The impact of systematic uncertainties in *N*-body simulations on the precision cosmology from galaxy clustering: a halo model approach

Hao-Yi Wu⋆ and Dragan Huterer⋆

Department of Physics, University of Michigan, 450 Church St, Ann Arbor, MI 48109-1040, USA

ABSTRACT

Dark matter *N*-body simulations provide a powerful tool to model the clustering of galaxies and help interpret the results of galaxy redshift surveys. However, the galaxy properties predicted from *N*-body simulations are not necessarily representative of the observed galaxy populations; for example, theoretical uncertainties arise from the absence of baryons in *N*-body simulations. In this work, we assess how the uncertainties in *N*-body simulations impact the cosmological parameters inferred from galaxy redshift surveys. Applying the halo model framework, we find that the velocity bias of galaxies in modelling the redshift-space distortions is likely to be the predominant source of systematic bias. For a deep, wide survey like BigBOSS, current 10 per cent uncertainties in the velocity bias limit k_{max} to $0.14 \, h \, \mathrm{Mpc}^{-1}$. In contrast, we find that the uncertainties related to the density profiles and the galaxy occupation statistics lead to relatively insignificant systematic biases. Therefore, the ability to calibrate the velocity bias accurately – from observations as well as simulations – will likely set the ultimate limit on the smallest length scale that can be used to infer cosmological information from galaxy clustering.

Key words: cosmological parameters – dark energy – dark matter – large-scale structure of Universe.

1 INTRODUCTION

The large-scale distribution of galaxies has been used to probe the structure and composition of the universe for over three decades. From the pioneering analyses of the Lick catalogue (Groth & Peebles 1977) and the CfA Redshift Survey (Huchra et al. 1983; Geller & Huchra 1989) revealing the cosmic web, the APM Galaxy Survey hinting the departure from the standard cold dark matter (CDM) model (Maddox et al. 1990) to the subsequent 2dF Galaxy Redshift Survey (Colless et al. 2001), the Sloan Digital Sky Survey (SDSS; York et al. 2000) and the VIMOS-VLT Deep Survey (Le Fèvre et al. 2005), galaxy redshift surveys have revolutionized the view of the large-scale structure of the universe. Recently, the WiggleZ Dark Energy Survey (Drinkwater et al. 2010) and the SDSS-III Baryon Oscillation Spectroscopic Survey (BOSS; Schlegel, White & Eisenstein 2009) have measured the galaxy clustering to unprecedented precision and provided stringent constraints on the cosmological parameters.

One of the most important features in the galaxy clustering is the baryon acoustic oscillations (BAO), originating from the waves in the primordial electron–photon plasma before the recombination. The sound horizon at the end of recombination is manifested as a peak in the real-space two-point correlation function or as wiggles in the Fourier-space power spectrum. This characteristic scale of BAO is considered as a standard ruler of the different evolution stages of the universe, and as a dark energy probe with relatively well-controlled systematics (Blake & Glazebrook 2003; Seo & Eisenstein 2003). Indeed, since its discovery (Miller, Nichol & Batuski 2001; Cole et al. 2005; Eisenstein et al. 2005), BAO has been providing ever improving constraints on cosmological parameters (e.g. Percival et al. 2010; Blake et al. 2011; Anderson et al. 2012).

Beyond the BAO feature, the full scale dependence of the clustering of galaxies contains much more information and can be used to constrain cosmology (e.g. Tegmark et al. 2006; Reid et al. 2010; Tinker et al. 2012; Cacciato et al. 2013) and the halo occupation statistics (e.g. Abazajian et al. 2005; Tinker et al. 2005; van den Bosch et al. 2007; Zheng & Weinberg 2007; Zehavi et al. 2011). From the perspective of power spectrum $P(k)$, the number of modes increases as k^3, and the information content increases dramatically as one goes to smaller scales. However, when one tries to draw information from high k, especially at low redshift, the density perturbations become non-linear and difficult to model (e.g. Smith et al. 2003; Heitmann et al. 2010; Jennings, Baugh & Pascoli 2011), which can introduce significant systematic errors in the recovered cosmological parameters (e.g. de la Torre & Guzzo 2012; Smith et al. 2012).

⋆E-mail: hywu@umich.edu (H-YW); huterer@umich.edu (DH)

The analysis of galaxy clustering often relies on *N*-body simulations and synthetic galaxy catalogues to model the non-linearity on small scales, as well as to estimate the cosmic and sample covariances. For example, the WiggleZ team has validated their model for the non-linear galaxy power spectrum using the GiggleZ Simulation[1] (Parkinson et al. 2012), while synthetic galaxy catalogues based on the Large Suite of Dark Matter Simulations (LasDamas[2]) have been used in the galaxy clustering analysis of SDSS (Chuang & Wang 2012; Xu et al. 2012).

For upcoming surveys, synthetic catalogues generated from *N*-body simulations will likely be routinely used to calibrate galaxy surveys. However, *N*-body simulations are not free from systematics. In *N*-body simulations, galaxies are assigned to haloes or dark matter particles based on models such as halo occupation distribution (HOD; Peacock & Smith 2000; Scoccimarro et al. 2001; Berlind & Weinberg 2002), abundance matching (Kravtsov et al. 2004; Vale & Ostriker 2004) or semi-analytic models (White & Frenk 1991; Kauffmann, White & Guiderdoni 1993; Somerville & Primack 1999; Cole et al. 2000). The galaxy populations predicted by simulations can be affected by intensive stripping in dense environment (e.g. Wetzel & White 2010) and the absence of baryons (e.g. Weinberg et al. 2008; Simha et al. 2012). On the other hand, when one uses dark matter particles to model the behaviour of galaxies, systematic errors may arise because the positions and velocities of galaxies do not necessarily follow those of dark matter particles (e.g. Wu et al. 2013a). Hydrodynamical simulations that include proper treatments of baryonic physics can be another avenue to predict the properties of galaxies more reliably; however, because these simulations are more computationally intensive, it is not yet practical to use them to achieve the statistics and high resolution required by upcoming large surveys.

In addition, it has been shown that galaxies predicted from *N*-body simulations cannot recover the spatial distribution of observed galaxies. For example, Wu et al. (in preparation) have shown that in high-resolution *N*-body simulations of galaxy clusters, subhaloes tend to be prematurely destroyed and fail to predict the location of galaxies (also see Appendix A). The need to include 'orphan galaxies' (galaxies not associated with subhaloes in simulations) to improve the completeness of predicted galaxies has been frequently addressed in the community (e.g. Gao et al. 2004; Wang et al. 2006; Guo et al. 2011); however, even including orphan galaxies does not lead to consistent galaxy clustering at all scales. For example, Guo et al. (2011) have shown that the galaxy population generated using the semi-analytic model applied to the Millennium Simulations overestimates the small-scale clustering (also see Contreras et al. 2013).

In this paper, we examine the impact of the systematics in *N*-body simulations on the predictions of galaxy clustering. We calculate the galaxy power spectrum based on the halo model, with inputs from the results of recent *N*-body simulations. We use the information of the full power spectrum of galaxies to forecast the cosmological parameter constraints and determine at which scale these systematics start to become relevant. We specifically explore how these uncertainties will limit our ability to utilize the cosmological information from small scale.

This paper is organized as follows. In Section 2, we review the halo model prediction for galaxy power spectrum. In Section 3, we present our fiducial assumptions and discuss the information

content associated with $P(k)$. Section 4 explores the self-calibration of HOD parameters. Section 5 addresses the impact of the uncertainties in the halo mass function on the cosmological constraints from galaxy clustering. Section 6 focuses on various systematics associated with the properties of galaxies in dark matter haloes in *N*-body simulations and presents the required control of these sources of systematic error. We conclude in Section 7. In Appendix A, we present the galaxy number density profile model used in this work. In Appendix B, we provide a detailed derivation of the galaxy power spectrum based on the halo model. In Appendix C, we derive the power spectrum covariance.

2 HALO MODEL AND GALAXY POWER SPECTRUM: A REVIEW

Throughout this work, we use the power spectrum of galaxies $P(k)$ as our clustering statistic. Possible alternatives include the three-dimensional correlation function $\xi(r)$ and its two-dimensional analogue – the angular two-point function $w(\theta)$ or the projected two-point function $w_p(r_p)$. While the Fourier-space power is more difficult to measure from the galaxy distribution, it is 'closest to theory' in the sense that the other aforementioned quantities are weighted integrals over $P(k)$. Therefore, it is easiest to see the effect of the uncertainties in theoretical modelling by using the power spectrum. While these different functions measured in a given galaxy survey contain the same information in principle, in data analysis sometimes discrepancies occur (e.g. Anderson et al. 2012).

2.1 Basic model

In this section, we provide the key equations of the galaxy power spectrum derived from the halo model, following Scherrer & Bertschinger (1991), Seljak (2000) and Cooray & Sheth (2002). The detailed derivation is provided in Appendix B.

The halo model assumes that all galaxies are inside dark matter haloes. To model the distribution of galaxies, we need the following distributions.

(i) Statistics and spatial distribution of dark matter haloes:

(a) Halo mass function, dn/dM, the number density of haloes as a function of the halo mass.
(b) Halo bias, $b^2(M) = P_{hh}(k)/P_{lin}(k)$, where P_{hh} is the power spectrum of haloes and P_{lin} is the linear matter power spectrum. We limit our use of $b(M)$ to large scales where $b(M)$ is scale independent.

(ii) Statistics and spatial distribution of galaxies in a halo:

(a) HOD function, $P(N|M)$, the probability distribution function of the number of galaxies in a halo of a given mass. The number of galaxies N is further split into the contribution from central galaxies N_{cen} (0 or 1) and from satellite galaxies N_{sat}.
(b) Galaxy number density profile, $u(r|M)$, the radial dependence of the galaxy number density inside a halo of a given mass. We normalize u such that $\int u(r|M)\, d^3r = 1$. We also use the density profile in Fourier space, $\tilde{u}(\boldsymbol{k}|M) = \int d^3\boldsymbol{x}\, u(\boldsymbol{x}|M)e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}$, and $\tilde{u} \to 1$ for small k.

The mean galaxy number density is given by

$$\bar{n}_{gal} \equiv \langle n_{gal} \rangle = \int dM \frac{dn}{dM} \langle N|M \rangle . \tag{1}$$

The power spectrum is contributed by two galaxies in two different haloes (the two-halo term, P_{gg}^{2h}) and two galaxies in the same halo

(the one-halo term, P_{gg}^{1h}):

$$P(k) = P_{gg}^{1h}(k) + P_{gg}^{2h}(k) \tag{2}$$

$$P_{gg}^{2h}(k) = \left[\frac{1}{\bar{n}_{gal}} \int dM \frac{dn}{dM} \langle N|M \rangle b(M) \tilde{u}(k|M) \right]^2 P_{lin}(k) \tag{3}$$

$$P_{gg}^{1h}(k) = \frac{1}{\bar{n}_{gal}^2} \int dM \frac{dn}{dM} \left\langle \binom{N}{2} \middle| M \right\rangle f(k|M). \tag{4}$$

Here $\langle X|M \rangle$ indicates the average value of quantity X at a given halo mass M. In the one-halo term,

$$\left\langle \binom{N}{2} \middle| M \right\rangle f(k|M) = \left[\langle N_{sat}|M \rangle \tilde{u}(k|M) \right.$$
$$\left. + \frac{1}{2} \langle N_{sat}(N_{sat} - 1)|M \rangle |\tilde{u}(k|M)|^2 \right],$$

which takes into account the contribution from central–satellite and satellite–satellite pairs (Berlind & Weinberg 2002).

2.2 Redshift-space distortions

In observations, one cannot recover the exact three-dimensional spatial distribution of galaxies, because the redshifts of galaxies are impacted by their motions due to the local gravitational field and do not reflect their true distances. On larger scales, galaxies tend to move towards high-density regions along filaments, and these motions tend to squash the galaxy distribution along the line of sight and boost the clustering, a phenomenon known as the Kaiser effect (Kaiser 1987). On small scales, the virial motions of galaxies inside a halo tend to make the galaxy distribution in the redshift space elongated along the line of sight, causing the so-called Fingers-of-God effect and reducing the small-scale power. In this section, we briefly describe the model we use for the redshift-space distortions (RSD) for $P(k)$, following Seljak (2001), White (2001) and Cooray & Sheth (2002). We adopt one of the simplified models – assuming the velocity distribution function to be Gaussian – and note that the improvement of the RSD model is currently an active research area.

Since the one-halo term involves the halo scale, we only consider the virial motions of galaxies inside a halo, which can be modelled as (Peacock 1999)

$$\tilde{\delta}_{gal}^z(\boldsymbol{k}) = \tilde{\delta}_{gal}(\boldsymbol{k}) e^{-\frac{1}{2}[k\sigma_v(M)\mu]^2}, \tag{5}$$

where $\tilde{\delta}_{gal}^z$ and $\tilde{\delta}_{gal}$ are the number density fluctuations of galaxies with and without the effect of RSD, $\sigma_v(M)$ is the velocity dispersion of galaxies inside a halo of mass M and $\mu = \hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{r}}$. We average over μ to obtain the angular averaged one-halo term

$$P_{gg}^{1h}(k) = \frac{1}{\bar{n}_{gal}^2} \int dM \frac{dn}{dM} \left\langle \binom{N}{2} \middle| M \right\rangle f_R(k|M), \tag{6}$$

where

$$\left\langle \binom{N}{2} \middle| M \right\rangle f_R(k|M) = \left[\langle N_{sat}|M \rangle \tilde{u}(k|M) R_1(M) \right.$$
$$\left. + \frac{1}{2} \langle N_{sat}(N_{sat} - 1)|M \rangle |\tilde{u}(k|M)|^2 R_2(M) \right].$$

The factor

$$R_p(M) = \frac{\sqrt{\pi}}{2} \frac{\mathrm{erf}[k\sigma_v(M)\sqrt{p/2}]}{k\sigma_v(M)\sqrt{p/2}} \tag{7}$$

comes from averaging over μ.

