set theory, haloes are identified with initial density peaks and the computation of the halo mass function (and thus as a derived quantity the halo bias) is mapped into a first passage process in the presence of a sharp barrier. This yields a deterministic halo bias, but cannot capture the full physical complications inherent to a realistic description of halo formation. In addition, numerical simulations show that there is not a good correspondence between peaks in the initial density field and collapsed haloes (see Katz, Quinn & Gelb 1993; Seljak & Warren 2004). Recently Maggiore & Riotto (2009) proposed to
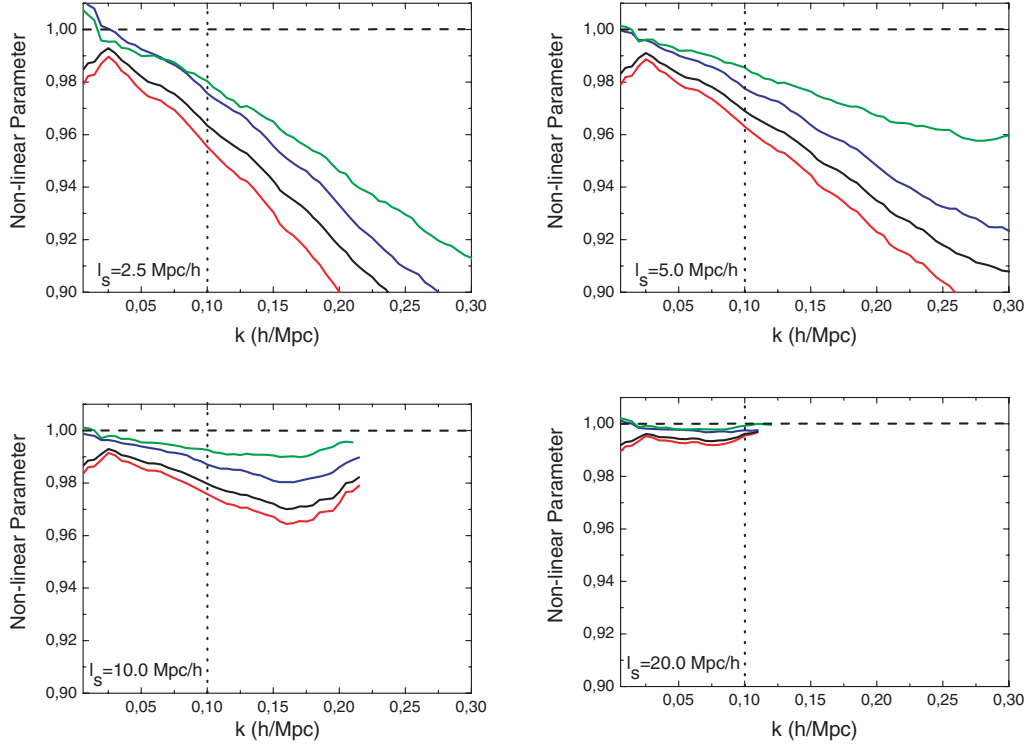
Figure 11. Non-linearity parameters, R (black line), $R_1(k)$ (red line), $R_2(k)$ (blue line) and $R_{12}(k)$ (green line), obtained from simulations for different smoothing lengths, as in Fig. 10. The subscripts 1 and 2 refer to M12 sample and M13 respectively, whereas the no subscript refers to the whole sample. The vertical dotted line, $k = 0.1\, h\,\mathrm{Mpc}^{-1}$, marks the limit of linear regime. The horizontal dashed line marks the maximum value for all Rs.
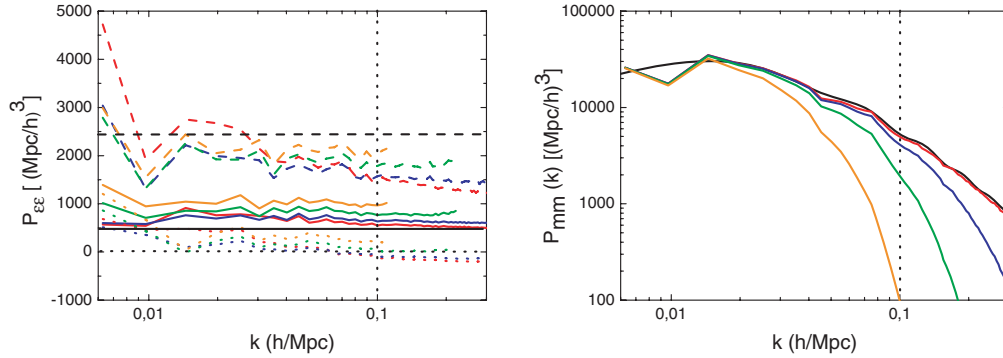

