BREAKING THE DEGENERACIES BETWEEN COSMOLOGY AND GALAXY BIAS

ZHENG ZHENG\(^1\)\(^2\) AND DAVID H. WEINBERG\(^3\)

Received 2005 December 2; accepted 2006 December 23

ABSTRACT

Adopting the framework of the halo occupation distribution (HOD), we investigate the ability of galaxy clustering measurements to simultaneously constrain cosmological parameters and galaxy bias. Starting with a fiducial cosmological model and galaxy HOD, we calculate spatial clustering observables on a range of length and mass scales, dynamical clustering observables that depend on galaxy peculiar velocities, and the galaxy-matter cross-correlation measurable by weak lensing. We then change one or more cosmological parameters and use \(\chi^2\) minimization to find the galaxy HOD that best reproduces the original clustering. Our parameterization of the HOD incorporates a flexible relation between galaxy occupation numbers and halo mass and allows spatial and velocity bias of galaxies within dark matter halos. Despite this flexibility, we find that changes to the HOD cannot mask substantial changes to the matter density \(\Omega_m\), the matter clustering amplitude \(\sigma_8\), or the shape parameter \(\Gamma\) of the linear matter power spectrum: cosmology and bias are not degenerate. With the conservative assumption of 10% fractional errors, the set of observables considered here can provide \(\sim 10\%\) \((1 \sigma)\) constraints on \(\sigma_8\), \(\Omega_m\), and \(\Gamma\), using galaxy clustering data alone. The combination \(\sigma_8 \Omega^{0.75}_m\) is constrained to \(\sim 5\%\). In combination with traditional methods that focus on large-scale structure in the “perturbative” regime, HOD modeling can greatly amplify the cosmological power of galaxy redshift surveys by taking advantage of high-precision clustering measurements at small and intermediate scales (from sub-Mpc to \(\sim 20 \, h^{-1} \, \text{Mpc}\)). At the same time, the inferred constraints on the galaxy HOD provide valuable tests of galaxy formation theory.

Subject headings: cosmology: theory — dark matter — galaxies: formation — galaxies: halos — large-scale structure of universe

Online material: color figures

1 INTRODUCTION

From the 1970s through the early 1990s, studies of galaxy clustering drove much of the progress in cosmology. Measurements of steadily improving dynamic range and precision demonstrated good agreement with the predictions of a cosmological model incorporating scale-invariant, Gaussian primeval fluctuations modulated by the transfer function expected in a universe dominated by cold dark matter (CDM) with \(\Omega_m h \sim 0.2\) (where \(\Omega_m\) is the matter density parameter and \(h = H_0/100 \, \text{km} \, \text{s}^{-1} \, \text{Mpc}^{-1}\)). The advent of multicolor galaxy redshift surveys and improved photometric input catalogs has dramatically improved the precision of clustering measurements over the last decade, beginning with the Las Campanas Redshift Survey (LCRS; Shectman et al. 1996) and continuing with the Two Degree Field Galaxy Redshift Survey (2dFGRS; Colless et al. 2001) and the Sloan Digital Sky Survey (SDSS; York et al. 2000; Abazajian et al. 2004). In parallel, numerical simulations and numerically tested analytic approximations have turned the task of calculating nonlinear dark matter clustering from specified initial conditions into an essentially solved problem. The principal obstacle to drawing cosmological inferences from galaxy clustering measurements is now the uncertainty in the relation between the distribution of observable galaxies and the underlying distribution of dark matter, the problem known as galaxy bias. Much of the cosmological progress in the last decade has been driven by observations that circumvent this complication, such as cosmic microwave background (CMB) anisotropies, weak gravitational lensing, the Ly\(\alpha\) forest, and the Type Ia supernova diagram.

