

SECOND MOMENT OF HALO OCCUPATION NUMBER

ASANTHA COORAY¹

Division of Physics, Mathematics and Astronomy, California Institute of Technology, MS 103-33, Pasadena, CA 91125; asante@caltech.edu

Received 2002 June 20; accepted 2002 July 29; published 2002 August 9

ABSTRACT

The halo approach to large-scale structure provides a physically motivated model to understand the clustering properties of galaxies. An important aspect of the halo model involves a description of how galaxies populate dark matter halos or what is now called the halo occupation distribution. We discuss a way in which clustering information, especially in the nonlinear regime, can be used to determine moments of this halo occupation number. We invert the nonlinear part of the real space power spectrum from the *IRAS* Point Source Catalog galaxy redshift survey to determine the second moment of the halo occupation distribution in a model-independent manner. The precise measurement of higher order correlations can eventually be used to determine successive higher order moments of this distribution.

Subject headings: cosmology: observations — cosmology: theory — galaxies: halos — large-scale structure of universe

1. INTRODUCTION

The halo approach to large-scale structure has now become a useful tool to study and understand clustering properties of dark matter and a number of tracers, including galaxies (see Cooray & Sheth 2002 for a recent review). This approach replaces the complex distribution of dark matter with a collection of collapsed dark matter halos. Thus, necessary inputs for a halo-based model include properties of this dark matter halo population, such as its mass function and the spatial profile of dark matter within each halo (Seljak 2000; Ma & Fry 2000; Scoccimarro et al. 2001; Cooray, Hu, & Miralda-Escudé 2000).

In order to describe clustering aspects beyond dark matter, it is necessary that one understands how the tracer property is related to the dark matter distribution in each halo. In the case of galaxies, an important input is a description of how galaxies populate halos. This is usually achieved by the so-called halo occupation number, in which one describes the mean number of galaxies in dark matter halos as a function of mass and its higher order moments (Jing, Mo, & Boerner 1998; Seljak 2000; Peacock & Smith 2000; Berlind & Weinberg 2002). For the two-point correlation function, one requires information up to the second moment of the halo occupation distribution, while higher order correlations successively depend on increasing moments.

The halo occupation distribution has been widely discussed in the literature in terms of semianalytical models of galaxy formation (e.g., Benson et al. 2001; Somerville et al. 2001). With the advent of a well-defined halo approach to clustering, observational constraints have also begun to appear (e.g., Scoccimarro et al. 2001; Moustakas & Somerville 2002). While expectations for constraints on the halo occupation distribution from current wide-field galaxy surveys, such as the Sloan Digital Sky Survey, are high (Berlind & Weinberg 2002; Scranton 2002), these constraints are all considered under a particular model.

Although descriptions of the halo occupation number based on a specific model are useful, it is probably more useful to consider constraints that are not restricted to a particular description. In this Letter, we consider such an approach and suggest that clustering information, especially in the nonlinear regime, can be used for a reconstruction of various moments of the halo occupation number. We discuss a possible inversion

for this purpose and use results on the nonlinear power spectrum from the *IRAS* Point Source Catalog galaxy redshift (PSCz) survey (Saunders et al. 2000) by Hamilton & Tegmark (2002) to provide a first estimate of the second moment of the halo occupation number. We discuss both the strengths and the limitations of our approach and provide a comparison with model-based descriptions of the halo occupation number.

We provide a general discussion of our method in the next section. When illustrating results in § 3, we take a flat Λ cold dark matter cosmology with parameters $\Omega_c = 0.3$, $\Omega_b = 0.05$, $\Omega_\Lambda = 0.65$, $h = 0.65$, $n = 1$, and $\delta_H = 4.2 \times 10^{-5}$.