Figure 12. Noise terms (left-hand panel) and power spectrum (right-hand panel) as a function of the scale obtained from simulations. In the left-hand panel, $\langle \epsilon_1(k)\epsilon_1^*(k)\rangle$ (M12) solid lines, $\langle \epsilon_2(k)\epsilon_2^*(k)\rangle$ (M13) dashed lines and $\langle \epsilon_1(k)\epsilon_2^*(k)\rangle$ dotted lines. The black lines are the prediction for a Poisson-like noise. In the right-hand panel, $P_{\mathrm{mm}}(k)$ from CAMB (black line) and $P_{\mathrm{mm}}(k)$ from simulations (colours lines). In both panels, the colours refer to the smoothing length, 2.5 (red), 5.0 (blue), 10.0 (green) and 20.0 Mpc h^{-1} (orange).

include these effects, at least at an effective level, by taking into account that the critical value for collapse is not a fixed constant but itself a stochastic variable. This will naturally lead to an extra 'noise' component in the halo bias.

In the Fig. 12 (right-hand panel), we show the signal component, $P_{\mathrm{mm}}(k)$, obtained from simulations for different smoothing lengths. The colour notation is the same as that in left-hand panel. In this case, the black line is the linear theory prediction for the same cosmological parameters used in our simulations. The effect of the sampling variance can be clearly seen at large scales. The enhancement of the clustering at small scales due to the nonlinear gravitational evolution is hidden by the smoothing of the density field.

In order to apply the findings from simulations to our model, we make the following assumptions.

(i) The bias is scale-independent and does not change with the smoothing scale. We take $\tilde{b} = 1.05$ for the whole sample, $\tilde{b}_1 = 0.97$ for M12 and $\tilde{b}_2 = 1.47$ for the M13 sample. This assumption is supported by Fig. 10.

(ii) The non-linearity parameters have a linear dependence with scale k. For each smoothing length, we fitted the best linear relation up to $k = 0.1\, h\,\mathrm{Mpc}^{-1}$.

(iii) The noise is Poisson-like. Therefore the diagonal terms of the noise are $N_{ii}(k) = \alpha_i W^2(k \cdot l_s)/\bar{n}_i$, where α_i is a parameter that takes into account the deviations from the ideal Poisson noise (see Fig. 12, left-hand panel), and $W(k \cdot l_s)$ is the smoothing filter. The

Table 2. Parameters values used in plots of Fig. 13 as function of the smoothing scales used.

l_s in Mpc h^{-1}	2.5	5.0	10.0	20.0
$\tilde{b}$	1.05	1.05	1.05	1.05
$\tilde{b}_1$	0.97	0.97	0.97	0.97
$\tilde{b}_2$	1.47	1.47	1.47	1.47
$R(k)$	$1.0 - 0.39208k$	$0.99737 - 0.26854k$	$0.99666 - 0.1585k$	$0.99587 - 0.01972k$
$R_1(k)$	$0.99939 - 0.4033k$	$0.99591 - 0.31089k$	$0.99567 - 0.18733k$	$0.99479 - 0.02141k$
$R_2(k)$	$1.00752 - 0.29446k$	$1.00145 - 0.22144k$	$0.99941 - 0.11259k$	$0.99876 - 0.02276k$
$R_{12}(k)$	$0.99922 - 0.16926k$	$0.99883 - 0.12419k$	$0.99922 - 0.06535k$	$0.99905 - 0.01082k$
α	1.250	1.325	1.675	2.10
α_1	1.365	1.428	1.785	2.100
α_2	0.680	0.680	0.760	0.840
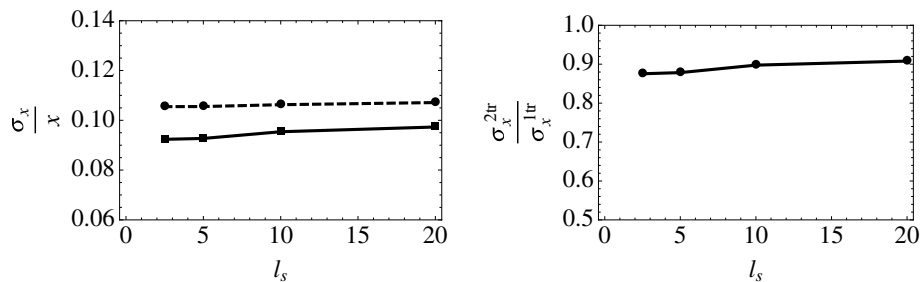
Figure 13. Left-hand panel: error of x as a function of the smoothing length l_s. The dashed line is for one-tracer model, whereas the solid line for the two-tracer one. Right-hand panel: ratio between the two models.