These new observables favor an inflationary, low-\(\Omega_m\), CDM-dominated model similar to that originally suggested by galaxy clustering, which in turn implies that the galaxies that dominate typical optically selected galaxy surveys must be approximately unbiased, in the sense that the rms galaxy count fluctuations are similar to the rms dark matter density fluctuations on large scales (see, e.g., Lahav et al. 2002). However, observed galaxy clustering varies systematically with galaxy luminosity, color, and spectral or morphological type (Norberg et al. 2002; Zehavi et al. 2005 and numerous references therein), and reproducing the observed galaxy correlation function in an inflationary CDM model requires that the bias of the correlation function vary with separation on scales below a few megaparsecs (Jenkins et al. 1998; Zehavi et al. 2004). The advances in observational cosmology over the last few years have also raised the stakes for galaxy clustering studies. We are no longer interested in, for example, distinguishing \(\Omega_m \sim 0.3\) from \(\Omega_m = 1\); instead, we want to constrain \(\Omega_m\) at the few percent level to increase the power of tests for the nature of dark energy. Despite improvements in semianalytic and numerical modeling of galaxy formation, it is not clear that these methods predict galaxy bias robustly enough for this kind of precision cosmology. Faced with these challenges, most cosmological applications of the 2dFGRS and the SDSS have focused on the linear or near-linear regime, where generic arguments suggest that the effects of galaxy bias should be relatively simple. These “perturbative” analyses of large-scale structure play a significant role in the current web of cosmological constraints (e.g., Percival et al. 2002; Spergel et al. 2003; Tegmark et al. 2004b, 2006; Cole et al. 2005; Seljak et al. 2005a), but they are restricted to large scales where even these enormous surveys have limited statistical precision.
In this paper we argue that recent developments in the theoretical description of galaxy bias allow a more aggressive approach to inferring cosmological constraints from galaxy clustering measurements. We work in the framework of the halo occupation distribution (HOD), which characterizes galaxy bias in terms of the probability distribution $P(N|M)$ that a dark matter halo of virial mass $M$ contains $N$ galaxies of a specified type, together with prescriptions for the spatial and velocity bias of galaxies within dark matter halos (Ma & Fry 2000; Peacock & Smith 2000; Seljak 2000; Scoccimarro et al. 2001; Berlind & Weinberg 2002; Cooray & Sheth 2002). Here the term “halo” refers to a bound dark matter structure of typical overdensity $\delta \sim 200$, in approximate dynamical equilibrium, which may be the individual halo of a single bright galaxy or the common halo of a galaxy group or cluster.\footnote{We have in mind the kinds of structures identified in $N$-body simulations by a friends-of-friends algorithm with linking length $\sim 0.15h^{-1}$-0.25$h^{-1}$ Mpc, but the precise definition of halo does not matter provided that one is consistent throughout all calculations.} The flow diagram in Figure 1, adapted from Weinberg (2002), sketches the interplay between the “cosmological model” and the “physics of galaxy formation” in determining observable galaxy clustering, which we take to include both the traditional statistics measured from redshift surveys and the galaxy-matter correlations measured by galaxy-galaxy lensing (e.g., Fischer et al. 2000; Hoekstra et al. 2001; Sheldon et al. 2004). On the left side, the cosmological model, which specifies the initial conditions and the energy and matter contents of the universe, determines the mass function, spatial correlations, and velocity correlations of the dark halo population. The intervening box indicates that the only features of the cosmological model that really matter in this context are $\Omega_m$, the amplitude and shape of the linear matter power spectrum $P(k)$, here represented by $\sigma_8$ (the rms linear matter fluctuation in 8 $h^{-1}$ Mpc spheres), the inflationary spectral index $n_s$, and the transfer function shape parameter $\Gamma$, which itself depends on the values of $\Omega_m$, $h$, and the baryon density (Bardeen et al. 1986; Hu & Sugiyama 1996). Other features of the cosmological model, such as the energy density and equation of state of the vacuum component, may have an important impact on other cosmological observables or on the history of matter clustering, but they have virtually no effect on the halo population at $z = 0$, if the shape of $P(k)$ and the present-day value of $\sigma_8$ are held fixed (see Zheng et al. 2002).