2. HALO APPROACH TO GALAXY CLUSTERING

The halo approach to galaxy clustering assumes that large-scale structure can be described by a distribution of dark matter halos while galaxies themselves form within these halos. At the two-point level of correlations, the contribution can be written as a sum of correlations of galaxies that occupy two different halos and the correlation of galaxies within a single halo. Assuming a halo population with a mass function given by $n(m)$ and that galaxies trace the dark matter in each halo randomly, we can write the total contribution to the power spectrum of galaxies as

$$P_{\text{gal}}(k) = P_{\text{gal}}^{1h}(k) + P_{\text{gal}}^{2h}(k), \quad (1a)$$

where

$$P_{\text{gal}}^{1h}(k) = \int dm n(m) \frac{\langle N_{\text{gal}}(N_{\text{gal}} - 1) | m \rangle}{\bar{n}_{\text{gal}}^2} |y(k|m)|^2, \quad (1b)$$

$$P_{\text{gal}}^{2h}(k) = P^{\text{lin}}(k) \left[\int dm n(m) b_1(m) \frac{\langle N_{\text{gal}} | m \rangle}{\bar{n}_{\text{gal}}} y(k|m) \right]^2. \quad (1c)$$

Here $y(k|m)$ represents the three-dimensional Fourier transform of the dark matter profile. In general, the 1-halo term dominates the total contribution to the power spectrum in the nonlinear regime while the 2-halo term captures the large-scale correlations in the linear regime.

In equations (1b) and (1c), $\langle N_{\text{gal}} | m \rangle$ and $\langle N_{\text{gal}}(N_{\text{gal}} - 1) | m \rangle$ are the first and second moments of the galaxy occupation distribution.

¹ Sherman Fairchild Senior Research Fellow.

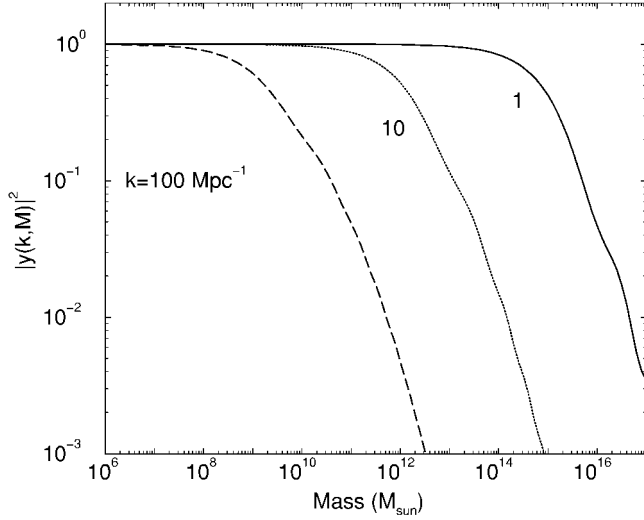


FIG. 1.—The $|y(k|m)|^2$ as a function of mass for three different values of k , as written on the plot, at redshift of zero. We assume dark matter profiles under the description of Navarro, Frenk, & White (1996).

bution, respectively, \bar{n}_{gal} is mean number density of galaxies,

$$\bar{n}_{\text{gal}} = \int dm n(m) \langle N_{\text{gal}} | m \rangle, \quad (2)$$

$b_1(m)$ is the first-order halo bias (Mo, Jing, & White 1997), $n(m)$ is the halo mass function, and $P^{\text{lin}}(k)$ is the linear power spectrum of the density field. On large scales where the 2-halo term dominates and $y(k|m) \rightarrow 1$, the galaxy power spectrum simplifies to

$$P_{\text{gal}}(k) \approx b_{\text{gal}}^2 P^{\text{lin}}(k), \quad (3)$$

where

$$b_{\text{gal}} = \int dm n(m) b_1(m) \frac{\langle N_{\text{gal}} | m \rangle}{\bar{n}_{\text{gal}}} \quad (4)$$

represents the large-scale constant bias factor of the galaxy population. Although not considered in detail, there are slight modifications to equations (1a)–(1b), especially if one is interested in accounting for other possibilities, such as a single galaxy always forming at the center of each dark matter halo, and the fact that galaxies may not trace the dark matter perfectly.