For the two-halo term, we multiply the large-scale and small-scale effects together (see Peacock 1999 and section 4 in Peacock & Dodds 1994)

$$\tilde{\delta}_{gal}^z(\boldsymbol{k}) = \left(\tilde{\delta}_{gal}(\boldsymbol{k}) + f(\Omega_M)\tilde{\delta}_m(\boldsymbol{k})\mu^2 \right) e^{-\frac{1}{2}[k\sigma_v(M)\mu]^2}. \tag{8}$$

The first part is the familiar Kaiser result with $f(\Omega_M) \equiv d \ln D/d \ln a$, where $D(a)$ is the linear growth function of density fluctuations and a is the scale factor. The density fluctuation of dark matter is denoted by $\tilde{\delta}_m$. The calculation thus includes not only the galaxy power spectrum, but also the matter power spectrum and the matter–galaxy cross power spectrum. After averaging over μ, we obtain

$$P_{gg}^{2h}(k) = \left(F_g^2 + \frac{2}{3} F_g F_v + \frac{1}{5} F_v^2 \right) P_{lin}(k), \tag{9}$$

where

$$F_g(k) = \frac{1}{\bar{n}_{gal}} \int dM \frac{dn}{dM} \langle N|M \rangle b(M) R_1(M) \tilde{u}(k|M) \tag{10}$$

comes from the contribution of $\tilde{\delta}_{gal}$ and

$$F_v(k) = f(\Omega_M) \frac{1}{\bar{\rho}} \int dM \frac{dn}{dM} M b(M) R_1(M) \tilde{u}_m(k|M) \tag{11}$$

comes from the contribution of $\tilde{\delta}_m$. Here $\tilde{u}_m(k|M)$ denotes the *dark matter* density profile normalized the same way as $\tilde{u}$ and $\bar{\rho}$ is the average matter density of the universe. We assume that $\tilde{u}_m(k|M)$ follows the Navarro–Frenk–White profile (Navarro, Frenk & White 1997) throughout the paper.

The left-hand panel of Fig. 1 shows an example of the contribution to the total galaxy power spectrum by the one-halo (blue) and two-halo (red) terms. The input of halo model will be detailed in Section 3.1. The solid and dashed curves correspond to including and excluding the effect of RSD. As can be seen, including RSD significantly reduces the power at small scale. We also note that the scale where one- and two-halo terms cross shifts very slightly due to RSD.

The right-hand panel of Fig. 1 presents the comparison between our model and one of the power spectra from the WiggleZ survey, provided by Parkinson et al. (2012). The green dashed/blue solid curve corresponds to the theoretical $P(k)$ before/after convolving with the window function of WiggleZ. We assume that the HOD is described by the five parameters in equation (27); we fit for these five parameters and show the model corresponding to the best-fitting parameters. This figure is only for the purposes of illustration; details of the fitting procedure will be presented in a future paper.

3 BASELINE MODEL AND FIDUCIAL DARK ENERGY CONSTRAINTS

In this section, we describe our inputs for the halo model, assumptions about the survey, predictions for the galaxy power spectrum, and Fisher matrix calculations of the statistical and systematic errors.

3.1 Baseline assumptions

We use the virial mass M_{vir} of dark matter haloes throughout this work and adopt the following functions in our halo model calculations.

(i) Mass function (dn/dM) and halo bias ($b(M, z)$): based on the fitting functions in Tinker et al. (2008, 2010), which are derived from N-body simulations and can achieve approximately 5 per cent accuracy for the mass function and 6 per cent for the halo bias.
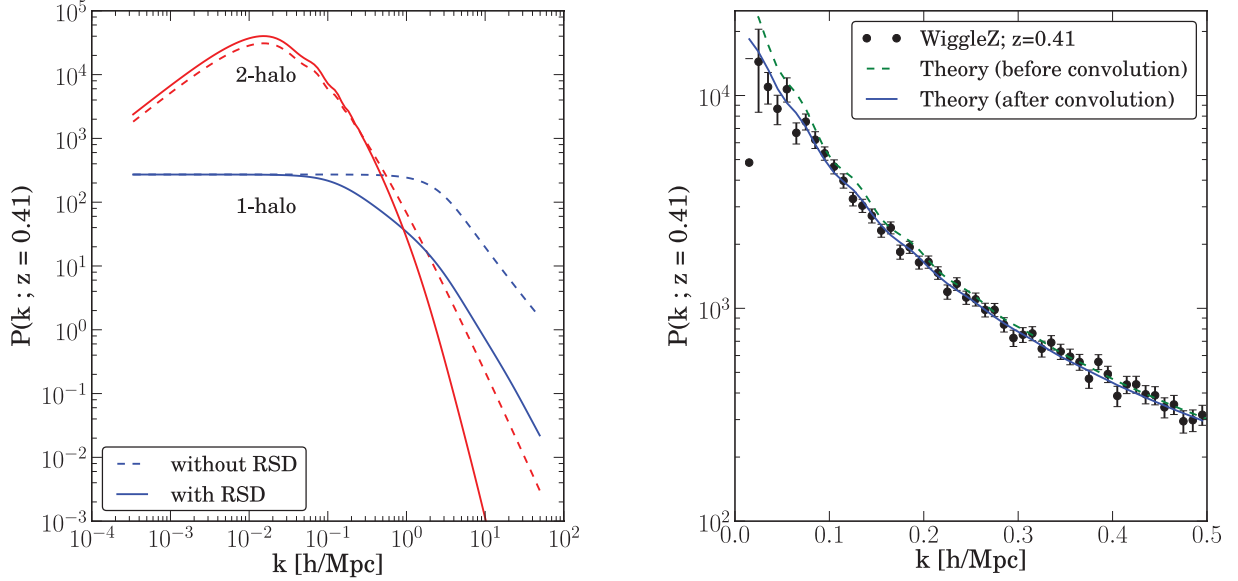
Figure 1. Galaxy power spectrum calculated based on the halo model. Left: the blue and red curves show the one- and two-halo terms, respectively. The solid curves include the RSD, while the dashed curves do not. RSD greatly reduce the power spectrum at small scale (the 'Fingers-of-God' effect), and only slightly shift the scale where one- and two-halo terms cross. Right: our model fit to WiggleZ data from Parkinson et al. (2012). The blue solid curve shows the theoretical $P(k)$ with the best-fitting HOD parameters and has been convolved with the observational window function.

(ii) Density profile: based on the universal NFW profile (Navarro et al. 1997), which is described by one concentration parameter $c_{\rm vir}$,

$$u_{\rm NFW}(r|M_{\rm vir}) \propto \frac{1}{(r/r_s)(1+r/r_s)^2} ,$$

$$c_{\rm vir}(M_{\rm vir}) = R_{\rm vir}/r_s . \tag{12}$$

(iii) Concentration–mass relation: based on the relation in Bhattacharya et al. (2013), which will be further discussed in Section 6.1. In the presence of significant scatter in the c–M relation, we perform the integration

$$u(r|M_{\rm vir}) = \int {\rm d}c_{\rm vir} P(c_{\rm vir}|M_{\rm vir})u(r|M_{\rm vir}(c_{\rm vir})) . \tag{13}$$

Throughout this paper, we assume that $c_{\rm vir}$ has a Gaussian distribution for a given $M_{\rm vir}$ with a scatter of 0.33, based on the finding of Bhattacharya et al. (2013).

(iv) Velocity dispersion: based on the scaling relation between dark matter velocity dispersion and halo mass from Evrard et al. (2008),

$$\sigma_{\rm v}^{\rm DM} = 1082.9 \left(\frac{h(z)M_{200}}{10^{15}\,{\rm M}_\odot} \right)^{0.3361} {\rm km\ s}^{-1} . \tag{14}$$

We convert the mass M_{200} to $M_{\rm vir}$ based on Hu & Kravtsov (2003). Since the scatter in the velocity dispersion is expected to be small (4 per cent), it is not included in our calculation.

(v) HOD: based on the parametrization from Zheng et al. (2005) and the fiducial parameters from Coupon et al. (2012), both of which will be discussed in detail in Section 4.

We assume a fiducial galaxy survey covering $f_{\rm sky} = 1/3$ of the full sky (about 14 000 square degrees), similar to the BigBOSS experiment.[3] We assume that the survey depth is comparable to the Canada–France–Hawaii Telescope Legacy Survey (CFHTLS) results presented in Coupon et al. (2012); specifically, we assume

five redshift bins in the range $0.2 < z < 1.2$, and the limiting magnitude in each bin is summarized in Table 1. We assume no uncertainties in the redshift measurements of galaxies. Given that the assumption of such a deep, wide spectroscopic survey *may* be somewhat optimistic, our required control of systematic errors may be somewhat more stringent than what BigBOSS needs.

We include seven cosmological parameters, whose fiducial values are based on the *Wilkinson Microwave Anisotropy Probe* 7 constraints (Komatsu et al. 2011): total matter density relative to critical $\Omega_{\rm M} = 0.275$; dark energy equation of state today and its variation with scale factor $w_0 = -1$ and $w_a = 0$, respectively; physical baryon and matter densities $\Omega_{\rm b}h^2 = 0.022\,55$ and $\Omega_{\rm M}h^2 = 0.1352$; spectral index $n_s = 0.968$; and the amplitude of primordial fluctuations $A = \Delta_\zeta^2(k = 0.002\,h^{-1}\,{\rm Mpc}) = 2.43 \times 10^{-9}$. We assume a flat universe; thus, dark energy density $\Omega_{\rm DE} = 1 - \Omega_{\rm M}$.

3.2 Likelihood function of $P(k)$ and error forecasting

Here we follow the derivations in Scoccimarro, Zaldarriaga & Hui (1999) and Cooray & Hu (2001) but use a different convention for the Fourier transform (see Appendix B). If we assume a thin shell in $\ln k$ space with width $\delta \ln k$ around $\ln k_i$, the power spectrum estimator reads

$$\hat{P}(k_i) = \int_{k_i} \frac{{\rm d}^3 \boldsymbol{k}}{V_s(k_i)} \delta(\boldsymbol{k})\delta(-\boldsymbol{k}) + \frac{1}{\bar{n}_{\rm gal}} , \tag{15}$$

where

$$V_s(k_i) = 4\pi k_i^3 \delta \ln k \tag{16}$$

and $1/\bar{n}_{\rm gal}$ accounts for the effect of shot noise. The first term of $\hat{P}$ is calculated based on the halo model results described in Section 2.2.

The covariance of power spectrum is given by

$$C_{ij} \equiv \left\langle \hat{P}(k_i)\hat{P}(k_j) \right\rangle - \left\langle \hat{P}(k_i) \right\rangle \left\langle \hat{P}(k_j) \right\rangle$$

$$= \frac{(2\pi)^3}{V_z} \frac{2P(k_i)^2}{V_s(k_i)} \delta_{ij} + \bar{T}_{ij} , \tag{17}$$

Table 1. Fiducial values for the HOD parameters, adopted from Coupon et al. (2012) based on CFHTLS.

Redshift	$M_g - 5\log_{10}h$	$\log_{10}M_{\min}$	$\log_{10}M_1$	$\log_{10}M_0$	$\sigma_{\log_{10}M}$	α_{sat}
$0.2 < z < 0.4$	-17.8	11.18	12.53	7.54	0.40	1.10
$0.4 < z < 0.6$	-18.8	11.48	12.66	10.96	0.43	1.09
$0.6 < z < 0.8$	-19.8	11.77	12.83	11.54	0.50	1.07
$0.8 < z < 1.0$	-20.8	12.14	13.21	12.23	0.35	1.12
$1.0 < z < 1.2$	-21.8	12.62	13.79	8.67	0.30	1.50

where the second term on the right-hand side is the contribution from the connected term given by the trispectrum describing the non-Gaussian nature of the random field,

$$\bar{T}_{ij} = \int_{k_i} \frac{\mathrm{d}^3 \boldsymbol{k}_1}{V_s(k_i)} \int_{k_j} \frac{\mathrm{d}^3 \boldsymbol{k}_2}{V_s(k_j)} T(\boldsymbol{k}_1, -\boldsymbol{k}_1, \boldsymbol{k}_2, -\boldsymbol{k}_2) . \tag{18}$$

We provide the detailed derivation in Appendix C. In equation (17), V_z is the volume of the redshift bin, $V_z = \Omega_{\mathrm{survey}} \int r^2(z)/H(z)\,\mathrm{d}z$, where the integral is performed over the redshift extent of the bin.

The calculation of $\bar{T}_{ij}$ involves four-point statistics, which is non-trivial to calculate. Fortunately, Cooray & Hu (2001) have shown that only the one-halo term dominates at the scale where the contribution of $\bar{T}_{ij}$ to C_{ij} is not negligible; therefore, we only need to calculate the one-halo contribution:

$$T^{1h}(k_1, k_2, k_3, k_4) = \frac{1}{V_z \bar{n}_{\mathrm{gal}}^4} \int \mathrm{d}M \frac{\mathrm{d}n}{\mathrm{d}M}$$

$$\times \left\langle \binom{N}{4} \Big| M \right\rangle f(k_1, k_2, k_3, k_4; M), \tag{19}$$

where

$$\left\langle \binom{N}{4} \Big| M \right\rangle f(k_1, k_2, k_3, k_4; M)$$

$$= \left\langle \binom{N_{\mathrm{sat}}}{3} \Big| M \right\rangle \frac{1}{4} \left(\Pi_{i=1}^3 \tilde{u}(k_i|M) + \mathrm{cyc.} \right)$$

$$+ \left\langle \binom{N_{\mathrm{sat}}}{4} \Big| M \right\rangle \Pi_{i=1}^4 \tilde{u}(k_i|M) . \tag{20}$$

Analogous to the case of P_{gg}^{1h} considered in Section 2.1, the first term accounts for quadruplets composed of one central and three satellite galaxies, and the second term accounts for the quadruplets composed of four satellite galaxies. We assume that $P(N_{\mathrm{sat}}|M)$ follows the Poisson distribution so that $\langle \binom{N_{\mathrm{sat}}}{3} | M \rangle = \langle N_{\mathrm{sat}} | M \rangle^3 / 3!$ and $\langle \binom{N_{\mathrm{sat}}}{4} | M \rangle = \langle N_{\mathrm{sat}} | M \rangle^4 / 4!$.