On the right side of Figure 1, the box titled “galaxy formation physics” represents the additional processes, such as shock heating, radiative cooling, star formation, feedback, and mergers, that are essential to producing galaxies and determining their masses, luminosities, diameters, colors, and morphologies. These physical processes operate in the background provided by the evolving halo population, so together with the cosmological model they determine the HODs of different classes of galaxies. The halo population and the HOD together determine galaxy clustering and galaxy-mass correlations. One can, of course, use hydrodynamic simulations or semianalytic models to predict galaxy clustering statistics directly, without computing the HOD as an intermediate step (e.g., Kauffmann et al. 1999; Benson et al. 2000; Cen & Ostriker 2000; Pearce et al. 2001; Weinberg et al. 2004). However, these predictions reflect the combination of the cosmological model and the galaxy formation theory, and unless one has complete confidence in the latter, one cannot draw secure conclusions about the former. We would therefore like to know how well cosmological parameters can be constrained without relying on a detailed theory of galaxy formation, by using the data themselves to determine the relation between galaxies and dark matter.

For this purpose, the HOD formulation of bias has two key strengths, both emphasized by Berlind & Weinberg (2002). The first is the division of labor implied by Figure 1: galaxy formation physics influences the HOD, but the properties of the halo population defined at $\rho/\rho_c \sim 200$ are determined almost entirely by the much simpler physics of gravitational clustering, which can be modeled accurately using $N$-body simulations or numerically tested analytic approximations. The second is completeness: given a cosmological model and a specified HOD, one can calculate any galaxy clustering statistic on any scale, by populate $N$-body halos or by drawing on a steadily expanding array of analytic techniques. The HODs of different galaxy classes therefore encode, statistically, all the aspects of galaxy formation physics that are relevant to predictions of galaxy clustering. (We discuss an important caveat to this statement below.) HOD modeling enables one to take full advantage of galaxy clustering data, with no restriction to large scales or particular statistics.

The goal of our empirical approach is to reverse the causal arrows in Figure 1, working backward from the data to properties of the HOD and the halo population, and from there to conclusions about galaxy formation physics and fundamental aspects of cosmology. Berlind & Weinberg (2002) considered a fixed cosmological model and showed how changes to the HOD affect many of the traditional statistical measures of galaxy clustering, such as correlation functions, the group multiplicity function, and redshift-space distortions. They argued that the complementary information from different statistics and different spatial scales should permit an accurate empirical reconstruction of the HOD for a specified cosmology, yielding physically informative tests of galaxy formation theories. Zheng et al. (2002) showed that $\Omega_m$, $\sigma_8$, and the power spectrum shape have nondegenerate effects on
the halo population, and they argued that changes to the galaxy HOD could not fully mask the effects of cosmological parameter changes. Here we complete this theoretical program by showing how galaxy clustering measurements can constrain simultaneous changes to the cosmological model and the galaxy HOD. Our results add quantitative teeth to the qualitative arguments and speculations in Berlind & Weinberg (2002), Zheng et al. (2002), and Weinberg (2002).

We define our cosmological parameter space by the values of \( \Omega_m, \sigma_8, \Gamma \). We concentrate mainly on \( \Omega_m \) and \( \sigma_8 \) because \( \Gamma \) (or, more generally, the shape of the linear power spectrum) can be well constrained by combining the large-scale galaxy power spectrum with other observables like CMB anisotropy and the Lyman \( \alpha \) forest. We first compute predicted values of a number of galaxy clustering observables, assuming a central model with observationally motivated choices of cosmological parameters and the galaxy HOD. We then change one or more cosmological parameters, and we test how well this new cosmological model can reproduce the original clustering “data” given complete freedom to vary the HOD within a very flexible parameterization. To gain insight, we first explore several interesting axes within the \((\Omega_m, \sigma_8, \Gamma)\) parameter space, varying parameters individually or in physically motivated combinations. We arrive at definite numbers by assuming that each of our 30 observables, some of which represent the same clustering statistic measured at multiple scales, can be measured with 10% uncorrelated fractional uncertainty from a survey like the SDSS. This assumption seems roughly plausible, but the ultimate strength of cosmological conclusions will depend on the precision and dynamic range of the measurements. Forecasting this precision for a survey like the SDSS, including the covariance of errors among different observables, is itself a major theoretical task, which we will not undertake here. For our theoretical investigation, we further restrict ourselves to observables for which we have reasonable analytic approximations, and we suspect that our quantitative conclusions will in the end prove overly pessimistic because we must omit some observables that contain significant additional information.