At nonlinear scales, $y(k|m) \ll 1$, and the 1-halo term dominates. Since $y(k|m)$ is no longer constant, if $y(k|m)$ is assumed a priori from a certain dark matter profile, we can consider a possible inversion of the power spectrum measurements to reconstruct $f(m) \equiv n(m) \langle N_{\text{gal}} (N_{\text{gal}} - 1) | m \rangle / \bar{n}_{\text{gal}}^2$. With information related to the halo mass function and mean density of galaxies, which comes directly from data, one can obtain a phenomenological estimate of $\langle N_{\text{gal}} (N_{\text{gal}} - 1) | m \rangle$. Note that in the large-scale regime, because of the scale-independent constant bias, the equation is noninvertible. Therefore, clustering information cannot be used to reconstruct accurately the mean of the halo occupation number.

Although our use of outside knowledge on $y(k|m)$ and $n(m)$ may make the extraction of the second moment model-dependent, one should note that these are effectively properties of the dark matter and are well determined from N -body nu-

merical simulations. The fact that we have a potential method to estimate information on the galaxy side (mainly details on the halo occupation number without resorting to any models) is extremely important. This is the main result of this Letter. We now consider the possibility for an inversion and an application of our suggestion to galaxy clustering data from the PSCz survey.

3. INVERSION

In order to invert the nonlinear power spectrum so as to estimate the information related to the second moment of the halo occupation number, we follow standard approaches in the literature (e.g., Dodelson et al. 2002; Cooray 2002). These inversions are usually considered in order to recover three-dimensional clustering information from two-dimensional angular clustering data. The inversion problem here is similar: given estimates of $P_{\text{gal}}^{2h}(k)$, and information related $y(k|m)$, we would like to estimate the function $f(m)$ defined earlier.

We can write the associated inversion equation as

$$\mathbf{P} = \mathbf{Y}\mathbf{M} + \mathbf{n}, \quad (5)$$

where \mathbf{P} is a vector containing the data related to $P_{\text{gal}}^{2h}(k)$, with an associated noise vector \mathbf{n} , and \mathbf{Y} is a matrix containing kernels at each k_i (where the nonlinear power spectrum is measured) and at each m_i , for which the galaxy-weighted mass function information is desired. The inversion problem, stated in terms of this equation, involves estimating \mathbf{M} given other vectors and the matrix \mathbf{Y} . Note that the matrix \mathbf{Y} differs from kernels defined by $|y(k|m)|^2$ because of an additional factor of dm . By appropriately renormalizing equation (5) with noise, following Dodelson et al. (2002), we can consider the minimum variance estimate of $f(m)$.

In Figure 1, we show $|y(k|m)|^2$ as a function of halo mass m for three different values of k and at a redshift of zero. The behavior of these kernels is important for the inversion since they determine how likely the inversion will be stable and to what extent information can be extracted from $f(m)$. The effective width of $|y(k|m)|^2$ increases with decreasing k , which is similar to the behavior one observes of kernels associated with the inversion of a simple galaxy correlation function. The kernels in the latter case involve a J_0 , the zeroth-order Bessel function, which, when plotted, looks similar in shape to associated kernels here, except for the ringing part associated with J_0 functions that oscillate between positive and negative values.

The general behavior of present kernels, as a function of k and m , is also consistent with how the halo model contributes to the nonlinear power spectrum: at small physical scales, contributions come from the low-mass halos, while at large physical scales or small k -values, contributions to the power spectrum come from the whole mass range. The overlap of kernels with decreasing k suggests that any estimate of the nonlinear power spectrum is likely to be highly correlated. This is, in fact, true when one considers the covariance resulting from the four-point correlation function or the trispectrum (e.g., Cooray & Hu 2001). Thus, any measurement of the nonlinear power spectrum should be considered with its associated covariance matrix, and correlations should be properly accounted for when inverting the power spectrum to determine any physical property, such as $f(m)$ defined above.

In terms of published analyses of clustering, this information, unfortunately, is not fully available to us from the literature. The best published estimate of the nonlinear galaxy power

spectrum, so far, comes from the PSCz survey by Hamilton & Tegmark (2002). Given the lack of knowledge of its covariance, as advised by one of the authors, we used the “prewhitened” power spectrum estimated by the same authors to carry out an inversion as part of this study. It is suggested in Hamilton & Tegmark (2002) that estimates of the prewhitened power spectrum are decorrelated, and we used estimates and their errors as part of this inversion without accounting for the correlations.