We employ the Fisher matrix formalism to forecast the statistical errors of the cosmological and nuisance parameters based on the fiducial survey. The Fisher matrix reads

$$F_{\alpha\beta} = \sum_z \sum_{i,j} \frac{\partial P_i}{\partial \theta_\alpha} \left[\frac{(2\pi)^3}{V_z} \frac{P_i^2}{2\pi k_i^3 \delta \ln k} \delta_{ij} + T_{ij} \right]^{-1} \frac{\partial P_j}{\partial \theta_\beta}$$

$$= \sum_z \sum_{i,j} \frac{\partial \ln P_i}{\partial \theta_\alpha} \left[\frac{(2\pi)^3}{V_z} \frac{1}{2\pi k_i^3 \delta \ln k} \delta_{ij} + \frac{T_{ij}}{P_i P_j} \right]^{-1} \frac{\partial \ln P_j}{\partial \theta_\beta} ,$$

where α and β are indices of model parameters, while i and j refer to bins in wavenumber which have a constant logarithmic width $\delta \ln k$ and extend out to the maximum wavenumber $k_{\max}$. We adopt $\delta \ln k = 0.1$, which has been tested to be small enough to ensure

convergence. The best achievable error in the parameter θ_α is given by

$$\sigma_{\theta_\alpha} = [(\mathbf{F}^{-1})_{\alpha\alpha}]^{1/2} . \tag{21}$$

Throughout this work, unless otherwise indicated, the full set of parameters considered is given by

$$\theta_{\mathrm{full}} = (w_0, w_a, \Omega_{\mathrm{DE}}, \Omega_M h^2, \Omega_b h^2, n_s, \ln A;$$

$$\log_{10} M_{\min}, \sigma_{\log_{10} M}, \log_{10} M_0, \log_{10} M_1, \alpha_{\mathrm{sat}}) . \tag{22}$$

The first seven are the cosmological parameters introduced in Section 3.1, while the last five are the nuisance parameters describing the HOD and will be discussed in Section 4.1.

3.3 Fiducial constraints without systematics

To represent the statistical power of an upcoming galaxy redshift survey, in the limiting case of no nuisance parameters, we consider the inverse of the square root of the dark energy figure of merit, originally defined as the inverse of the forecasted 95 per cent area of the ellipse in the w_0–w_a plane (Huterer & Turner 2001; Albrecht et al. 2006). In other words, our parameter of interest is $\sqrt{\sigma(w_a)\sigma(w_p)}$, where w_p is the pivot that physically corresponds to $w(a)$ evaluated at the scale factor where the constraint is the best. This quantity takes into account the temporal variation of dark energy, and the square root serves to compare it fairly to the constant w; the two quantities, $\sigma(w)$ and $\sqrt{\sigma(w_0)\sigma(w_p)}$, tend to show very similar behaviour. For our fiducial survey, the statistical error in our parameter combination of interest is $\sqrt{\sigma(w_a)\sigma(w_p)} = 0.4$ (or 0.003) for $k_{\max} = 0.1$ (or 1) $h\,\mathrm{Mpc}^{-1}$, without external priors. When we add the Planck Fisher matrix (Hu, private communication), $\sqrt{\sigma(w_a)\sigma(w_p)}$ becomes 0.002 (or 0.0002) for $k_{\max} = 0.1$ (or 1) $h\,\mathrm{Mpc}^{-1}$.

Fig. 2 presents the expected dark energy constraints as a function of $k_{\max}$, without nuisance parameters or systematic errors for the moment, for three levels of sophistication in the theory. We proceed in steps: the blue curve corresponds to no RSD (Section 2.1) with a Gaussian likelihood function. In this case, the dark energy constraints increase sharply with $k_{\max}$, indicating that these assumptions are unrealistic. The green curve includes the RSD (Section 2.2), which reduce the dark energy information from small scales. The red curve further includes the effect of non-Gaussian likelihood [$\bar{T}_{ij}$ from equation (18)], which reduces the information at high k even more.

3.4 Systematic bias in model parameters

In this work, we estimate the systematic shifts in parameter inference caused by using an inadequate model. In particular, if we assume a problematic model that produces a power spectrum $P_{\mathrm{sys}}(k)$ that systematically deviates from the truth $P_{\mathrm{fid}}(k)$, we will obtain parameters that systematically deviate from their true values:
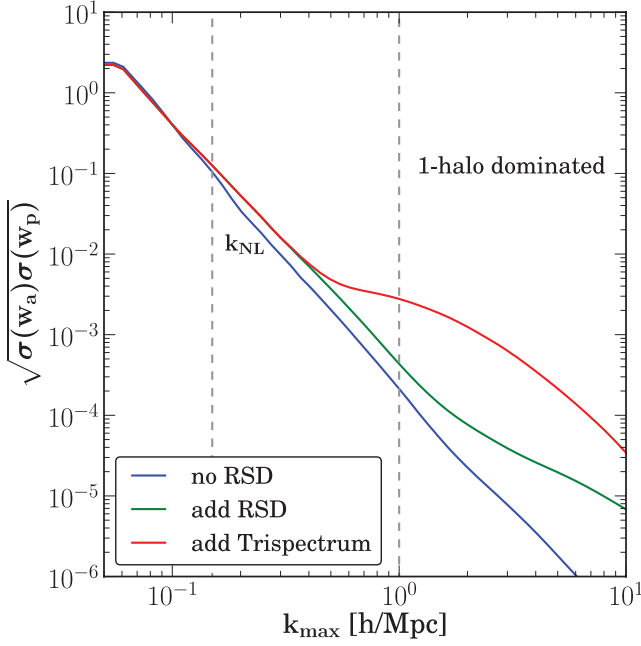
Figure 2. Dark energy information content from $P(k)$, based on our fiducial survey assumptions. The x-axis corresponds to the largest k (smallest scale) assumed to be reliably measured and interpreted. The blue curve corresponds to a Gaussian likelihood function and assumes no RSD; it leads to unrealistically tight constraints for large k_{max}. The green curve includes RSD, and the red curve further includes the trispectrum correction to the covariance matrix. As can be seen, including these two effects reduces the small-scale information.

$\theta_{sys} = \theta_{fid} + \Delta\theta$. The systematic shifts in parameters can be obtained through a modified Fisher matrix formalism (Knox, Scoccimarro & Dodelson 1998):

$$\Delta\theta_\alpha = \sum_\beta (F^{-1})_{\alpha\beta} G_\beta , \qquad (23)$$

where

$$G_\beta \equiv \sum_z \sum_{i,j} (\ln P_{sys,i} - \ln P_{fid,i})$$
$$\times \left[\frac{(2\pi)^3}{V_z} \frac{1}{2\pi k_i^3 \delta \ln k} \delta_{ij} + \frac{T_{ij}}{P_i P_j} \right]^{-1} \frac{\partial \ln P_j}{\partial \theta_\beta} . \qquad (24)$$

To determine the significance of systematic errors, we calculate the systematic shifts $\Delta\chi_{tot}^2$ in the full high-dimensional parameter space,

$$\Delta\chi_{tot}^2 = \Delta\theta^T \mathbf{F} \Delta\theta , \qquad (25)$$

where $\Delta\theta$ is the vector of the systematic shifts of parameters. Both $\Delta\theta$ and the Fisher matrix $\mathbf{F}$ include cosmological and nuisance parameters. The systematic bias is considered significant if the inferred θ_{sys} lies outside the 68.3 per cent confidence interval of the Gaussian likelihood function centred on θ_{fid}; in other words, the bias is 'greater than the 1σ dispersion'. For example, in a full 12-dimensional parameter space considered here, the 68.3 per cent confidence interval corresponds to $\Delta\chi_{tot}^2 = 13.7$.

4 SELF-CALIBRATION OF HOD PARAMETERS

In this section, we focus on the efficacy of self-calibrating the HOD parameters, that is, determining these parameters from the survey

concurrently with cosmological parameters. Since these HOD parameters are not known a priori, one usually marginalizes over them along with cosmological parameters (e.g. Tinker et al. 2012), which inevitably increases the uncertainties in cosmological parameters. Here we focus on the statistical uncertainties and assume no systematic error; in the next section, we will compare these statistical errors with systematic shifts of parameters.

We focus on two parametrization of HOD: one is based on Zheng et al. (2005) and the other is based on a piecewise continuous parametrization.

4.1 Zheng et al. parametrization

The HOD describes the probability distribution of having N galaxies in a halo of mass M. In principle, the HOD is specified by the full distribution $P(N|M)$; in practice, modelling of the two-point statistics only requires $\langle N_{cen}|M \rangle$, $\langle N_{sat}|M \rangle$ and $\langle N_{sat}(N_{sat} - 1)|M \rangle$. We follow the HOD parametrization from Zheng et al. (2005), which separates the contribution from central and satellite galaxies:

$$\langle N_{cen}|M \rangle = \frac{1}{2} \left[1 + \mathrm{erf} \left(\frac{\log_{10} M - \log_{10} M_{min}}{\sigma_{\log_{10} M}} \right) \right] \qquad (26)$$

$$\langle N_{sat}|M \rangle = \langle N_{cen}|M \rangle \times \left(\frac{M - M_0}{M_1} \right)^{\alpha_{sat}} . \qquad (27)$$

The first equation describes the contribution from the central galaxy; M_{min} corresponds to the threshold mass where a halo can start to host a galaxy that is observable to the survey and $\sigma_{\log_{10} M}$ describes the transition width of this threshold. The second equation describes the contribution from satellite galaxies, whose number is assumed to follow a power law, and M_0 is the cutoff mass. In addition, we make the widely adopted assumption that $P(N_{sat}|M)$ follows a Poisson distribution, i.e.

$$\langle N_{sat}(N_{sat} - 1)|M \rangle = \langle N_{sat}|M \rangle^2 . \qquad (28)$$

We adopt the fiducial values from Coupon et al. (2012), which are constrained using the projected angular two-point correlation function $w(\theta)$ from the CFHTLS out to $z = 1.2$. We use the same binning and limiting magnitude as in Coupon et al. (2012); the values are summarized in Table 1. We do not use the error bars quoted there as our priors because we would like all parameters to be self-calibrated consistently.

Under these assumptions, we have five nuisance parameters ($\log_{10} M_{min}$, $\sigma_{\log_{10} M}$, $\log_{10} M_0$, $\log_{10} M_1$, α_{sat}) for each of the five redshift bins, i.e. 25 parameters in total. We assume that each of the five distinct nuisance parameters varies coherently across the five redshift bins, and is therefore described by a single parameter. Under this assumption, instead of 25 nuisance parameters, we only use 5 nuisance parameters to describe the uncertainties of all HOD parameters. We parametrize the variations around the fiducial values:

$$\theta_i^{HOD}(z) = h_i \theta_i^{HOD,fid}(z) \quad (i = 1, \ldots, 5), \qquad (29)$$

where h_i are the dimensionless parameters describing the uncertainties of the aforementioned five HOD parameters. We note that this choice of five HOD parameters only represents one possible model; depending on the data available and the astrophysical motivation, in principle one can use a more general model to describe the evolution of HOD. Increasing the number of degrees of freedom describing the evolution of HOD will inevitably lead to degradation in the dark energy constraints, and it will be very important to establish the
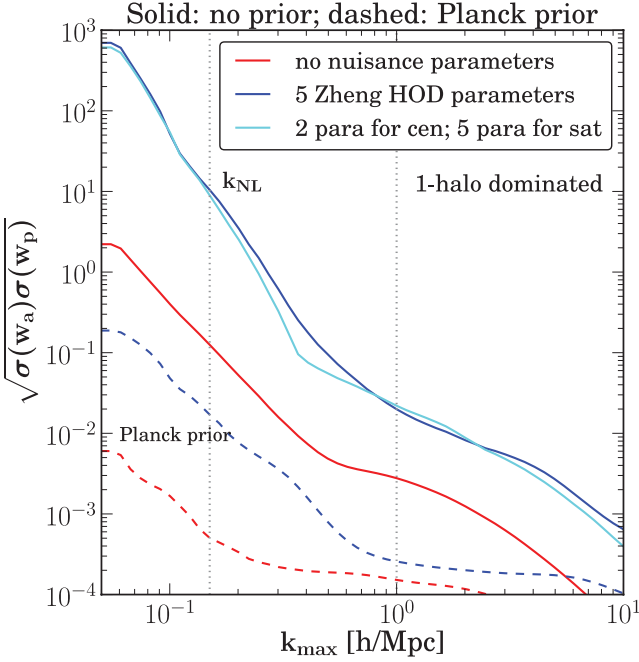
Figure 3. Self-calibration of HOD parameters. We show the dark energy constraints as a function of the highest k used in the survey. The red curve corresponds to no nuisance parameters. The dark blue curve corresponds to five nuisance parameters based on the parametrization in Zheng et al. (2005), while the cyan curve corresponds to a piecewise continuous parametrization for satellite galaxies, with one parameter in each of the five mass bins. Including nuisance parameters in either parametrization systematically increases the dark energy uncertainties by one or two orders of magnitude. The dashed curves include the Planck prior and assume the same nuisance parameters as their solid-curve counterparts.

total number of degrees of freedom necessary to model the HOD and its uncertainties.

We explore how well these parameters can be self-calibrated by $P(k)$ without the aid of priors. Fig. 3 shows the dark energy constraints as a function of k_{max}, with fixed nuisance parameters (red) and with these five marginalized nuisance parameters (dark blue). The RSD and the full covariances of $P(k)$ are included in this calculation. Clearly, the dark energy constraints are weakened by approximately about one or two orders of magnitude when we marginalize over HOD parameters.