As discussed by Berlind & Weinberg (2002), the completeness of the HOD as a description of bias rests on the assumption that the galaxy content of a halo of virial mass \( M \) is statistically independent of the halo’s larger scale environment. Berlind et al. (2003) show that this assumption accurately describes the galaxy population in the Weinberg et al. (2004) hydrodynamic cosmological simulation. It has fundamental theoretical roots in the excursion set model of Bond et al. (1991), which predicts that the statistical features of a halo’s assembly history depend only on its present mass; indeed, all semianalytic galaxy formation models that use statistically generated merger trees make this assumption implicitly. Since the halo mass function itself varies with environment, with high-mass halos absent in low-density regions, models that tie galaxy populations to halo masses still predict strong correlations between galaxy properties and the large-scale environment, in good agreement with observations (e.g., Benson 2001; Berlind et al. 2005; Zehavi et al. 2005). However, while Lemson & Kauffmann (1999) show that N-body halos of mass \( M \gtrsim 10^{13} M_\odot \) have similar properties and formation histories in different environments, recent studies show a substantial correlation of halo formation redshift with large-scale overdensity for lower mass halos (\( M \lesssim 10^{12.5} M_\odot \)), contradicting the simplest form of the Bond et al. (1991) model (Gao et al. 2005; Harker et al. 2006; see also Sheth & Tormen 2004). As discussed below in \( \S \) 5, we expect the impact of such dependence on the clustering statistics of mass- or luminosity-thresholded galaxy samples to be small, but given the level of precision we eventually hope to reach, it will probably be necessary to allow for it when fitting the data. Strategies for incorporating environmental variations into HOD modeling will require guidance from the next generation of hydrodynamic and semianalytic models of galaxy clustering, and we reserve this task for future investigation. Ultimately, the assumption of an environment-independent HOD, or of any particular parameterization of environmental dependence, must be tested empirically by checking that the adopted model can consistently explain the variations of galaxy clustering with large-scale environment.

Many current efforts to constrain cosmological parameters with galaxy clustering data make (explicitly or implicitly) the much stronger assumption that galaxy bias can be adequately described by a linear model, \( \delta_g = b \delta_m \), on the relevant scales. Here \( \delta_g \) and \( \delta_m \) are the galaxy and dark matter density contrast, respectively, and the linear bias factor \( b \) may depend on galaxy type but is assumed to be independent of scale. In this case, \( P_g(k) = b^2 P_m(k) \), and the shape of the galaxy power spectrum provides powerful cosmological constraints in combination with CMB data even if \( b \) is unknown (Percival et al. 2002; Tegmark et al. 2004b, 2006). Large-scale redshift-space distortions measure \( \beta \equiv \Omega_{m,b}^{0.8}/\Omega_{m} \) (Kaiser 1987; Hamilton 1998; Hawkins et al. 2003), which in combination with clustering measurements constrains \( \sigma_8 \Omega_{m,b}^{0.8} = \Omega_{m,b}^{0.9} \sigma_{8,b} / b \), where \( \sigma_{8,b} \) is the measured rms fluctuation of the galaxy density contrast at 8 \( h^{-1} \) Mpc. The mass function or X-ray temperature function of galaxy clusters constrains a similar combination of \( \sigma_8 \) and \( \Omega_m \) (White et al. 1993) by a route that is independent of galaxy bias. At low redshift, the combination of galaxy-galaxy lensing with the galaxy-galaxy correlation function constrains somewhat different combination of these parameters, \( \Omega_m / b \propto \sigma_8 / \Omega_m \), assuming a linear bias model with \( \xi_{gg}(r) = b \xi_{mm}(r) = b^2 \xi_{mm}(r) \). The triangle shape dependence of the reduced galaxy bispectrum can constrain \( b \) directly (Fry 1994; Verde et al. 2002), in this case assuming a quadratic bias model \( \delta_g = b b_m + b_2 \delta_m^2 + const \) to relate the galaxy and mass density fields at second order (Fry 1994; Juszkiewicz et al. 1995).