For the purpose of this inversion, we define the nonlinear part as the power spectrum from $k > 1 \ h \text{ Mpc}^{-1}$. At small k -values, we expect a fractional contribution from the 2-halo term or from the correlated part between galaxies in different halos. This contribution, however, decreases rapidly with an increasing value for k . We use power spectrum estimates out to $k \sim 300 \ h \text{ Mpc}^{-1}$, although estimates beyond $\sim 100 \ h \text{ Mpc}^{-1}$ from the PSCz survey are noise-dominated. Following the redshift distribution of the PSCz survey as measured by Saunders et al. (2000), we assume a redshift of 0.028 for the full nonlinear power spectrum. It is likely that estimates of the nonlinear power are redshift-dependent (Scranton & Dodelson 2000); however, we have no useful information to account for such variations.

When converting $f(m)$ estimates to the second moment of the halo occupation number, we require information on the number density of PSCz galaxies at the mean redshift of the estimated nonlinear power spectrum. We obtained this information using the luminosity function of PSCz galaxies as calculated by Seaborn et al. (1999; see also Saunders et al. 1990), but we have not attempted to account for any variations in the density from the original PSCz survey to the one utilized in estimating the power spectrum by Hamilton & Tegmark (2002). However, we expect any changes on this aspect, if any at all, to be minor. Also, in order to convert $f(m)$ to the second moment, we assume the halo mass function as described by Press & Schechter (1974) theory. Again, we make no attempt to incorporate any uncertainties in the mass function.

We summarize our results on the second moment of the halo occupation distribution in Figure 2. Here we plot $\langle N_{\text{gal}}(N_{\text{gal}} - 1) | m \rangle^{1/2}$ as a function of the halo mass m . We present two separate sets of estimates of this quantity, with bands (on the mass axis) of the second set shifted relative to the bands of the first set. This shifting demonstrates the robustness of the inversion and follows from the now well-known analysis technique introduced by the cosmic microwave background experimentalists. Note that our estimates are likely to be contaminated by assumptions used in the analysis. One of the main drawbacks is the lack of an accounting of full covariance, or associated correlations, between power spectrum measurements. Another is the assumption that galaxies are randomly distributed in each halo, with no preference for the halo center. The latter assumption is likely to affect estimates at the low-mass end, especially when $\langle N_{\text{gal}}(N_{\text{gal}} - 1) | m \rangle^{1/2}$ goes below 1. Since we find this to be the case only in our first two bins, we have not attempted to correct for this any further.

The second moment of the halo occupation number estimated here should be considered as the value at the mean redshift of the PSCz survey. The estimates clearly show the power-law behavior of the halo occupation number. If the distribution can be described by a Poisson distribution, then we can write $\langle N_{\text{gal}}(N_{\text{gal}} - 1) | m \rangle^{1/2} = \langle N_{\text{gal}} | m \rangle$. For comparison, we also plot predictions for the mean number, $\langle N_{\text{gal}} | m \rangle$, following various model descriptions using results from the semianalytical models of galaxy formation by Scranton (2002) and from power-law fits to the Automatic Plate Measuring (APM) Facility clus-