4.2 Piecewise continuous parametrization of HOD parameters

One potential worry with the parametrization in equation (27) is whether $\langle N_{sat}|M \rangle$ is accurately described by a power law. To address this, we propose a less model-dependent, piecewise continuous parametrization for $\langle N_{sat}|M \rangle$. We divide the halo mass range into n_{bins} bins and assign a parameter describing the uncertainties of HOD in each bin. That is,

$$\langle N_{sat}|M \rangle = \sum_{i=1}^{n_{bins}} \Theta_i(M) f_i \langle N_{sat}|M \rangle_{fid} , \quad (30)$$

where $\Theta_i(M)$ defines the binning and equals 1 in $[M_i, M_{i+1}]$ and 0 elsewhere, while f_i is the free parameter in bin i and describes the uncertainty of $\langle N_{sat}|M \rangle$ in this bin.

We still assume $P(N_{sat}|M)$ to be a Poisson distribution, which now implies

$$\langle N_{sat}(N_{sat}-1)|M \rangle = \langle N_{sat}|M \rangle^2$$

$$= \sum_{i=1}^{n_{bins}} \Theta_i(M) f_i^2 \langle N_{sat}|M \rangle_{fid}^2 . \quad (31)$$

We start with one parameter per decade in mass, using $n_{bins} = 5$ parameters between 10^{11} and $10^{16}\, h^{-1}\, M_\odot$, equally spaced in $\log_{10} M$. We assume these parameters to be independent of redshift. The cyan curve in Fig. 3 corresponds to marginalizing over these five piecewise continuous parameters for $\langle N_{sat}|M \rangle$ and two parameters $(\log_{10} M_{min}, \sigma_{\log_{10} M})$ for $\langle N_{cen}|M \rangle$, with no prior on them.

Fig. 4 shows the dependence of dark energy constraints on the number of parameters describing $\langle N_{sat}|M \rangle$ per decade of mass, $N_{per\,decade}$. The three panels correspond to $k_{max} = 0.1$, 0.4 and $1\, h\,Mpc^{-1}$. The Planck prior is included in this calculation. The black curve corresponds to no prior on f_i and shows strong degradation with increasing $N_{per\,decade}$ as one would expect. When k_{max} is small, the prior knowledge of HOD is important to improve the dark energy constraints. On the other hand, when k_{max} is large, HOD can be well self-calibrated, and the prior is not as important.

To enable a fair comparison of priors, however, we would like to increase the freedom in the HOD model while fixing the overall uncertainty per decade. To do this, we impose a fixed prior *per decade* of mass:

$$\sigma_{f_i} = \sigma_0 \sqrt{N_{per\,decade}} \quad (32)$$

so that the total prior per unit $\log_{10} M$, when we add the Fisher information from all f_i, is σ_0 regardless of the value of $N_{per\,decade}$.

The red/green/blue curves in Fig. 4 correspond to imposing $\sigma_0 = 1/0.1/0.01$. For $k_{max} = 0.1\, h\,Mpc^{-1}$, the dark energy constraints converge when we use one parameter per decade of mass regardless of the prior on nuisance parameters. When $k_{max} > 0.1\, h\,Mpc^{-1}$, a few more parameters per decade in mass are required for the results to converge. For example, for $k_{max} = 0.4$ $(1.0)\, h\,Mpc^{-1}$, we need two (three) parameters per decade to ensure convergence. The required number of parameters also somewhat depends on the prior.

We note that the HOD parameters are progressively better self-calibrated when we go to higher k_{max}; when $k_{max} = 1\, h\,Mpc^{-1}$, self-calibrating the five HOD parameters only moderately degrades the dark energy constraints. This finding encourages future surveys to further push towards high k_{max} for rich cosmological and astrophysical information.

5 SYSTEMATIC ERRORS DUE TO THE UNCERTAINTIES IN THE HALO MASS FUNCTION

In this section, we focus on the effect of the uncertainties in the halo mass function on the cosmological constraints from galaxy clustering. The mass function has been widely explored analytically (e.g. Press & Schechter 1974) as well as numerically using dark matter *N*-body simulations (e.g. Sheth & Tormen 1999; Jenkins et al. 2001; Sheth, Mo & Tormen 2001; Evrard et al. 2002; Reed et al. 2003, 2013; Warren et al. 2006; Lukić et al. 2007, 2009; Cohn & White 2008; Tinker et al. 2008; Crocce et al. 2010; Watson et al. 2013; Bhattacharya et al. 2013) and hydrodynamical simulations (e.g. Rudd, Zentner & Kravtsov 2008; Stanek et al. 2010; Cui et al. 2012). The different fitting formulae for the mass function are often based on different halo identification methods and mass definitions;
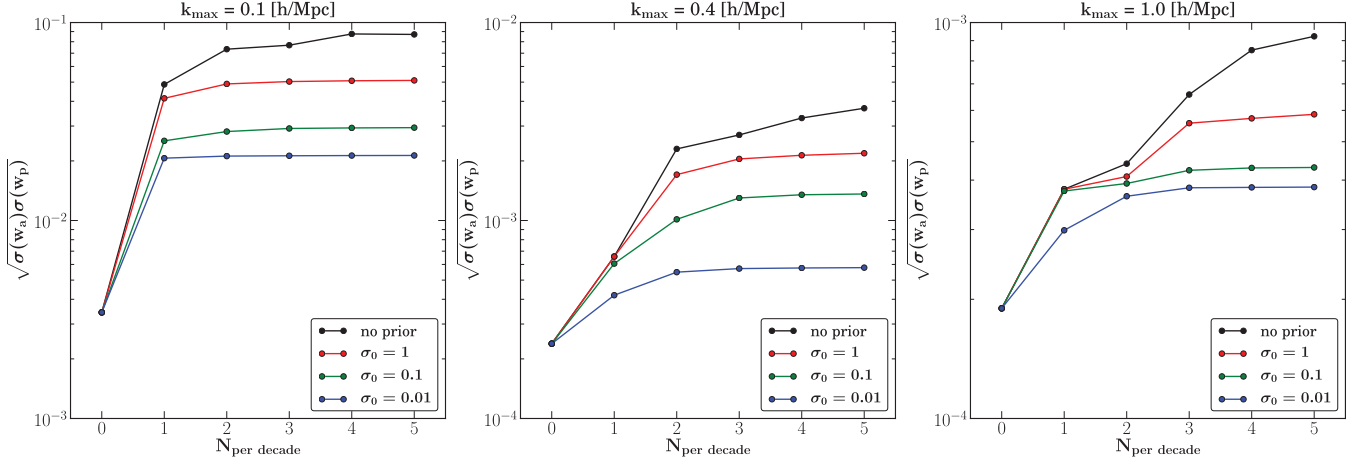
Figure 4. Dark energy constraints with self-calibrated piecewise continuous HODs. The three panels correspond to $k_{max} = 0.1$, 0.4 and $1\,h\,\mathrm{Mpc}^{-1}$. The *x*-axis corresponds to the number of parameters used to describe $\langle N_{sat}|M\rangle$ per decade of mass, and the *y*-axis corresponds to the dark energy constraints. The black curve corresponds to no prior, and the constraints are degraded with larger number of parameters. The other curves correspond to consistently adding *a fixed total* prior per decade of mass; that is, $\sigma_{f_i} = \sigma_0\sqrt{N_{per\ decade}}$, where $\sigma_0 = 1$, 0.1 or 0.01. We note that one parameter per decade is sufficient for $k_{max} = 0.1\,h\,\mathrm{Mpc}^{-1}$, while two or three parameters are needed for higher k_{max}. Note that for higher k_{max}, the HOD parameters are better self-calibrated, and the dark energy constraints are less dependent on the prior on HOD nuisance parameters.

therefore, instead of drawing a direct comparison between different fitting formulae, we choose one specific fiducial model and explore the uncertainties relative to this model.[4]

We use the fitting function from Tinker et al. (2008, described in Section 3.1), which has been calibrated based on a large suite of simulations implementing several different *N*-body algorithms and different versions of ΛCDM cosmology; therefore, it is likely to fairly represent the uncertainties in the mass function calibration. Tinker et al. (2008) quoted a statistical uncertainty of $\lesssim 5$ per cent at $z = 0$ (~ 1 per cent around M_*). However, the uncertainties are presumably larger at higher redshift and can further increase if the effects of baryons are taken into account.

We explore the effect of a small constant shift of the halo mass function, parametrized as

$$\left(\frac{\mathrm{d}n}{\mathrm{d}M}\right) = (1+\epsilon)\left(\frac{\mathrm{d}n}{\mathrm{d}M}\right)_{fid}. \qquad (33)$$

The main panel of Fig. 5 shows the impact of $\epsilon = 0.05$ on $P(k)$. For the two-halo term (small k), $P(k)$ changes by less than 1 per cent, because a constant shift in the mass function only affects F_v describing the large-scale RSD (see equation 9). For the one-halo term (large k), $P(k)$ changes by -5 per cent, which can be easily seen from equation (6); the numerator includes one integration of $\mathrm{d}n/\mathrm{d}M$ (galaxy pairs in one halo) while the denominator includes the square of such an integration.

We next see how this systematic shift in $P(k)$ impacts cosmological parameters. We use $\Delta\chi^2_{tot} = 13.7$ [1σ errors in a 12-dimensional parameter space; see equation (25)] as our criterion of significant impact from the systematic error. We calculate $\Delta\chi^2_{tot}$ using the Fisher matrix for seven cosmological parameters (Section 3.1) and five HOD parameters (Section 4.1). Throughout this and the next section, we use the Planck prior but *no* priors on HOD parameters. We believe that these two assumptions reflect reality in the next

[4] It has been shown that for surveys of cluster abundance such as the Dark Energy Survey, ~ 1 per cent accuracy in the mass function is required to avoid significant degradation in dark energy constraints (see Cunha & Evrard 2010; Wu, Zentner & Wechsler 2010). Here we would like to explore whether the same accuracy is sufficient for surveys of galaxy clustering.
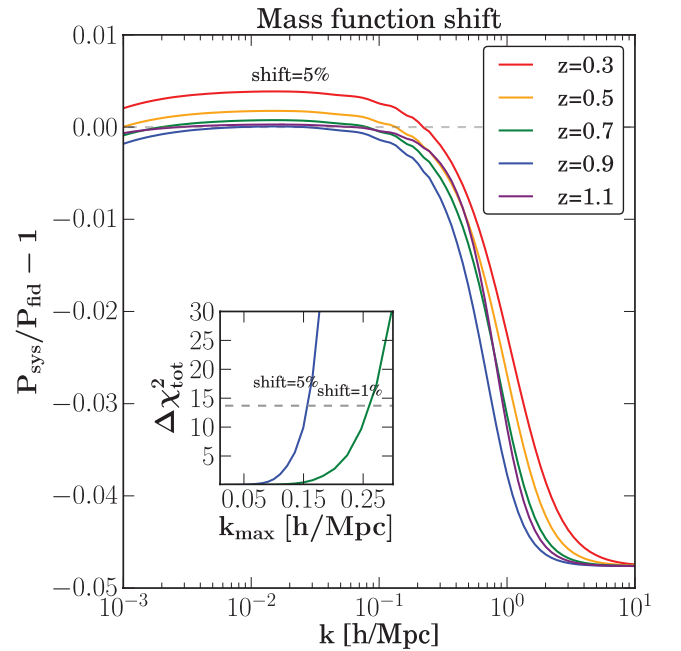

Figure 5. Impact of the uncertainty in the halo mass function on $P(k)$. The main panel shows the systematic shifts of $P(k)$ when the mass function is shifted by a constant 5 per cent (independent of mass and redshift). The inset shows $\Delta\chi^2_{tot}$ as a function of k_{max} when the mass function is shifted by 1 or 5 per cent. The horizontal dashed line marks $\Delta\chi^2_{tot} = 13.7$, the 1σ deviation in the 12-dimensional parameter space. As can be seen, 5 per cent (1 per cent) allows k_{max} up to 0.15 (0.25) $h\,\mathrm{Mpc}^{-1}$.

5–10 years, when Planck data will firmly pin down certain combinations of cosmological parameters, while the determination of the nuisance HOD quantities will keep improving. We note that unbiased priors always *decrease* the resulting systematic bias (for a proof, see appendix of Bernstein & Huterer 2010) and make the theoretical requirements less stringent. Thus, any prior on HOD parameters will alleviate the systematic biases and make the required accuracy of theory less stringent.

The inset of Fig. 5 shows how $\Delta \chi^2$ depends on k_{max}, for 5 per cent (blue) and 1 per cent (green) systematic shifts in the mass function. As can be seen, a 5 per cent (1 per cent) shift in the mass function can cause a significant systematic error at $k_{max} = 0.15$ (0.25) $h\,\mathrm{Mpc}^{-1}$. We note that at these scales, $P(k)$ is still dominated by the two-halo term; therefore, the systematic shifts caused by the mass function are mainly related to the large-scale redshift distortion (the Kaiser effect). However, this large-scale effect can be mitigated by prior knowledge of σ_8 (e.g. σ_8 constraints from galaxy cluster counts; Rozo et al. 2010), with which the large-scale galaxy bias can be calibrated. Therefore, by calibrating the large-scale clustering amplitude, one can in principle reduce the impact of the uncertainties in the mass function.

Finally, we note that the halo bias $b(M)$ is also currently being actively studied (e.g. Tinker et al. 2010; Ma et al. 2011; Manera & Gaztañaga 2011; Paranjape, Sheth & Desjacques 2013). The uncertainty in the halo bias is related to the uncertainty in the mass function; for example, Tinker et al. (2010) have indicated that their fitting function for the halo bias has an ~6 per cent uncertainty, which is related to the uncertainty of their mass function. In addition, the uncertainties and systematics in $b(M)$ will lead to a constant shift in the two-halo term (see equation 9). In this case, holding the galaxy bias fixed will cause a huge systematic shift in cosmological parameters (for example, σ_8); therefore, it is necessary to fit the overall galaxy bias to the large-scale clustering data. In this work, we do not specifically explore the impact of uncertainties of the halo bias because the halo bias determines the large-scale clustering amplitude, which can be observationally calibrated when combined with independent knowledge of σ_8. On the other hand, we note that a scale-dependent bias can arise from the primordial non-Gaussianity (e.g. Dalal et al. 2008) or small-scale non-linearity (e.g. Smith, Scoccimarro & Sheth 2007). In this case, one could resort to multiple tracers of large-scale structure (e.g. Seljak 2009; Cacciato et al. 2013), knowledge of primordial non-Gaussianity from the cosmic microwave background (e.g. Planck Collaboration 2013) or higher order statistics (e.g. Marín et al. 2013) to better calibrate the scale dependence of the galaxy bias.