α values used are shown in Table 2. According to the number of haloes of the two tracers and the volume of the simulation, we have $\bar{n} = 2.5 \times 10^{-3} \, h^3/\mathrm{Mpc}^3$, $\bar{n}_1 = 2.1 \times 10^{-3} \, h^3/\mathrm{Mpc}^3$ and $\bar{n}_2 = 4.0 \times 10^{-4} \, h^3/\mathrm{Mpc}^3$.

(iv) We take the off-diagonal noise terms to be zero. This has some support from the simulations (Fig. 12, left-hand panel), where this term is $N_{12}/\sqrt{N_{11}N_{22}} < 0.2$

(v) Finally, we assume that the cross-correlation terms between the ϵ field and the matter field are 0.

We summarize all these assumptions in Table 2. Applying these conditions yield the results shown in Fig. 13. In the left-hand panel, the fractional errors for one- (dashed line) and two-tracer model (solid line) are shown as a function of the smoothing scale l_s, and the ratio of these two errors is shown in the right-hand panel.

The improvement between the two cases under these assumptions is rather modest: $\sigma_{2\mathrm{tr}}/\sigma_{1\mathrm{tr}} \simeq 0.9$. This is because it is mainly dominated by the ratio of biases – we have $\tilde{b}_2/\tilde{b}_1 \simeq 1.5$ – if we are in a region where S/N $\simeq 10$. This result is robust to small changes in the bias modelling, e.g. perfectly linear bias. We also have tried to fit the off-diagonal term to some non-zero value, but we have found that doing this does not produce any significant change in the plots of Fig. 13.

There may be some merit in splitting the sample in a different way; by choosing samples with very different biases, the gains should be larger, but in practice to do this almost certainly requires one sample to be of rare objects, which will have very high shot noise.

4 DISCUSSION & CONCLUSIONS

We have revisited the method of circumventing sample variance in the measurement of $f = \mathrm{d} \ln D/\mathrm{d} \ln a$ (D being the linear growth factor), based on comparing the clustering properties of two differently biased tracers of the dark matter distribution. This method was recently investigated by McDonald & Seljak (2009), although a similar technique in a different context was presented in Pen (2004). Along the same lines, Slosar (2009) and Seljak (2009) propose to compare clustering of differently biased tracers to circumvent sample variance in the measurement of primordial non-Gaussianity.

Most of the statistical power of these measurements comes from very large scales where cosmic variance is the dominant contribution to the statistical error-bars. By suppressing cosmic variance, this approach promises to reduce drastically error-bars on cosmologically very important quantities; for example it would allow for a high-precision determination of growth of structure as a function of redshift, as encoded in $f D$, and an improvement of dark energy figures of merit by large factors.

All these approaches assume that the observed objects (i.e. galaxies) trace the dark matter deterministically; the galaxy density field is assumed to be proportional to the dark matter field with the constant of proportionality given by a single parameter, the bias. This goes under the name of the linear bias model. An important underlying assumption is therefore made, that there is no stochasticity between the tracer field and the dark matter field on the scales of interest, which is expected to breakdown at some level, at least on small scales. While the linear bias model has been extremely successful in cosmology (e.g. Reid et al. 2010 and references therein), it is well known that the linear bias

model might provide a good description for the galaxy power spectrum even if the relation between the galaxy and dark matter overdensities is not that of a linear bias (e.g. Heavens, Matarrese & Verde 1998).