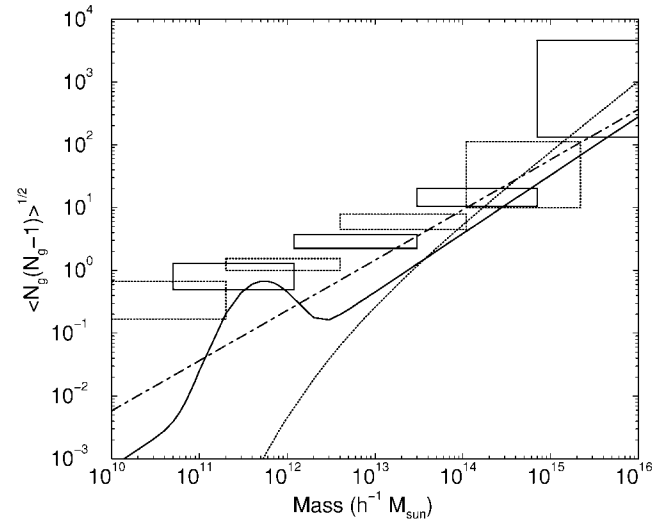


FIG. 2.—Second moment of the halo occupation number, plotted as $[\langle N_{\text{gal}}(N_{\text{gal}} - 1) | m \rangle]^{1/2}$, as a function of halo mass m . The two sets of error boxes (solid and dotted lines) show two independent sets of estimates with bins of one set shifted relative to those of the other. For comparison, we show three estimates of the second moment of the halo occupation: the blue (solid line) and red (dotted line) galaxy model of Scranton (2002) and the best-fit blue galaxy model of Scoccimarro et al. (2001) based on APM data (dot-dashed line) with normalization following Sheth & Diaferio (2001).

tering data by Scoccimarro et al. (2001). The latter can be described as $\langle N_{\text{gal}} | m \rangle = A(m/m_0)^{0.8}$. We normalize this curve using $m_0 = 4 \times 10^{12} \ h^{-1} M_\odot$ and $A = 0.7$ following Sheth & Diaferio (2001).

At intermediate-mass ranges and below, estimated band powers generally lie above these descriptions, suggesting that $\langle N_{\text{gal}}(N_{\text{gal}} - 1) | m \rangle^{1/2}$ may follow as $\alpha(m) \langle N_{\text{gal}} | m \rangle$. There are descriptive models available in the literature for $\alpha(m)$ (Scoccimarro et al. 2001; Scranton 2002). These, however, indicate $\alpha(m) < 1$ such that the halo occupation number is sub-Poissonian at the low-mass end. This is contrary to the estimates that we have obtained. In general, Scranton’s blue galaxy model is more consistent with our estimates, although, a priori, there is no reason to believe why PSCz galaxies should be described by such a model.

Assuming the second moment can be described by a simple power law such that $\langle N_{\text{gal}}(N_{\text{gal}} - 1) | m \rangle^{1/2} = (m/m_0)^p$, we fitted our band power estimates to constrain (m_0, p) . When estimating the likelihood, we use the full covariance matrix of our estimates of $\langle N_{\text{gal}}(N_{\text{gal}} - 1) | m \rangle^{1/2}$. We show 1σ , 2σ , and 3σ contours in Figure 3. In general, our estimates are consistent with a wide range of values of m_0 , from 10^{10} to 10^{13} , and p , from 0.2 to 1, at the 3σ level, while power laws with slopes less than 0.8 are consistent at the 2σ level.

The present technique can be extended for higher order correlations as well. Note that in the nonlinear regime, the p -point power spectrum for equal length configurations can be written as

$$T_p^{1h}(k) = \int dm n(m) \frac{\langle N_{\text{gal}}(N_{\text{gal}} - 1), \dots, (N_{\text{gal}} - n) | m \rangle}{\bar{n}_{\text{gal}}^p} |y(k|m)|^p, \quad (6)$$

where $n = p - 1$. The inversion technique is effectively similar and can be used to determine related higher order moments of the galaxy halo occupation number. And, once again, such an

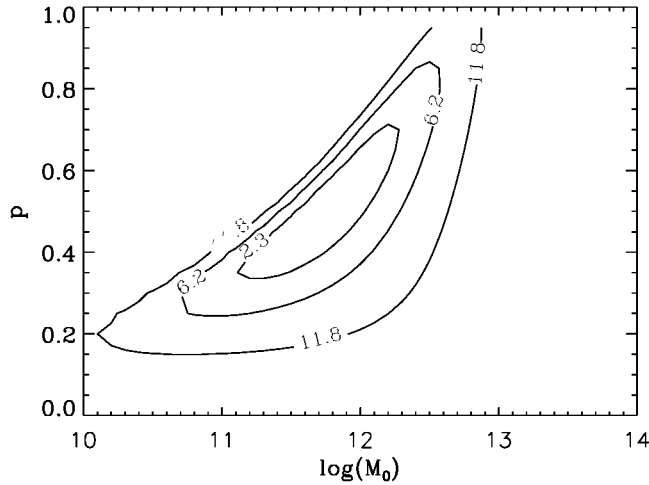


FIG. 3.—Constraints on the slope, p , and normalization, m_0 , of the halo occupation number from the PSCz data. We show 1σ , 2σ , and 3σ contours levels that are labeled as $\Delta\chi^2 = 2.3, 6.2$, and 11.8 , respectively.

approach will require higher order p -point power estimates down to the highly nonlinear regime and, more importantly, a proper accounting of associated errors. As always, we look forward to the day such a study can be carried out with observational measurements from ongoing and upcoming wide-field surveys such as the Sloan Digital Sky Survey and many others.