General theoretical arguments suggest that linear bias should be a good approximation for the power spectrum on sufficiently large scales, provided that the efficiency of galaxy formation is determined by the local \( r < \text{a few megaparsecs} \) environment (Coles 1993; Fry & Gaztañaga 1993; Weinberg 1995; Mann et al. 1998; Scherrer & Weinberg 1998; Narayan et al. 2000). Numerical experiments using such “local” bias models also show that the “\( b \)” multiplying the power spectrum amplitude should be similar to the “\( b \)” affecting redshift-space distortions (Berlind et al. 2001). Thus, the combination of these perturbative galaxy clustering analyses can in principle yield separate constraints on \( \Gamma \) (measured directly from the power spectrum shape), \( \Omega_m \), and \( \sigma_8 \), with the degeneracy of \( \Omega_m \) and \( \sigma_8 \) broken either by the bispectrum analysis or by the different \( \Omega_m \) dependence of the galaxy-galaxy lensing constraint and the \( \beta \) or cluster normalization constraints. The applications of these techniques to 2dFGRS and SDSS data show an impressive degree of internal consistency and good agreement with external cosmological constraints (Percival et al. 2002; Verde et al. 2002; Spergel et al. 2003; Tegmark et al. 2004b, 2006; Cole et al. 2005; Seljak et al. 2005a). However, reliance on the linear or quadratic bias approximations restricts these analyses to large scales, and it is not clear that these approximations hold consistently at the level of accuracy desired for further improvements (better than 10%, say), since scatter in the relation between galaxy and mass densities can have different effects on different statistics and at different scales (Pen 1998; Dekel & Lahav 1999).
The HOD approach to galaxy clustering analysis substitutes a much more general model of galaxy bias, and it makes use of high-precision measurements from small and intermediate scales (from sub-Mpc to \( \sim 20\ h^{-1}\) Mpc) in addition to the lower precision measurements in the perturbative regime. Furthermore, regardless of the strength of the cosmological constraints, the HOD parameters derived from the data themselves provide detailed tests of galaxy formation models.

The program of HOD-based interpretation of observed galaxy clustering is, in fact, well underway. Jing & Börner (1998) and Jing et al. (2002) used HOD-type bias models to interpret the correlation functions and pairwise velocity dispersions measured from the LCRS and the Point Source Catalog Redshift Survey (PSCz; Saunders et al. 2000). Peacock & Smith (2000), Marinoni & Hudson (2002), and Kochanek et al. (2003) used the group multiplicity function to constrain the galaxy occupations of high-mass halos. Guzik & Seljak (2002), Seljak et al. (2005a), and Mandelbaum et al. (2006) applied HOD modeling to galaxy-galaxy lensing measurements from the SDSS. Zehavi et al. (2004), analyzing a volume-limited sample of bright (\( M_r < -21\) SDSS galaxies, showed that HOD models naturally explain the observed deviation from a power-law correlation function. Zehavi et al. (2005) measured the luminosity and color dependence of the SDSS galaxy correlation function and used it to infer the luminosity and color dependence of the galaxy HOD, finding results in good qualitative agreement with theoretical predictions (Zheng et al. 2005). Magliocchetti & Porciani (2003) used a similar approach to investigate the halo occupations of early- and late-type galaxies in the 2dFGRS, while Collister & Lahav (2005) derived HODs of red and blue 2dFGRS galaxies from the group catalog of Eke et al. (2004), testing the consistency of their result by comparing predicted and observed correlation functions. Abazajian et al. (2005) used the SDSS measurements for \( M_r < -21\) galaxies in conjunction with CMB anisotropy data to infer simultaneous constraints on HOD and cosmological parameters. Tinker et al. (2005) used HOD modeling of the Zehavi et al. (2005) clustering measurements to predict cluster mass-to-light (M/L) ratios, and they inferred constraints on \( \sigma_8\Omega_m^{0.6} \) by comparing to published M/L measurements. HOD models have been applied to the interpretation of high-redshift clustering by Bullock et al. (2002), Moustakas & Somerville (2002), Yan et al. (2003), Zheng (2004), Ouchi et al. (2005), Lee et al. (2006), Coil et al. (2006), and Cooray (2006). Finally, van den Bosch et al. (2003a) have initiated a comprehensive program similar to the one described here, based on the closely related conditional luminosity function (CLF) formalism (see also van den Bosch et al. 2003b, 2004, 2005, 2006; Yang et al. 2003, 2004, 2005; Mo et al. 2004; Wang et al. 2004). We discuss the similarities and differences between the HOD and CLF approaches in § 6.