6 SYSTEMATIC ERRORS DUE TO THE UNCERTAINTIES IN HALO PROPERTIES

In this section, we explore the impact of four sources of theoretical uncertainties related to the properties of dark matter haloes coming from N-body simulations on the constraining power of $P(k)$. These sources of systematics are as follows:

(i) concentration–mass relation,
(ii) deviation of $\tilde{u}$ from the NFW profile,
(iii) deviation of N_{sat} from the Poisson distribution and
(iv) velocity bias.

In particular, we address the following points.

(i) With the current level of uncertainties, what are the systematic errors in the prediction of $P(k)$? What are the biases in the parameter inference caused by these systematics?
(ii) What is the smallest scale (largest k_{max}) allowed by the current level of uncertainties?
(iii) What is the required reduction of these uncertainties if we would like to push to higher k_{max}?

We again use $\Delta \chi^2_{tot}$ to assess the impact of systematic errors on the cosmological parameters, as described in the previous section. The

summary of the impact of these systematics is presented in Fig. 6 and Table 2.

6.1 Concentration–mass relation

In the halo model, the one-halo term depends on the number density profile of galaxies, $\tilde{u}(k|M)$. We assume that the galaxy distribution follows the dark matter distribution, which is well described by an NFW profile. We then use the concentration–mass relation of dark matter haloes from the literature to compute $\tilde{u}(k|M)$.

The concentration–mass relation has been calibrated with dark matter N-body simulations (e.g. Bullock et al. 2001; Neto et al. 2007; Duffy et al. 2008; Macciò, Dutton & van den Bosch 2008; Prada et al. 2012; Bhattacharya et al. 2013; Kwan et al. 2013) and hydro-dynamical simulations (e.g. Lau, Kravtsov & Nagai 2009; Duffy et al. 2010; Rasia et al. 2013). Several observational programmes are also working towards pinning down this relation (e.g. Coe et al. 2012; Oguri et al. 2012). However, 10–20 per cent of uncertainties in the concentration–mass relation remain, and the concentration–mass relation also varies with cosmology and the implementation of baryonic physics (see, e.g., the review in Bhattacharya et al. 2013).

We investigate the impact of uncertainties in the concentration–mass relation by comparing the models from Bullock et al. (2001, B01 hereafter) and the recent calibration from Bhattacharya et al. (2013, B13 hereafter). These two models represent two extreme cases of the concentration–mass relation; therefore, using these two extreme cases sets the upper limit of the systematic bias caused by the c–M relation. We assume a scatter of 0.33 for the c–M relation in both cases. Ignoring this scatter will lead to an approximately 0.5 per cent difference in $P(k)$ at $k \approx 1\,h\,\mathrm{Mpc}^{-1}$.

Our baseline model is from the recent formula given by B13 (based on virial overdensity):

$$c(\nu) = D(z)^{0.78} 7.9 \nu^{-0.28}$$

$$\nu = \frac{1}{D(z)} \left[1.12 \left(\frac{M_{vir}}{5 \times 10^{13}\,h^{-1}\,\mathrm{M}_\odot} \right)^{0.3} + 0.53 \right]. \tag{34}$$

We compare it with the model from B01:

$$c(M_{vir}) = \frac{9}{1+z} \left(\frac{M_{vir}}{M_*(z)} \right)^{-0.13}. \tag{35}$$

These two calibrations agree near M_* at $z = 0$.

The top-left panel of Fig. 6 shows the relative change in the power spectrum $P(k)$, evaluated at five redshifts, due to the difference between B01 and B13. We find that $P(k)$ based on B01 is in general lower than that based on B13, because B01 predict lower concentrations at the high-mass end. Although B01 predict higher concentrations at the low-mass end, these haloes rarely contribute to the one-halo term and thus do not significantly boost clustering.

The inset in this panel shows the systematic shifts in the parameter space caused by different models, which are characterized by $\Delta \chi^2_{tot}$. It can be seen that the systematic error starts to be comparable to the statistical error ($\Delta \chi^2_{tot} = 13.7$, marked by a horizontal dashed line) at $k_{max} = 1.2\,h\,\mathrm{Mpc}^{-1}$, which makes it a relatively unimportant source of systematic error.

We would now like to study the effects of improved calibration in the c–M relation. A natural way to do this is to assume that the difference between the two extreme predictions has been reduced by some constant factor, and that the new value interpolates between the two original extremes. We define the interpolated value as

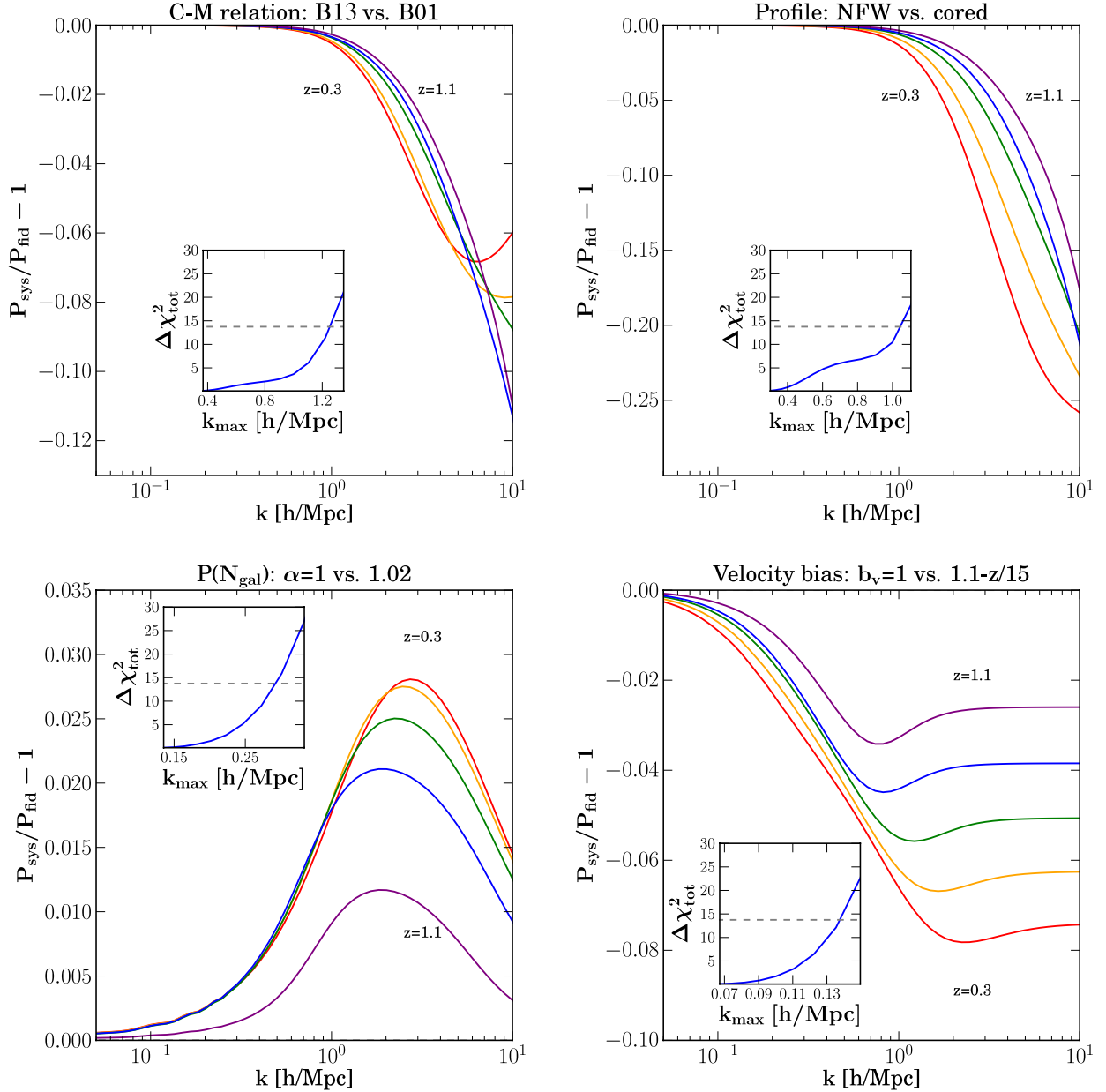$$c_{interp}(M) = c_{fid}(M) + f_{sys} (c_{alt}(M) - c_{fid}(M)), \tag{36}$$

Figure 6. Systematic differences in $P(k)$ caused by the four sources of errors discussed in Section 6. In each panel, the main figure shows the fractional difference in $P(k)$ in the five redshift bins, while the inset shows the systematic error $\Delta\chi^2_{tot}$ as a function of k_{max}. The 1σ deviation in the 12-dimensional parameter space, $\Delta\chi^2_{tot} = 13.7$, is marked by the horizontal dashed line in each inset.

Table 2. Summary of the effects of the four sources of systematic error considered in Section 6. Note that $\langle\Delta P/P\rangle$ is calculated at $k = k_{max}$ and averaged over the five redshift bins, and 'f_{sys} req.' is the required reduction factor in the amplitude of the systematic difference so that it becomes a 1σ effect in the full parameter space.

	Systematic difference	k_{max} allowed	$k_{max} = 0.3\,h\,{\rm Mpc}^{-1}$			$k_{max} = 1\,h\,{\rm Mpc}^{-1}$		
			$\langle\Delta P/P\rangle$	Deviation (σ)	f_{sys} req.	$\langle\Delta P/P\rangle$	Deviation (σ)	f_{sys} req.
c–M relation	B13 versus B01	1.3	0.000 19	0	None	0.0039	0.014	None
Profile	NFW versus cored	1	0.0004	0	None	0.0074	0.56	None
$P(N_{sat})$	$\alpha = 1$ versus 1.02	0.29	0.0037	1.3	0.93	0.016	15	0.23
Velocity bias	$b_v = 1$ versus $1.1-z/15$	0.14	0.026	32	0.11	0.052	108	0.034

where $c_{fid}(M)$ and $c_{alt}(M)$ are, respectively, the fiducial (say, B13) and the alternate (say, B01) models for the concentration–mass relation. Here f_{sys} is a tunable parameter that allows us to assess the effect of a fraction of the full systematics. The limiting cases are

$$f_{sys} = 0 \iff \text{no systematics}$$

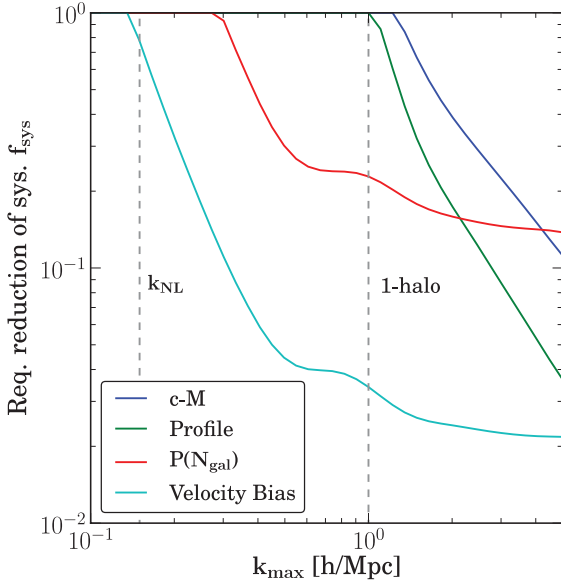$$f_{sys} = 1 \iff \text{fiducial systematics.}$$

Figure 7. Required reduction of systematic errors, shown as the fraction of the current errors, for the four sources of systematic errors discussed in this paper, as a function of the maximum wavenumber considered in the survey. Note that the velocity bias requires the greatest improvement relative to the current knowledge.

For a higher k_{max}, the tolerance of systematics is smaller, and f_{sys} provides a measure for required reduction of systematics. For a given k_{max}, we search for the appropriate f_{sys} value that makes the systematic negligible.[5]

The blue curve in Fig. 7 shows the requirement on f_{sys} from the c–M relation as a function of k_{max}. For all practical k_{max} values, c–M does not require more precise calibrations from N-body simulations. The results are summarized in the 'c–M relation' row of Table 2.

6.2 Galaxy number density profile: deviation from NFW

Our fiducial model assumes that the galaxy distribution inside a halo is described by the NFW profile. However, N-body simulations have shown that the distribution of subhaloes in cluster-size haloes tends to be shallower than the NFW profile and also shallower than the observed galaxy number density profile (e.g. Diemand, Moore & Stadel 2004; Nagai & Kravtsov 2005). These deviations could be related to insufficient resolution or the absence of baryons in N-body simulations – the so-called overmerging issue. Several authors have proposed models for 'orphan galaxies' to compensate the overmerging issue; however, these models do not always recover the observed galaxy clustering (e.g. Guo et al. 2011). The exact cause for these issues is still uncertain; nevertheless, the uncertainties associated with the distribution of subhaloes will likely impact the modelling of galaxy clustering. Based on the comparisons between dark matter and hydrodynamical simulations (e.g. Macciò et al. 2006; Weinberg et al. 2008), the observed galaxy density profile is likely to be bracketed by the density profiles of the subhalo number and dark matter.

---

[5] Note that some fraction f_{sys} of the systematics does *not* trivially lead to the same fractional shift in $P(k)$ because the c–M relation (and most other systematics) enters non-linearly into $P(k)$. We therefore need to perform a separate calculation of $P(k)$ for each f_{sys}.