Galaxies are believed to form inside dark matter haloes, but their formation probability as function of halo mass and their exact radial distribution is still the subject of active research. The process of halo occupation by galaxies is expected to be stochastic to some extent, but the details of the galaxy distribution within haloes are expected to become increasingly unimportant on large scales. Here, we simplify the issue by assuming (possibly with an over-simplification) that dark matter haloes can be used as tracers. The linear deterministic bias model however is known not to be a perfect description of halo clustering and that the relation between dark matter and haloes and between haloes of different masses is stochastic. For example, Seljak & Warren (2004) point out that 'the fluctuations between haloes and the initial or final matter fields are never below 10–20 per cent' and that 'the scatter between the fields in individual modes is significant and one cannot assume that the fields are simply proportional to one another'. This was further explored and quantified by Bonoli & Pen (2009). Note that the halo overdensity field is expected to have a stochastic component even if it was a perfect Poisson sampling of a linearly biased dark matter field, but the above references and *N*-body simulations show that there are additional sources of stochasticity beyond shot noise.

We have thus set out to generalize the approach of McDonald & Seljak (2009), by assuming that the bias of haloes may not be perfectly linear and allowing for some stochasticity. We have computed the expected error on the quantity fD achievable by comparing clustering of differently biased tracers (thus suppressing cosmic variance) and by combining the different tracers in a single sample (thus reducing shot noise and stochasticity but carrying along sample variance in full).

We have analysed how the bias, the noise, the non-linearity and stochasticity affect the measurements of fD and explored in which S/N regime it is significantly advantageous to split a galaxy sample in two differently biased tracers. We used results from simulations to set plausible values for these parameters to see how great the gains may be in practice. We find that even small amount of stochasticity (either in the form of Poisson noise or in more general form) and of non-linearity can limit significantly the performance of the two-tracers approach. In our analysis, we also have assumed a scale-independent bias. This may be enough, for the mass range studied, if we consider only dark matter haloes as tracers. This is indeed what we have seen in simulations. On the other hand, it is also true that more realistic approaches, which account for galaxies as tracers instead of haloes, should include a scale-dependent bias. However, including this in our formalism can only reduce the gain achievable by splitting the sample. We expect the ratio of errors increases from the current value of 0.9 to even closer values to 1 if the bias is strongly scale-dependent.

We have shown that only in the very high S/N regime it is significantly advantageous to split the sample and that, even though the gain is maximized by increasing the *ratio* of the biases of the two tracers, both tracers should be well sampled. We have explored different ways of selecting and splitting dark matter haloes obeying a Λ CDM mass function and found that one can achieve up to a 40 per cent reduction of the error on fD. While this would correspond to the gain from a three times larger survey volume if the two tracers were not to be split, it is much smaller that the improvement forecasted in the absence of stochasticity and bias non-linearity.

In addition, we should note that these findings apply to dark matter haloes as tracers, while realistic surveys would select galaxies: the galaxy–host halo relation is likely to introduce extra stochasticity which would reduce the gain further. The formalism we have developed, however, is general enough that can be used to optimize survey design and tracers selection and optimally split (or combine) tracers to minimize the error on the cosmologically interesting quantities.

REFERENCES

Bacon D. J. et al., 2005, MNRAS, 363, 723
Bonoli S., Pen U. L., 2009, MNRAS, 396, 1610
Davis M., Peebles P. J. E., 1983, ApJ, 267, 465
Dekel A., Lahav O., 1999, ApJ, 520, 24
Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23
Feldman H. A., Frieman J. A., Fry J. N., Scoccimarro R., 2001, Phys. Rev. Lett., 86, 1434
Fisher R. A., 1935, J. R. Statistical Soc., 98, 39
Fry J. N., 1994, Phys. Rev. Lett., 73, 215
Guzzo L. et al., 2008, Nat, 451, 541
Hamilton A. J. S., 1998, ed., Astrophys. Space Sci. Library, Vol. 231, The Evolving Universe. Kluwer, Dordrecht, p. 185
Hawkins E. et al., 2003, MNRAS, 346, 78
Heavens A. F., Matarrese S., Verde L., 1998, MNRAS, 301, 797
Hoekstra H., van Waerbeke L., Gladders M. D., Mellier Y., Yee H. K. C., 2002, ApJ, 577, 604
Kaiser N., 1984, ApJ, 284, L9
Kaiser N., 1987, MNRAS, 227, 1