In the next section we define our class of cosmological models, list the analytic approximations we use for properties of the halo population, and describe our flexible parameterization of the galaxy HOD. In § 3 we describe the methods and approximations that we adopt for analytic calculation of galaxy clustering observables. Section 4 presents our main results, showing the ability of complementary galaxy clustering measurements to constrain the galaxy HOD and cosmological parameters simultaneously. Section 5 discusses the issue of environmental variations of the HOD. Section 6 summarizes our findings and discusses the overall prospects for application of this approach. A reader familiar with the field who wants an overview of our main results can read the first paragraph of § 2.1, skim § 2.2 to understand our HOD parameterization, then skip to § 4, paying particular attention to §§ 4.2, 4.5, and 4.6 and Figures 5, 14, and 15.

---

2. COSMOLOGICAL MODEL AND HOD PARAMETERIZATION

2.1. Cosmological Parameters and Halo Properties

Throughout this paper we adopt spatially flat \( \Lambda \)CDM cosmological models with Gaussian initial density fluctuations. The cosmological model is defined by the mass density parameter \( \Omega_m \), the mass fluctuation power spectrum \( P(k) \propto k^n T^2(k; \Gamma) \), where \( k \) is the wavenumber, \( n_s \) is the spectral index of the inflationary power spectrum, and \( T(k; \Gamma) \) is the transfer function with shape parameter \( \Gamma \). We parameterize each cosmological model in this paper by \( \Omega_m, \sigma_8, n_s \), and \( \Gamma \), where \( \sigma_8 \) is the rms fluctuation of the linear density field filtered with a top-hat filter of radius \( 8\ h^{-1}\) Mpc. We adopt the parameterization of Efstathiou et al. (1992) for the transfer function \( T(k; \Gamma) \), which approximates the evolution of adiabatic primordial fluctuations. This formulation suffices for our purpose of investigating how galaxy clustering data constrain HOD and cosmological parameters, with a single parameter \( \Gamma \) encoding the combined effects of the matter-radiation transition (at the scale \( \lambda \propto \Omega_m h \)) and the suppression of fluctuation growth in baryon or massive neutrino components. Observational analyses should use a more accurate transfer function, e.g., from CMBFAST (Seljak & Zaldarriaga 1996), but this would not change the sensitivity to the shape and amplitude of \( P(k) \). It is important to note that we treat \( \Omega_m \) and \( \Gamma \) as independently variable quantities; we do not automatically change the power spectrum shape when we change \( \Omega_m \). In the context of \( \Lambda \)CDM, our approach implicitly assumes that a change in \( \Omega_m \) is compensated by a change in \( h \) or some other parameter to keep the power spectrum at its empirically constrained shape.

Our calculations of galaxy clustering statistics rely on analytic descriptions of the density profiles, mass function, spatial clustering, and velocity statistics of dark matter halos. We draw on the extensive literature that presents numerically tested analytic approximations or numerically calibrated fitting formulae for these quantities.