In this section, we investigate whether the uncertainties in the galaxy number density profile lead to a significant systematic bias. To model the possibility that the galaxy distribution is shallower than dark matter in the inner region of clusters, we adopt the subhalo number density profile measured from Wu et al. (in preparation), which is also illustrated in Appendix A. Fig. A1 presents one example of the galaxy number density profile measured from an N-body simulation. Based on this result, we model the subhalo number density profile as

$$u(r|M) = f_{surv}(M, r) \times u_{NFW}(r|M), \qquad (37)$$

where $u_{NFW}(r|M)$ is the NFW profile and f_{surv} is the 'surviving fraction' of galaxies given by

$$f_{surv} = 1 - 0.99 e^{-a(r/R_{vir})}$$

$$a \equiv 0.005 \left(\ln \frac{M_{vir}}{1000\, h^{-1}\, M_{\odot}} \right)^2. \qquad (38)$$

We note that f_{surv} is smaller for higher host halo mass and smaller radius, where the effect of overmerging is stronger.

The top-right panel of Fig. 6 shows the difference in $P(k)$ caused by this cored profile. As expected, the deficit of the galaxy number at small scales leads to lower power at high k. In addition, the suppression is stronger at low redshift because massive clusters are more abundant at low z. The inset shows the corresponding $\Delta\chi^2_{tot}$ as a function of k_{max}; the systematic shifts dominate at $k_{max} = 1\, h\, Mpc^{-1}$.

We model the interpolated systematic error in the density profile as

$$\tilde{u}_{interp}(k|M) = \tilde{u}_{fid}(k|M) + f_{sys}\, [\tilde{u}_{alt}(k|M) - \tilde{u}_{fid}(k|M)]. \qquad (39)$$

Here our fiducial model is the NFW profile, and the alternative profile is given by equation (37). We again search for the required f_{sys} as a function of k_{max}. The result is shown by the green curve in Fig. 7. Like the c–M relation, the density profile of galaxies does not require more precise calibrations for all practical k_{max}. The results are summarized in the 'Profile' row of Table 2.

6.3 Deviation from the Poisson distribution

In our fiducial model, $P(N_{sat}|M)$ is assumed to be Poisson distributed; that is, the second moment is given by $\langle N_{sat}(N_{sat} - 1)\rangle = \langle N_{sat}\rangle^2$, or $\alpha \equiv \sqrt{\langle N_{sat}(N_{sat}-1)\rangle}/\langle N_{sat}\rangle = 1$. However, Boylan-Kolchin et al. (2010) have shown that the number of subhaloes for a given halo mass deviates from the Poisson distribution (their fig. 8). In addition, Wu et al. (2013a) have shown that the extra-Poisson scatter depends on how subhaloes are chosen and depends on the resolution. Therefore, it is still unclear whether $P(N_{sat}|M)$ follows a Poisson distribution. To assess the impact of the possible extra-Poisson scatter, we adopt $\alpha = 1.02$ in our one-halo term (following Boylan-Kolchin et al. 2010), noting that this choice of α brackets the various possibilities explored in Wu et al. (2013a, fig. 3 therein).

The bottom-left panel in Fig. 6 shows the impact of $\alpha = 1.02$ on $P(k)$, relative to the fiducial Poisson case with $\alpha = 1$. The extra-Poisson scatter only impacts the one-halo term; therefore, the large-scale $P(k)$ is unaffected. At small scales, $P(k)$ is boosted by less than 3 per cent. For different redshifts, $\Delta P/P$ takes off at different k, reflecting the varying scale where one-halo and one-halo terms cross. We also note that at high k, $\Delta P/P$ bends downwards, reflecting the fact that the one-halo term includes $\langle N_{sat}|M\rangle\tilde{u} + \frac{1}{2}\langle N_{sat}(N_{sat} - 1)|M\rangle\tilde{u}^2$. When $P(N_{sat}|M)$ is super-Poisson, more galaxy pairs are expected, and the one-halo term

gets more weighting of $\tilde{u}^2$ ($\tilde{u} < 1$); thus, $P(k)$ becomes lower at high k.

The inset in the bottom-left panel of Fig. 6 shows that the systematic shifts dominate statistical errors at $k_{max} = 0.3 \, h \, \mathrm{Mpc}^{-1}$. We model the partial uncertainties in α as

$$\alpha_{interp} = \alpha_{fid} + f_{sys}(\alpha_{alt} - \alpha_{fid}). \tag{40}$$

The red curve in Fig. 7 shows the required f_{sys} as a function of k_{max}; for $k_{max} = 1 \, h \, \mathrm{Mpc}^{-1}$, the required $f_{sys} = 0.2$. The results are summarized in the $P(N_{sat})$ row of Table 2.

6.4 Velocity bias

The small-scale RSD (also known as the 'Fingers-of-God' effect) are usually modelled as an exponential suppression of power with the term $e^{-(k\sigma_v \mu)^2}$. Here σ_v is the velocity dispersion of *galaxies* inside a cluster $\sigma_v^{gal}(M_{vir})$. Assuming that the motions of galaxies trace those of dark matter particles, we use the velocity dispersion of dark matter particles inside a halo, $\sigma_v^{DM}(M_{vir})$, which has been well established using simulations (Evrard et al. 2008). However, the velocity dispersion of *galaxies* inside a cluster σ_v^{gal} is not necessarily the same as σ_v^{DM}. The ratio between the two is defined as the velocity bias,

$$b_v = \frac{\sigma_v^{gal}}{\sigma_v^{DM}}. \tag{41}$$

The exact value of b_v and its redshift dependence are still under debate. Subhaloes from N-body simulations have shown $b_v > 1$ (e.g. Colín, Klypin & Kravtsov 2000). In addition, Wu et al. (2013b) have shown that the exact value of b_v depends on the selection criteria applied to subhaloes, on the resolution of simulations and on the location of subhaloes. On the other hand, a simulated galaxy population based on assigning subhaloes to dark matter particles (e.g. Faltenbacher & Diemand 2006) or based on hydrodynamical simulations with cooling and star formation (e.g. Lau, Nagai & Kravtsov 2010; Munari et al. 2013) tends to have unbiased velocities.

Since this paper focuses on the possible systematics from N-body simulations, we adopt $b_v > 1$ observed in N-body simulations. Based on the recent calibration from Munari et al. (2013), we adopt the value of the velocity bias to be

$$b_v(z) = 1.1 - \frac{z}{15} \tag{42}$$

(estimated from the dotted curve in their fig. 7A, which corresponds to subhaloes in their N-body simulations.) The bottom-right panel of Fig. 6 shows the systematic error in $P(k)$ caused by this velocity bias. Introducing higher velocity dispersion of galaxies clearly leads to larger suppression on small scales. Note that each curve showcases a dip near $k \approx 1 \, h \, \mathrm{Mpc}^{-1}$, which roughly corresponds to the scale where one-halo and two-halo terms cross. As shown in Section 2.2, the exponential suppression of RSD enters the one-halo and two-halo terms differently; modifying the RSD will therefore slightly change the scale of one-halo to two-halo transition. Also note that the shift in $P(k)$ does not vanish even for very small k, because the exponential suppression enters the two-halo term as well.

The inset in the bottom-right panel of Fig. 6 shows that the systematic shifts associated with velocity bias (difference between no velocity bias and positive velocity bias) dominate the statistical error even for $k_{max} = 0.14 \, h \, \mathrm{Mpc}^{-1}$. Because the deviation of $P(k)$ starts at large scales and increases towards small scales, the velocity bias is dominant among the four sources of systematic errors studied in this paper.

As before, we consider values of the velocity bias that interpolate between the two extreme values considered:

$$b_{v,\,interp} = b_{v,\,fid} + f_{sys}(b_{v,\,alt} - b_{v,\,fid}), \tag{43}$$

where $f_{sys} = 0$ corresponds to $b_v = b_{fid} = 1.0$ while $f_{sys} = 1.0$ corresponds to $b_v = b_{v,\,alt} = 1.1 - z/15$. The cyan curve in Fig. 7 shows the required reduction of f_{sys} for a given k_{max}. For example, to extend the survey just out to the usually conservative wavenumber $k_{max} = 0.3 \, h \, \mathrm{Mpc}^{-1}$, better-than-current knowledge of the velocity bias ($f_{sys} = 0.11 < 1$) is required.[6]

Given that a biased b_v value can lead to a significant systematic error, it is necessary to marginalize over b_v to mitigate the systematic bias. We find that marginalizing over an additional parameter b_v in the Fisher matrix calculation does not significantly degrade the dark energy constraints; the statistical error $\sqrt{\sigma(w_a)\sigma(w_p)}$ is increased by a factor of 2 at most. Since $P(k)$ is sensitive to the change in b_v (as shown in the last panel of Fig. 6), it is not surprising that b_v can be well constrained by data when set free. In addition, the effect of b_v does not seem to be degenerate with the effects of other nuisance parameters and is likely to be well constrained.

While the preparation of this paper was near completion, we learned about the related work from Linder & Samsing (2013). These authors have focused on a particular RSD model from Kwan, Lewis & Linder (2012) and assess the impact of uncertainties in this model on cosmological constraints. These authors have found that, if the model parameters are fixed, they often require sub-per cent accuracy; on the other hand, if these model parameters are self-calibrated using the data, they do not significantly degrade the cosmological constraints. This trend is consistent with our findings regarding fixing versus marginalizing over the velocity bias.

We emphasize that the main goal of this paper is to see to what extent the theoretical uncertainties associated with calibrating galaxy clustering using N-body simulations lead to errors in the cosmological parameters. Given the difficulty of predicting clustering beyond $k \simeq 0.5 \, h \, \mathrm{Mpc}^{-1}$ using purely theoretical methods (e.g. the perturbation theory), resorting to calibration with N-body simulations is required, and this will remain to be the case for years to come. Our findings suggest that the velocity information of galaxies predicted from N-body simulations is likely to generate biases.

7 SUMMARY

As the interpretation of the galaxy clustering measurements from deep, wide redshift surveys often relies on synthetic galaxy catalogues from N-body simulations, the systematic uncertainties in N-body simulations are likely to lead to systematic errors in the cosmological results. In this paper, we have studied several theoretical uncertainties in the predictions of N-body simulations, including the statistics, the spatial distribution and the velocity dispersion of subhaloes. In particular, we have applied the halo model to calculate the galaxy power spectrum $P(k)$, with inputs from recent N-body simulations. We have investigated how the uncertainties from these inputs impact the cosmological interpretation of $P(k)$, and how well these systematics need to be controlled for future surveys. Our main findings can be summarized as follows.

---

[6] Note that Colín et al. (2000) have shown that b_v is scale dependent. Since our scale-independent assumption has already introduced significant systematic shifts, we do not further consider the possible scale dependence of velocity bias in this work but note that the possible scale dependence will further complicate the systematic error.

(i) We have found that the inclusion of the RSD and the covariances between different k modes (the trispectrum contribution to the covariance matrix) is essential to accurately model the information content at small scale.

(ii) Uncertainties in the halo mass function and bias tend to affect $P(k)$ on large scales and can lead to significant systematic errors. However, these effects can be mitigated by measurements of galaxy bias at large scales combined with an independent measurement of σ_8.

(iii) Uncertainties in predicting the halo concentration–mass relation, as well as the deviation from an NFW profile, are unlikely to be a dominant source of systematic error for $k_{max} < 1\,h\,\mathrm{Mpc}^{-1}$.

(iv) Possible deviation of $P(N_{sat})$ from the Poisson distribution, at its current uncertainty level (2 per cent), could be significant for $k_{max} > 0.3\,h\,\mathrm{Mpc}^{-1}$.

(v) Velocity bias is likely to be the most important source of systematic error for upcoming surveys. The current uncertainty of 10 per cent at $z = 0$ is likely to introduce 3 (5) per cent difference in $P(k)$ for $k_{max} = 0.3$ (1) $h\,\mathrm{Mpc}^{-1}$, thus leading to a significant bias in cosmological parameters. Given its predominant role in the systematics, the velocity bias will need to be calibrated internally from the survey or externally with follow-up campaigns.

The sensitivity of $P(k)$ to velocity bias leads to the question of what can be done to alleviate the potential systematic bias. Calibration through both observations and simulations is certainly one obvious solution. Another trick that is increasingly being used for large-scale structure surveys is to self-calibrate the systematic error(s); in the velocity bias case, this would mean marginalizing over b_v. With this marginalization, we expect to be left with vastly diminished biases and only a modest degradation in the cosmological parameters. We do not expect the bias to vanish completely, however, since second-order effects (e.g. redshift and scale dependence of b_v) will remain and will cause systematic shifts. Given that we currently do not have a good model of $b_v(z, k)$, we have not attempted the full self-calibration exercise, but we definitely expect this to be modus operandi of galaxy clustering analyses in the future.