Katz N., Quinn T., Gelb J. M., 1993, MNRAS, 265, 689
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Maggiore M., Riotto A., 2009, preprint (arXiv:0903.1250)
McDonald P., Seljak U., 2009, J. Cosmol. Astropart. Phys., 10, 7
Mo H. J., Jing Y. P., White S. D. M., 1997, MNRAS, 284, 189
Peebles P. J. E., 1980, The Large-Scale Structure of the Universe. Princeton University Press, Princeton, NJ, p. 435
Pen U.-L., 2004, MNRAS, 350, 1445
Reid B. A. et al., 2010, MNRAS, 404, 60
Scoccimarro R., Sheth R. K., Hui L., Jain B., 2001, ApJ, 546, 20
Seljak U., 2009, Phys. Rev. Lett., 102, 021302
Seljak U., Warren M. S., 2004, MNRAS, 355, 129
Seljak U. et al., 2005, Phys. Rev. D, 71, 043511
Seljak U., Hamaus N., Desjacques V., 2009, Phys. Rev. Lett., 103, 091303
Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
Slosar A., 2009, J. Cosmol. Astropart. Phys., 3, 4
Smith R. E., Scoccimarro R., Sheth R. K., 2007, Phys. Rev. D, 75, 063512
Springel V., 2005, MNRAS, 364, 1105
Tegmark M., Taylor A. N., Heavens A. F., 1997, ApJ, 480, 22
Verde L. et al., 2002, MNRAS, 335, 432

APPENDIX A: RELATIONS BETWEEN ONE-TRACER CASE AND TWO-TRACER CASE

The parameters $R(r), R_1(r), R_2(r)$ and $R_{12}(r)$ are not fully independent, the same is true for the set of parameters $\tilde{\beta}(r)$, $\tilde{\beta}_2(r)$ and $\tilde{\beta}_1(r)$, and also the different noise matrices. Here we make explicit their relation.

A1 Relation between the *R* parameters

Let the tracer (galaxies, haloes...) overdensity $\delta_g(\boldsymbol{x})$ be defined as

$$\delta_g(\boldsymbol{x}) \equiv \frac{\rho_g(\boldsymbol{x})}{\bar{\rho}_g} - 1 = \frac{n_g(\boldsymbol{x})}{\bar{n}_g} - 1, \tag{A1}$$

where $\rho_g(\boldsymbol{x})$ is the tracer density at $\boldsymbol{x}$. In the second equality, we have used the fact that $\rho_g(\boldsymbol{x}) \propto n_g(\boldsymbol{x})$ with $n_g(\boldsymbol{x})$ the number of tracers at $\boldsymbol{x}$. The total number density of galaxies is $n_g(\boldsymbol{x}) = n_{g1}(\boldsymbol{x}) + n_{g2}(\boldsymbol{x})$. Defining the ratio of number of galaxies as

$$Y \equiv \frac{\bar{n}_1}{\bar{n}_2}, \tag{A2}$$

we can write the overdensity of galaxies as

$$\delta_g(\boldsymbol{x}) = \frac{\delta_{g1}(\boldsymbol{x})}{1 + Y^{-1}} + \frac{\delta_{g2}(\boldsymbol{x})}{1 + Y}. \tag{A3}$$

Recalling the definitions of $R(r), R_{12}(r), R_1(r), R_2(r)$ we find that

$$R(r) = \frac{R_1(r)Y\tilde{\beta}_2(r) + R_2(r)\tilde{\beta}_1(r)}{\sqrt{\tilde{\beta}_2^2(r)Y^2 + \tilde{\beta}_1^2(r) + 2R_{12}(r)Y\tilde{\beta}_1(r)\tilde{\beta}_2(r)}}. \tag{A4}$$

Note that, as expected, when $Y \to 0, R(r) \to R_2(r)$ and when $Y \to \infty, R(r) \to R_1(r)$.

This last equation is also valid in k-space according to the definitions of $\tilde{\beta}_i(k)$, $\tilde{b}_i(k)$, $R_i(k)$ and $R_{12}(k)$. This is because the definitions of all these parameters are mathematically symmetric in Fourier and configuration space.