The density profile of a dark matter halo of mass \( M \) is assumed to have the Navarro-Frenk-White (NFW) form (Navarro et al. 1995, 1996, 1997),

\[
\rho_n(r, M) = \frac{\rho_s}{(r/r_s)(1 + r/r_s)^2},
\]

where the characteristic radius \( r_s \) is related to the virial radius \( R_{\text{vir}} \) of the halo through the concentration parameter \( c = R_{\text{vir}}/r_s \), and

\[
\rho_s = \frac{\delta_{\text{vir}}}{3} \frac{c^3}{\ln(1 + c) - c/(1 + c)}.
\]

We define the virial radius by the condition \( \delta_{\text{vir}} = 200 \), independent of cosmology, where \( \delta_{\text{vir}} \) is the average mass density of the halo within \( R_{\text{vir}} \) in units of the mean matter density \( \rho_0 \). The concentration parameter \( c \) is a function of halo mass. Here we adopt the relation given by Bullock et al. (2001) after modifying it according to our constant-\( \delta_{\text{vir}} \) halo definition,

\[
c(M) = c_0 \left( \frac{M}{M_*} \right)^{\beta_c},
\]

where \( c_0 = 11, \beta_c = -0.13 \), and \( M_* \) is the nonlinear mass. The mass function of halos can be expressed as

\[
\frac{dn(M)}{dM} \frac{dM}{dM} = \frac{\rho_0}{M f(u)} \frac{d\ln \sigma(M)}{dM},
\]
where \(n(M)\) is the space density of halos with mass higher than \(M\), \(\sigma(M)\) is the rms fluctuation of the mass overdensity at a mass scale \(M\), \(\nu = \delta_c(\sigma(M))\) by \(\delta_c \approx 1.686\) is the threshold density contrast for collapse, and \(f(\nu)\) is a dimensionless function. The mass \(M\), in equation (3) is defined by the condition \(\sigma(M) = \delta_c\) (i.e., \(\nu = 1\)). The function \(f(\nu)\) can be approximately derived using the Press-Schechter (Press & Schechter 1974) or excursion set (Bond et al. 1991) formalism, or it can be calibrated by fitting the results of \(N\)-body simulations (e.g., Sheth & Tormen 1999; Jenkins et al. 2001). Here we use the fitting formula of Sheth & Tormen (1999)

\[
f(\nu) = A \sqrt{\frac{2a}{\pi}} \left[ 1 + (av^2)^{-c} \right] \exp \left( -\frac{av^2}{2} \right),
\]

where \(A = 0.3222\), \(a = 0.707\), and \(\rho = 0.3\). This formula is consistent with the ellipsoidal collapse model (Sheth et al. 2001), and it provides an excellent match to results from simulations (see, e.g., Reed et al. 2003).

At large scales, the halo two-point correlation function \(\xi_{ab}(r)\) is biased relative to the matter two-point correlation function \(\xi_{mm}(r)\) by a mass-dependent factor \(b^2(M)\). We adopt the Sheth et al. (2001) formula for \(b_b(M)\), based on the ellipsoidal collapse model,

\[
b_b(M) = 1 + \frac{1}{\sqrt{\delta_c}} \left[ \sqrt{a(\nu^2)} + \sqrt{ab(\nu^2)^{1-c}} - \frac{(\nu^2)^c}{(\nu^2) + b(1-c)(1-c/2)} \right],
\]

where \(a = 0.707\), \(b = 0.5\), \(c = 0.6\), and \(\nu\) and \(\delta_c\) have the same meaning as in equation (4). Our calculation of the galaxy two-point correlation function, discussed in §3.1, also incorporates scale dependence of halo bias. Recent numerical work (Seljak & Warren 2004; Tinker et al. 2005) suggests that the bias factors given by equation (6) are systematically too high. Since we use the same approximations to calculate the “observations” of our central model and the “predictions” of others, we do not expect moderate errors in \(b_b(M)\) to alter our conclusions. However, comparisons to real observational data should use the most accurate available bias factors.