REFERENCES

Abazajian K. et al., 2005, ApJ, 625, 613
Albrecht A. et al., 2006, preprint (astro-ph/0609591)
Anderson L. et al., 2012, MNRAS, 427, 3435
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Bernstein G., Huterer D., 2010, MNRAS, 401, 1399
Bhattacharya S., Habib S., Heitmann K., Vikhlinin A., 2013, ApJ, 766, 32 (B13)
Blake C., Glazebrook K., 2003, ApJ, 594, 665
Blake C. et al., 2011, MNRAS, 418, 1707
Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., 2010, MNRAS, 406, 896
Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001, MNRAS, 321, 559 (B01)
Cacciato M., van den Bosch F. C., More S., Mo H., Yang X., 2013, MNRAS, 430, 767
Chuang C.-H., Wang Y., 2012, MNRAS, 426, 226
Coe D. et al., 2012, ApJ, 757, 22
Cohn J. D., White M., 2008, MNRAS, 385, 2025
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Cole S. et al., 2005, MNRAS, 362, 505
Colín P., Klypin A. A., Kravtsov A. V., 2000, ApJ, 539, 561
Colless M. et al., 2001, MNRAS, 328, 1039
Contreras S., Baugh C., Norberg P., Padilla N., 2013, MNRAS, 432, 2717
Cooray A., Hu W., 2001, ApJ, 554, 56
Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Coupon J. et al., 2012, A&A, 542, A5
Crocce M., Fosalba P., Castander F. J., Gaztañaga E., 2010, MNRAS, 403, 1353
Cui W., Borgani S., Dolag K., Murante G., Tornatore L., 2012, MNRAS, 423, 2279
Cunha C. E., Evrard A. E., 2010, Phys. Rev. D, 81, 083509
Dalal N., Doré O., Huterer D., Shirokov A., 2008, Phys. Rev. D, 77, 123514
de la Torre S., Guzzo L., 2012, MNRAS, 427, 327
Diemand J., Moore B., Stadel J., 2004, MNRAS, 352, 535
Dolag K., Borgani S., Murante G., Springel V., 2009, MNRAS, 399, 497
Drinkwater M. J. et al., 2010, MNRAS, 401, 1429
Duffy A. R., Schaye J., Kay S. T., Dalla Vecchia C., 2008, MNRAS, 390, L64
Duffy A. R., Schaye J., Kay S. T., Dalla Vecchia C., Battye R. A., Booth C. M., 2010, MNRAS, 405, 2161
Eisenstein D. J. et al., 2005, ApJ, 633, 560
Evrard A. E. et al., 2002, ApJ, 573, 7
Evrard A. E. et al., 2008, ApJ, 672, 122
Faltenbacher A., Diemand J., 2006, MNRAS, 369, 1698
Gao L., De Lucia G., White S. D. M., Jenkins A., 2004, MNRAS, 352, L1
Geller M. J., Huchra J. P., 1989, Sci, 246, 897
Groth E. J., Peebles P. J. E., 1977, ApJ, 217, 385
Guo Q. et al., 2011, MNRAS, 413, 101
Heitmann K., White M., Wagner C., Habib S., Higdon D., 2010, ApJ, 715, 104
Hu W., Kravtsov A. V., 2003, ApJ, 584, 702
Huchra J., Davis M., Latham D., Tonry J., 1983, ApJS, 52, 89
Huterer D., Turner M. S., 2001, Phys. Rev. D, 64, 123527
Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, MNRAS, 321, 372
Jennings E., Baugh C. M., Pascoli S., 2011, MNRAS, 410, 2081
Johnston D. E. et al., 2007, preprint (arXiv:0709.1159)
Kaiser N., 1987, MNRAS, 227, 1
Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201
Knox L., Scoccimarro R., Dodelson S., 1998, Phys. Rev. Lett., 81, 2004
Komatsu E. et al., 2011, ApJS, 192, 18
Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, ApJ, 609, 35
Kwan J., Lewis G. F., Linder E. V., 2012, ApJ, 748, 78
Kwan J., Bhattacharya S., Heitmann K., Habib S., 2013, ApJ, 768, 123
Lau E. T., Kravtsov A. V., Nagai D., 2009, ApJ, 705, 1129
Lau E. T., Nagai D., Kravtsov A. V., 2010, ApJ, 708, 1419
Le Fèvre O. et al., 2005, A&A, 439, 845
Linder E. V., Samsing J., 2013, J. Cosmol. Astropart. Phys., 2013, 25
Lukić Z., Heitmann K., Habib S., Bashinsky S., Ricker P. M., 2007, ApJ, 671, 1160
Lukić Z., Reed D., Habib S., Heitmann K., 2009, ApJ, 692, 217
Macciò A. V., Moore B., Stadel J., Diemand J., 2006, MNRAS, 366, 1529
Macciò A. V., Dutton A. A., van den Bosch F. C., 2008, MNRAS, 391, 1940
Maddox S. J., Efstathiou G., Sutherland W. J., Loveday J., 1990, MNRAS, 242, 43P
Ma C.-P., Maggiore M., Riotto A., Zhang J., 2011, MNRAS, 411, 2644
Manera M., Gaztañaga E., 2011, MNRAS, 415, 383
Marín F. A. et al., 2013, MNRAS, 432, 2654
Miller C. J., Nichol R. C., Batuski D. J., 2001, ApJ, 555, 68
Munari E., Biviano A., Borgani S., Murante G., Fabjan D., 2013, MNRAS, 430, 2638
Nagai D., Kravtsov A. V., 2005, ApJ, 618, 557

Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Neto A. F. et al., 2007, MNRAS, 381, 1450
Oguri M., Bayliss M. B., Dahle H., Sharon K., Gladders M. D., Natarajan P., Hennawi J. F., Koester B. P., 2012, MNRAS, 420, 3213
Paranjape A., Sheth R. K., Desjacques V., 2013, MNRAS, 431, 1503
Parkinson D. et al., 2012, Phys. Rev. D, 86, 103518
Peacock J. A., 1999, Cosmological Physics. Cambridge Univ. Press, Cambridge, UK
Peacock J. A., Dodds S. J., 1994, MNRAS, 267, 1020
Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144
Percival W. J. et al., 2010, MNRAS, 401, 2148
Planck Collaboration, 2013, preprint (arXiv:1303.5084)
Prada F., Klypin A. A., Cuesta A. J., Betancort-Rijo J. E., Primack J., 2012, MNRAS, 423, 3018
Press W. H., Schechter P., 1974, ApJ, 187, 425
Rasia E., Borgani S., Ettori S., Mazzotta P., Meneghetti M., 2013, preprint (arXiv:1301.7476)
Reed D., Gardner J., Quinn T., Stadel J., Fardal M., Lake G., Governato F., 2003, MNRAS, 346, 565
Reed D. S., Smith R. E., Potter D., Schneider A., Stadel J., Moore B., 2013, MNRAS, 431, 1866
Reid B. A. et al., 2010, MNRAS, 404, 60
Rozo E. et al., 2010, ApJ, 708, 645
Rudd D. H., Zentner A. R., Kravtsov A. V., 2008, ApJ, 672, 19
Scherrer R. J., Bertschinger E., 1991, ApJ, 381, 349
Schlegel D., White M., Eisenstein D., 2009, preprint (arXiv:0902.4680)
Scoccimarro R., Zaldarriaga M., Hui L., 1999, ApJ, 527, 1
Scoccimarro R., Sheth R. K., Hui L., Jain B., 2001, ApJ, 546, 20
Seljak U., 2000, MNRAS, 318, 203
Seljak U., 2001, MNRAS, 325, 1359
Seljak U., 2009, Phys. Rev. Lett., 102, 021302
Seo H.-J., Eisenstein D. J., 2003, ApJ, 598, 720
Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1
Simha V., Weinberg D. H., Davé R., Fardal M., Katz N., Oppenheimer B. D., 2012, MNRAS, 423, 3458
Smith R. E. et al., 2003, MNRAS, 341, 1311
Smith R. E., Scoccimarro R., Sheth R. K., 2007, Phys. Rev. D, 75, 063512
Smith R. E., Reed D. S., Potter D., Marian L., Crocce M., Moore B., 2012, preprint (arXiv:1211.6434)
Somerville R. S., Primack J. R., 1999, MNRAS, 310, 1087
Stanek R., Rasia E., Evrard A. E., Pearce F., Gazzola L., 2010, ApJ, 715, 1508
Tegmark M. et al., 2006, Phys. Rev. D, 74, 123507
Tinker J. L., Weinberg D. H., Zheng Z., Zehavi I., 2005, ApJ, 631, 41
Tinker J. L., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709
Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, ApJ, 724, 878
Tinker J. L. et al., 2012, ApJ, 745, 16
Vale A., Ostriker J. P., 2004, MNRAS, 353, 189
van den Bosch F. C. et al., 2007, MNRAS, 376, 841
Wang L., Li C., Kauffmann G., De Lucia G., 2006, MNRAS, 371, 537
Warren M. S., Abazajian K., Holz D. E., Teodoro L., 2006, ApJ, 646, 881
Watson W. A., Iliev I. T., D'Aloisio A., Knebe A., Shapiro P. R., Yepes G., 2013, MNRAS, 433, 1230
Weinberg D. H., Colombi S., Davé R., Katz N., 2008, ApJ, 678, 6
Wetzel A. R., White M., 2010, MNRAS, 403, 1072
White M., 2001, MNRAS, 321, 1
White S. D. M., Frenk C. S., 1991, ApJ, 379, 52
Wu H.-Y., Zentner A. R., Wechsler R. H., 2010, ApJ, 713, 856
Wu H.-Y., Hahn O., Wechsler R. H., Behroozi P. S., Mao Y.-Y., 2013a, ApJ, 767, 23
Wu H.-Y., Hahn O., Evrard A. E., Wechsler R. H., Dolag K., 2013b, preprint (arXiv:1307.0011)
Xu X., Cuesta A. J., Padmanabhan N., Eisenstein D. J., McBride C. K., 2012, MNRAS, 431, 2834
York D. G. et al., 2000, AJ, 120, 1579

Zehavi I. et al., 2011, ApJ, 736, 59
Zheng Z., Weinberg D. H., 2007, ApJ, 659, 1
Zheng Z. et al., 2005, ApJ, 633, 791

APPENDIX A: GALAXY NUMBER DENSITY PROFILE

Fig. A1 presents the galaxy number density profile based on which we model its theoretical uncertainties. The coloured curves are based on the Consuelo simulation – an N-body simulation with 1400^3 particles in a volume of side length $420\,h^{-1}\,\mathrm{M}_\odot$. The mass resolution is $1.9 \times 10^9\,h^{-1}\,\mathrm{M}_\odot$ and the force resolution is $8\,h^{-1}$ kpc. We assign each subhalo a luminosity value using the v_{max}^{pk}–luminosity relation based on a subhalo abundance matching model (Behroozi, private communication), where v_{max}^{pk} is the subhalo's peak maximum circular velocity in its history.

We compare the simulated galaxy density profiles with the results from the SDSS maxBCG cluster catalogue as presented in Tinker et al. (2012). The grey dashed curves correspond to three of the richness bins of maxBCG. From the Consuelo simulation, we select clusters in a way that they have approximately the same mass distribution as the maxBCG cluster sample (Johnston et al. 2007). Each maxBCG cluster is assigned a richness value N_{200}, which is the number of red-sequence galaxies brighter than $^{0.25}M_i = -19.2$ within r_{200}. Here r_{200} is defined as the radius within which the density of galaxies is 200 times the mean density of galaxies. At large radii $(>0.5\,h^{-1}\,\mathrm{Mpc})$, the simulation and observation agree well. This agreement naturally comes from our mass selection and abundance matching without tuning the normalization.

However, discrepancy between simulation and observation occurs at small radius. As can be seen, the subhalo number density
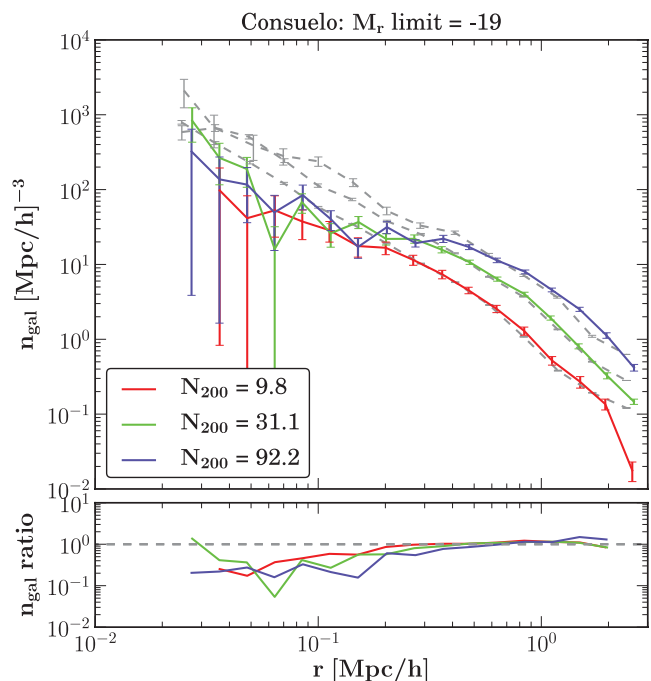

Figure A1. Galaxy number density profile from the Consuelo simulation. The coloured curves show galaxies in the simulation, while the grey dashed curves correspond to the observations of SDSS from Tinker et al. (2012). As can be seen, the simulated galaxy population near the centre of clusters tends to have a shallower distribution than the real galaxy population. The difference is larger for more massive clusters. This figure is adapted from Wu et al. (in preparation).

profile measured from the simulation is shallower than the galaxy density profile measured from SDSS and is also shallower than the NFW profile. This discrepancy is stronger for more massive host haloes. Wu et al. (in preparation) further demonstrate that (1) the discrepancy is also stronger for dimmer galaxies, (2) the trend exists in several state-of-the-art *N*-body simulations using different algorithms and resolutions, and (3) the incompleteness of subhaloes depends on the radius, the mass of the host halo and the mass of the subhalo. It has been shown that the deficit of simulated galaxies near the centre of massive haloes can be alleviated in hydrodynamical simulations that include cooling and star formation (e.g. Weinberg et al. 2008; Dolag et al. 2009). Therefore, it is highly likely that this deficit presents a fundamental limitation of *N*-body simulations and needs to be taken into account when we use *N*-body simulations to model the galaxy population in massive clusters.

APPENDIX B: DERIVATION OF THE GALAXY POWER SPECTRUM

In this appendix, we provide the detailed derivation of the galaxy power spectrum, mainly following the derivations in Scherrer & Bertschinger (1991), Seljak (2000) and Cooray & Sheth (2002), in order to clarify possible confusions originated from different conventions. Let us assume that dark matter halo i with mass M_i is located at x_i. It has N_i galaxies, whose spatial distribution is described by $u(x - x_i|M_i)$ [normalized so that $\int d^3x\, u(x|M) = 1$]. The galaxy number density field can be described by summing over all haloes in the universe:

$$n_{\text{gal}}(x) = \sum_i N_i\, u(x - x_i|M_i)$$

$$= \sum_i \int dM \delta_D(M - M_i)$$

$$\times \int d^3x' \delta_D(x' - x_i) N_i u\left(x - x'|M\right), \quad \text{(B1)}$$

where we insert Dirac delta functions for M and x'. If we define

$$\left\langle \sum_i \delta_D(M - M_i)\delta_D(x' - x_i)\, N_i \right\rangle \equiv n(M)\, \langle N|M \rangle, \quad \text{(B2)}$$

then the mean galaxy number density is given by

$$\bar{n}_{\text{gal}} = \langle n_{\text{gal}} \rangle = \int dM n(M)\, \langle N|M \rangle, \quad \text{(B3)}$$

where we write $n(M) = dn/dM$ for the halo mass function.