A2 Relation between the $\tilde{\beta}$ parameters

Similarly, we can derive the relation between $\tilde{\beta}(r)$, $\tilde{\beta}_1(r)$ and $\tilde{\beta}_2(r)$. Recalling the definition of $\tilde{\beta}$s as,

$$\tilde{\beta}_i(r) \equiv \frac{f}{\tilde{b}_i(r)} \qquad \tilde{\beta}(r) \equiv \frac{f}{\tilde{b}(r)} \tag{A5}$$

then the relation between $\tilde{\beta}$s is

$$\tilde{\beta}_i(r) = \tilde{\beta}(r) \left[\frac{\tilde{b}^2(r)}{\tilde{b}_i^2(r)} \right]^{1/2}. \tag{A6}$$

From this we obtain

$$\tilde{\beta}^{-2}(r) = \frac{\tilde{\beta}_1^{-2}(r)}{(1 + 1/Y)^2} + \frac{\tilde{\beta}_2^{-2}(r)}{(1 + Y)^2} + \frac{2R_{12}(r)\tilde{\beta}_1^{-1}(r)\tilde{\beta}_2^{-1}(r)}{(1 + Y)(1 + 1/Y)}. \tag{A7}$$

Again, this relation is valid for both configuration space and k-space because of reasons of symmetry in the definitions of the parameters. Clearly, we can also write down a relation between the $\tilde{b}$ parameters:

$$\tilde{b}^2 = \frac{\tilde{b}_1^2}{(1 + 1/Y)^2} + \frac{\tilde{b}_2^2}{(1 + Y)^2} + \frac{2R_{12}\tilde{b}_1\tilde{b}_2}{(1 + Y)(1 + 1/Y)}. \tag{A8}$$

A3 Relation between the noise terms

Let the noise matrix for one-tracer model be

$$N_{1\mathrm{tr}} = N \tag{A9}$$

and for the two-tracer model,

$$N_{2\mathrm{tr}} = \begin{pmatrix} N_{11} & N_{12} \\ N_{12} & N_{22} \end{pmatrix} \tag{A10}$$

both in k-space.

Assuming Poisson noise terms, and distinct populations, the off-diagonal terms are zero, $N_{12} = 0$. Setting $\bar{n}_1$ and $\bar{n}_2$ to be the number density of galaxies of type 1 and 2 respectively, and $\bar{n} = \bar{n}_1 + \bar{n}_2$ the total number of galaxies, we can say that

$$(N_{11})^{-1} = \bar{n}_1 \qquad (N_{22})^{-1} = \bar{n}_2 \qquad (N)^{-1} = \bar{n}. \tag{A11}$$

Finally, we obtain

$$N_{11} = N(1 + Y^{-1}) \qquad N_{22} = N(1 + Y). \tag{A12}$$

A4 Constraints between the non-linearity parameters

Given the definitions of the non-linearity coefficients, using the Cauchy–Schwarz inequality we find that

$$-1 \leq R(r), R_1(r), R_2(r), R_{12}(r) \leq 1. \tag{A13}$$

However, the negative values for these Rs parameters represent a negative bias for tracers relative to the dark matter, with a doubtful physical connection. For this reason, we restrict the possible values for these parameters to the range

$$0 \leq R(r), R_1(r), R_2(r), R_{12}(r) \leq 1. \tag{A14}$$

The parameters $R_1(r), R_2(r)$ and $R_{12}(r)$ are not totally independent but are related by the condition of equation (B11),

$$1 - R_1^2(r) - R_2^2(r) - R_{12}^2(r) + 2R_1(r)R_2(r)R_{12}(r) \geq 0. \tag{A15}$$

If we isolate $R_{12}(r)$ as a function of $R_1(r)$ and $R_2(r)$, the last equation becomes

$$R_1(r)R_2(r) - \sqrt{R_1^2(r)R_2^2(r) + 1 - R_1^2(r) - R_2^2(r)} \leq R_{12}(r) \leq R_1(r)R_2(r) + \sqrt{R_1^2(r)R_2^2(r) + 1 - R_1^2(r) - R_2^2(r)}. \tag{A16}$$

This equation is enough if we are working only with the two-tracer model. However, if we want to compare this model with the one-tracer model, we have to make sure that also the $R(r)$ parameter is between 0 and 1 (see equation A4):

$$0 \leq \frac{R_1(r)Y\tilde{\beta}_2(r) + R_2(r)\tilde{\beta}_1(r)}{\sqrt{\tilde{\beta}_2^2(r)Y^2 + \tilde{\beta}_1^2(r) + 2R_{12}(r)Y\tilde{\beta}_1(r)\tilde{\beta}_2(r)}} \leq 1. \tag{A17}$$

Since $R_1(r), R_2(r), Y$ and $\tilde{\beta}_i(r)$ are always positive, the first inequality always holds, and we find that

$$R_{12}(r) \geq \frac{(R_1(r)Y\tilde{\beta}_2(r) + \tilde{\beta}_1(r)R_2(r))^2 - Y^2\tilde{\beta}_2^2(r) - \tilde{\beta}_1^2(r)}{2Y\tilde{\beta}_1(r)\tilde{\beta}_2(r)} \tag{A18}$$

to satisfy the second.