4. SUMMARY

The halo approach to large-scale structure provides a physically motivated technique for studying the clustering properties of galaxies (Cooray & Sheth 2002 and references therein). A necessary, and important, ingredient for a halo-based clustering calculation involves the description of how galaxies populate dark matter halos, the so-called halo occupation number. We have raised the possibility for a direct study of moments of the halo occupation number using an inversion of the nonlinear clustering power (and p -point) spectrum measurements. We have considered an application of this suggestion utilizing the PSCz power spectrum estimated by Hamilton & Tegmark (2002). Our estimates on the second moment of the halo occupation number are consistent with power-law models over five decades in mass and with certain model descriptions in the literature. With an expected increase in the measurements of the clustering in the nonlinear regime and in the associated measurements of covariance, we expect that an analysis like the one suggested here will eventually make it possible for a detailed understanding of the nature of galaxy occupation in halos.

This research was supported at Caltech by a senior research fellowship from the Sherman Fairchild foundation and a DOE grant. We thank Ryan Scranton and Andrew Hamilton for useful suggestions and appreciate the quite atmosphere at the Aspen Center for Physics where this work was initiated.

REFERENCES

- Benson, A. J., Pearce, F. R., Frenck, C. S., Baugh, C. M., & Jenkins, A. 2001, *MNRAS*, 320, 261
- Berlind, A. B., & Weinberg, D. H. 2002, *ApJ*, 575, 587
- Cooray, A. 2002, *MNRAS*, submitted (astro-ph/0206068)
- Cooray, A., & Hu, W. 2001, *ApJ*, 554, 56
- Cooray, A., Hu, W., & Miralda-Escudé, J. 2000, *ApJ*, 535, L9
- Cooray, A., & Sheth, R. 2002, *Phys. Rep.*, in press (astro-ph/0206508)
- Dodelson, S., et al. 2002, *ApJ*, 572, 140
- Hamilton, A. J. S., & Tegmark, M. 2002, *MNRAS*, 330, 506
- Jing, Y. P., Mo, H. J., & Boerner, G. 1998, *ApJ*, 494, 1
- Ma, C.-P., & Fry, J. N. 2000, *ApJ*, 543, 503
- Mo, H. J., Jing, Y. P., & White, S. D. M. 1997, *MNRAS*, 284, 189
- Moustakas, L. A., & Somerville, R. S. 2002, *ApJ*, in press (astro-ph/0110584)
- Navarro, J., Frenck, C., & White, S. D. M. 1996, *ApJ*, 462, 563
- Peacock, J. A., & Smith, R. E. 2000, *MNRAS*, 318, 1144
- Press, W. H., & Schechter, P. 1974, *ApJ*, 187, 425
- Saunders, W., Rowan-Robinson, M., Lawrence, A., Efstathiou, G., Kaiser, N., Ellis, R. S., & Frenk, C. S. 1990, *MNRAS*, 242, 318
- Saunders, W., et al. 2000, *MNRAS*, 317, 55
- Scoccimarro, R., Sheth, R. K., Hui, L., & Jain, B. 2001, *ApJ*, 546, 20
- Scranton, R. 2002, *ApJ*, submitted (astro-ph/0205517)
- Scranton, R., & Dodelson, S. 2000, *MNRAS*, submitted (astro-ph/0003034)
- Seaborne, M. D., et al. 1999, *MNRAS*, 309, 89
- Seljak, U. 2000, *MNRAS*, 318, 203
- Sheth, R. K., & Diaferio, A. 2001, *MNRAS*, 322, 901
- Somerville, R. S., Lemson, G., Sigad, Y., Dekel, A., Kauffmann, G., & White, S. D. M. 2001, *MNRAS*, 320, 289