The number density fluctuation of galaxies is defined as

$$\delta_{\text{gal}}(x) = \frac{n_{\text{gal}}(x)}{\bar{n}_{\text{gal}}} - 1. \quad \text{(B4)}$$

The two-point statistics follows the definition:

$$\left\langle \sum_i \delta_D(M_1 - M_i)\delta_D(x_1 - x_i)N_i \sum_j \delta_D(M_2 - M_j)\delta_D(x_2 - x_j)N_j \right\rangle$$

$$\equiv n(M_1)\, \langle N|M_1 \rangle\, n(M_2)\, \langle N|M_2 \rangle$$

$$\times [1 + \xi_{\text{hh}}(M_1, M_2, |x_2 - x_1|)] \quad (i \neq j)$$

$$+ n(M_1) \left\langle \binom{N}{2} \Big| M_1 \right\rangle \delta_D(M_1 - M_2)\delta_D(x_1 - x_2) \quad (i = j), \quad \text{(B5)}$$

where ξ_{hh} is the two-point correlation function of dark matter contributed by two different haloes. The two-point correlation function

for galaxies reads

$$\xi_{\text{gg}}(r) = \left\langle \delta_{\text{gal}}(x)\delta_{\text{gal}}(x + r) \right\rangle$$

$$= \frac{1}{\bar{n}_{\text{gal}}^2} \int dM_1 \int dM_2 \int d^3x_1 \int d^3x_2\, u(x - x_1|M_1)u$$

$$\times (x + r - x_2|M_2) \left\langle \sum_i \delta_D(M_1 - M_i)\delta_D(x_1 - x_i)N_i \right.$$

$$\times \left. \sum_j \delta_D(M_2 - M_j)\delta_D(x_2 - x_j)N_j \right\rangle$$

$$= \xi_{\text{gg}}^{\text{1h}}(r) + \xi_{\text{gg}}^{\text{2h}}(r), \quad \text{(B6)}$$

where

$$\xi_{\text{gg}}^{\text{1h}}(r) = \frac{1}{\bar{n}_{\text{gal}}^2} \int dM n(M) \left\langle \binom{N}{2} \Big| M \right\rangle \int d^3x\, u(x|M)u(x + r|M),$$

$$\xi_{\text{gg}}^{\text{2h}}(r) = \frac{1}{\bar{n}_{\text{gal}}^2} \int dM_1 n(M_1) \langle N|M_1 \rangle \int dM_2 n(M_2) \langle N|M_2 \rangle$$

$$\times \int d^3x_1 \int d^3x_2\, u(x_1|M_1)u(x_2 - r|M_2)$$

$$\times (1 + \xi_{\text{hh}}(M_1, M_2; |x_2 - x_1 - r|)). \quad \text{(B7)}$$

We now turn to the Fourier space. We follow this convention of the Fourier transform

$$\tilde{\delta}(k) = \frac{1}{\sqrt{V}} \int d^3x\, \delta(x)\, e^{-ik\cdot x}. \quad \text{(B8)}$$

The Dirac delta function in k space is defined as

$$\delta_D(k) = \frac{1}{V} \int \frac{d^3x}{(2\pi)^3}\, e^{-ik\cdot x} \quad \text{(dimensionless)}. \quad \text{(B9)}$$

From this convention, the relation between the correlation function and the power spectrum follows:

$$\xi(r) = \frac{1}{(2\pi)^3} \int d^3k\, P(k)e^{-ik\cdot r}. \quad \text{(B10)}$$

The Fourier transform of the density perturbation reads

$$\tilde{\delta}_{\text{gal}}(k) = \frac{1}{\sqrt{V}} \int d^3x\, \delta_{\text{gal}}(x)\, e^{-ik\cdot x}$$

$$= \frac{1}{\bar{n}_{\text{gal}}\sqrt{V}} \sum_i N_i \tilde{u}(k|M_i)e^{-ik\cdot x_i} - (2\pi)^3\sqrt{V}\delta_D(k), \quad \text{(B11)}$$

where

$$\tilde{u}(k|M) = \int d^3x u(x|M)e^{-ik\cdot x}. \quad \text{(B12)}$$

Based on this definition, $\tilde{u} \to 1$ when $k \to 0$ and is dimensionless. Applying the Fourier transform to equation (B5), we obtain

$$\left\langle \sum_i \delta_D(M_1 - M_i)e^{-ik_1\cdot x_i} N_i \sum_j \delta_D(M_2 - M_j)e^{+ik_2\cdot x_j} N_j \right\rangle$$

$$\equiv n(M_1)\, \langle N|M_1 \rangle\, n(M_2)\, \langle N|M_2 \rangle\, (2\pi)^6\, V^2 \delta_D(k_1)\delta_D(k_2)$$

$$+ (2\pi)^3\, V n(M_1)\, \langle N|M_1 \rangle\, n(M_2)\, \langle N|M_2 \rangle\, P_{\text{hh}}(M_1, M_2; k)\delta_D(k_1$$

$$- k_2) + (2\pi)^3\, V n(M_1) \left\langle \binom{N}{2} \Big| M_1 \right\rangle \delta_D(M_1 - M_2)\delta_D(k_1 - k_2). \quad \text{(B13)}$$

We note that under our convention of the Fourier transform, $\int d^3k \delta_D(k) = 1/V$ and $\delta_D(k)\delta_D(k) = \delta_D(k)/(2\pi)^3$.

We are now ready to compute the galaxy power spectrum. Applying the trick of inserting Dirac delta functions and then using equation (B13), we obtain

$$P_{gg}(k) = \frac{1}{(2\pi)^3} \left\langle \tilde{\delta}_{gal}(\boldsymbol{k}) \tilde{\delta}_{gal}^*(\boldsymbol{k}) \right\rangle$$

$$= \frac{1}{(2\pi)^3\, V \bar{n}_{gal}^2} \left\langle \sum_i \tilde{u}(\boldsymbol{k}|M_i) e^{-i\boldsymbol{k}\cdot\boldsymbol{x}_i} N_i \sum_j \tilde{u}^*(\boldsymbol{k}|M_j) e^{i\boldsymbol{k}\cdot\boldsymbol{x}_j} N_j \right\rangle$$

$$= P_{gg}^{1h}(k) + P_{gg}^{2h}(k), \tag{B14}$$

where the two-halo term reads

$$P_{gg}^{2h}(k) = \left[\frac{1}{\bar{n}_{gal}} \int dM \frac{dn}{dM} \langle N|M\rangle\, b(M) \right]^2 P_{lin}(k), \tag{B15}$$

and the one-halo term reads

$$P_{gg}^{1h}(k) = \frac{1}{\bar{n}_{gal}^2} \int dM \frac{dn}{dM} \left\langle \binom{N}{2} \middle| M \right\rangle f(k|M). \tag{B16}$$

Here $\langle \binom{N}{2}|M\rangle f(k|M)$ is the galaxy pair-weighted profile, including the contribution from central and satellite galaxies (Berlind & Weinberg 2002),

$$\left\langle \binom{N}{2} \middle| M \right\rangle f(k|M)$$

$$= \left[\langle N_{sat}|M\rangle\, \tilde{u}(k|M) + \frac{1}{2} \langle N_{sat}(N_{sat}-1)|M\rangle\, |\tilde{u}(k|M)|^2 \right]. \tag{B17}$$

APPENDIX C: DERIVATION OF THE COVARIANCE MATRIX

We now derive the covariance of power spectra at different wavenumbers in equation (17). First, recall the definitions for the power spectrum and trispectrum:

$$\langle \delta(\boldsymbol{k}_1)\delta(\boldsymbol{k}_2) \rangle = (2\pi)^3 \delta_D(\boldsymbol{k}_{12}) P(k_1)$$

$$\langle \delta(\boldsymbol{k}_1)\delta(\boldsymbol{k}_2)\delta(\boldsymbol{k}_3)\delta(\boldsymbol{k}_4) \rangle_c = (2\pi)^3 \delta_D(\boldsymbol{k}_{1234}) T(k_1, k_2, k_3, k_4), \tag{C1}$$

where the subscript c indicates the 'connected' term. Under our convention, $[P] = L^3$ and $[T] = L^6$. For a given realization of the density field $\delta(\boldsymbol{k})$, the estimator of the binned power spectrum is

$$\hat{P}(k_i) = \int_{k_i} \frac{d^3\boldsymbol{k}}{V_s(k_i)} \delta(\boldsymbol{k})\delta(-\boldsymbol{k}), \tag{C2}$$

where $V_s(k_i) = 4\pi k_i^3 \delta \ln k$. Its covariance is

$$C_{ij} = \left\langle \hat{P}(k_i)\hat{P}(k_j) \right\rangle - \left\langle \hat{P}(k_i) \right\rangle \left\langle \hat{P}(k_j) \right\rangle$$

$$= \frac{(2\pi)^3}{V} \frac{2P(k_i)^2}{V_s(k_i)} \delta_{ij} + \bar{T}(k_i, k_j), \tag{C3}$$

where

$$\bar{T}(k_i, k_j) \equiv \int_{k_i} \frac{d^3\boldsymbol{k}_1}{V_s(k_i)} \int_{k_j} \frac{d^3\boldsymbol{k}_2}{V_s(k_j)} T(\boldsymbol{k}_1, -\boldsymbol{k}_1, \boldsymbol{k}_2, -\boldsymbol{k}_2). \tag{C4}$$

Below we provide the derivation. The first term in equation (C3) can be calculated as

$$\left\langle \hat{P}(k_i)\hat{P}(k_j) \right\rangle = \int_{k_i} \frac{d^3\boldsymbol{k}_1}{V_s(k_i)} \int_{k_j} \frac{d^3\boldsymbol{k}_2}{V_s(k_j)} \langle \delta(\boldsymbol{k}_1)\delta(-\boldsymbol{k}_1)\delta(\boldsymbol{k}_2)\delta(-\boldsymbol{k}_2) \rangle, \tag{C5}$$

where the integrand reads

$$\langle \delta(\boldsymbol{k}_1)\delta(-\boldsymbol{k}_1)\delta(\boldsymbol{k}_2)\delta(-\boldsymbol{k}_2) \rangle$$

$$= \left\langle \delta_1 \delta_1^* \delta_2 \delta_2^* \right\rangle_c + \left\langle \delta_1 \delta_1^* \right\rangle \left\langle \delta_2 \delta_2^* \right\rangle + \left\langle \delta_1 \delta_2 \right\rangle \left\langle \delta_1^* \delta_2^* \right\rangle + \left\langle \delta_1 \delta_2^* \right\rangle \left\langle \delta_2 \delta_1^* \right\rangle$$

$$= (2\pi)^3 \delta_D(0) T(\boldsymbol{k}_1, -\boldsymbol{k}_1, \boldsymbol{k}_2, -\boldsymbol{k}_2) \tag{C6}$$

$$+ (2\pi)^6 \delta_D(0) P(\boldsymbol{k}_1)\delta(0)P(\boldsymbol{k}_2) \tag{C7}$$

$$+ (2\pi)^6 \delta_D(\boldsymbol{k}_1 + \boldsymbol{k}_2) P(\boldsymbol{k}_1)\delta(\boldsymbol{k}_1 + \boldsymbol{k}_2)P(\boldsymbol{k}_1) \tag{C8}$$

$$+ (2\pi)^6 \delta_D(\boldsymbol{k}_1 - \boldsymbol{k}_2) P(\boldsymbol{k}_1)\delta(\boldsymbol{k}_1 - \boldsymbol{k}_2)P(\boldsymbol{k}_1). \tag{C9}$$

We note that $\delta_D(0) = \frac{1}{(2\pi)^3}$. Then the contribution from each term reads

$$(C6) \Rightarrow \int_{k_i} \frac{d^3\boldsymbol{k}_1}{V_s(k_i)} \int_{k_j} \frac{d^3\boldsymbol{k}_2}{V_s(k_j)} T(\boldsymbol{k}_1, -\boldsymbol{k}_1, \boldsymbol{k}_2, -\boldsymbol{k}_2) \equiv \bar{T}(k_i, k_j)$$

$$(C7) \Rightarrow \left\langle \hat{P}(k_i) \right\rangle$$

$$\left\langle \hat{P}(k_j) \right\rangle \quad \text{[cancels the second term of equation (C3)]}$$

$$(C8) = (C9) \Rightarrow \int_{k_i} \frac{d^3\boldsymbol{k}_1}{V_s(k_i)} \int_{k_j} \frac{d^3\boldsymbol{k}_2}{V_s(k_j)} (2\pi)^3 \delta_D(\boldsymbol{k}_1 - \boldsymbol{k}_2) P(\boldsymbol{k}_1)P(\boldsymbol{k}_1)$$

$$= \int_{k_i} \frac{d^3\boldsymbol{k}_1}{V_s(k_i)} P(k_1)^2 (2\pi)^3$$

$$\times \int_{k_j} \frac{d^3\boldsymbol{k}_2}{V_s(k_j)} \delta_D(\boldsymbol{k}_1 - \boldsymbol{k}_2) \quad \text{(only non-zero if } k_i = k_j\text{)}$$

$$= \frac{(2\pi)^3}{V_z} \frac{\left\langle \hat{P}(k_i)^2 \right\rangle}{V_s(k_i)} \delta_{ij} \approx \frac{(2\pi)^3}{V_z} \frac{P(k_i)^2}{V_s(k_i)} \delta_{ij}. \tag{C10}$$

The expression of $\bar{T}(k_i, k_j)$ (equation 19) can be obtained using equation (B11) and is similar to the derivation of $P(k)$.

This paper has been typeset from a TeX/LaTeX file prepared by the author.