This minimum value could be lower or higher than the one given by the equation (A16) depending on the values of the other parameters. Therefore, the limits for $R_{12}(r)$ are

$$R_{12}(r) \geq \max\left\{ R_1(r)R_2(r) - \sqrt{R_1^2(r)R_2^2(r) + 1 - R_1^2(r) - R_2^2(r)}, \frac{[R_1(r)Y\tilde{\beta}_2(r) + \tilde{\beta}_1(r)R_2(r)]^2 - Y^2\tilde{\beta}_2^2(r) - \tilde{\beta}_1^2(r)}{2Y\tilde{\beta}_1(r)\tilde{\beta}_2(r)} \right\}$$

$$R_{12}(r) \leq R_1(r)R_2(r) + \sqrt{R_1^2(r)R_2^2(r) + 1 - R_1^2(r) - R_2^2(r)}. \tag{A19}$$

As we said before, also this last equation is valid for k-space parameter because of reasons of symmetry in the definitions.

APPENDIX B: CONSTRAINTS ON THE NON-LINEAR COEFFICIENTS

Suppose there are three possibly correlated fields, x, y, z (in our application these would correspond to $\delta_{g1}, \delta_{g2}, \delta_m$), and the corresponding non-linear coefficients are

$$r_1^2 = \frac{\langle xz \rangle^2}{\langle x^2 \rangle \langle z^2 \rangle}, \tag{B1}$$

$$r_2^2 = \frac{\langle yz \rangle^2}{\langle y^2 \rangle \langle z^2 \rangle}, \tag{B2}$$

$$r_3^2 = \frac{\langle xy \rangle^2}{\langle x^2 \rangle \langle y^2 \rangle}. \tag{B3}$$

Using the Cauchy–Schwarz inequality, we can state that

$$0 \leq r_i^2 \leq +1. \tag{B4}$$

We want to know what are the constraints on the triplet r_1, r_2 and r_3. To solve this, consider

$$C \equiv \langle (x + \lambda y + \mu z)^2 \rangle \geq 0. \tag{B5}$$

This is at least zero for all λ and μ, and in particular for the values which minimize C, namely λ' and μ':

$$\frac{\partial C}{\partial \lambda}\bigg|_{\lambda', \mu'} = \langle (x + \lambda' y + \mu' z)y \rangle = 0, \tag{B6}$$

$$\frac{\partial C}{\partial \mu}\bigg|_{\lambda', \mu'} = \langle (x + \lambda' y + \mu' z)z \rangle = 0. \tag{B7}$$

The system has an unique solution if and only if $r_1^2 \leq 1$. In that case, the values are

$$\lambda' = \frac{1}{D} \left(\langle yz \rangle \langle xz \rangle - \langle xy \rangle \langle z^2 \rangle \right) \tag{B8}$$

$$\mu' = \frac{1}{D} \left(\langle y^2 \rangle \langle xz \rangle - \langle xz \rangle \langle yz \rangle \right), \tag{B9}$$

where $D = \langle y^2 \rangle \langle z^2 \rangle - \langle yz \rangle^2$. Substituting these values into equation (B5) for C gives

$$(1 - r_1^2)(1 - r_1^2 - r_2^2 - r_3^2 + 2r_1 r_2 r_3) \geq 0. \tag{B10}$$

Provided that $r_1^2 \neq 1$, we can write

$$1 - r_1^2 - r_2^2 - r_3^2 + 2r_1 r_2 r_3 \geq 0. \tag{B11}$$

Because of symmetry reasons we can say that this last equation holds if at least one of the rs is different from ± 1. Note that if two of the rs are equal to one, so is the third. In our application, r_1, r_2, r_3 correspond to R_1, R_2, R_{12}.

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