The rms three-dimensional (3D) space velocity of halos of mass \(M\) can be approximated as

\[
\sigma_b(M) = H_0 \Omega_m^{0.6} \sigma_{-1} \sqrt{1 - \sigma_0^4/\sigma_1^2},
\]

where

\[
\sigma^2 = \frac{1}{2\pi^2} \int_0^\infty dk k^{2+2/3} P(k) W^2[kR(M)],
\]

\(W[kR(M)]\) is the Fourier transform of the top-hat filter on mass scale \(M\), and the smoothing radius \(R\) is determined through \(M = 4\pi \rho_0 R^3/3\) (Sheth & Diaferio 2001). The one-dimensional (1D) rms velocity difference of halos of masses \(M_1\) and \(M_2\) at a separation \(r\) is

\[
\sigma^2_{1,1D}(M_1, M_2|r) = \sigma^2_{h,1D}(M_1) + \sigma^2_{h,1D}(M_2) + \Psi(M_1, M_2|r),
\]

where \(\Psi(M_1, M_2|r)\) represents the correlation term. In our calculations (see §3.2), we ignore the correlation term, which is an acceptable simplification (Sheth & Diaferio 2001). We assume that the 1D velocity dispersion of dark matter within a halo of mass \(M\) has the value \((GM/2R_{vir})^{1/2}\) expected for a singular isothermal sphere, independent of radius. Moderate departures from this assumption can be accommodated by the velocity bias parameter \(\alpha_v\), described below.

### 2.2. HOD Parameterization

The HOD of a specified class of galaxies is defined by the probability distribution \(P(N|M)\) that a halo of mass \(M\) hosts \(N\) galaxies, together with prescriptions for the spatial and velocity distributions of galaxies within halos relative to those of the dark matter particles. For our purpose in this paper, we want a flexible parameterization of the HOD that could describe the relevant features of any physically reasonable model. We want the constraints on the HOD to be imposed by the clustering measurements rather than by our parameterization, and we want to give each cosmological model its best shot at reproducing these measurements before rejecting it.

Conceptually, it is useful to separate \(P(N|M)\) into a mean occupation function \(\langle N(M) \rangle\), defined by \(\langle N(M) \rangle = \sum_N NP(N|M)\), and a distribution \(P(N|\langle N \rangle)\) at fixed halo mass. There have been numerous investigations of the predictions of semianalytic models, hydrodynamic simulations, and high-resolution \(N\)-body simulations for \(P(N|M)\) (e.g., Kauffmann et al. 1997; Benson et al. 2000; Seljak 2000; Scoccimarro et al. 2001; Sheth & Diaferio 2001; White et al. 2001; Yoshikawa et al. 2001; Cooray & Sheth 2002; Guzik & Seljak 2002; Scranton 2003; Berlind et al. 2003; Kravtsov et al. 2004; Zheng et al. 2005; the last three are the most extensive and the most relevant to our considerations here). The predicted HOD depends on the defining characteristics of the galaxy sample, such as luminosity, color, and morphological type. For the most sensitive cosmological investigations, it seems best to work with volume-limited samples defined by luminosity thresholds, which are the largest homogeneous samples (i.e., with the same distribution of galaxy types at all distances) that one can construct from an apparent magnitude–limited survey. In this case, \(\langle N(M) \rangle\) should be a more or less monotonic function of halo mass, since more massive halos contain, on average, either more galaxies or brighter galaxies. Below some cutoff mass, \(\langle N(M) \rangle\) should drop to zero, as low-mass halos do not contain enough cooled baryons to make a galaxy above the luminosity threshold. Instead of a luminosity threshold, one can impose a threshold in stellar mass, estimated for each galaxy from its luminosity and its color or spectral energy distribution (e.g., Bell & de Jong 2001; Kauffmann et al. 2003). This approach should reduce the scatter between the galaxy observable and the host halo mass by reducing the impact of stellar population variations, thus sharpening the cutoff in \(\langle N(M) \rangle\) and minimizing its environmental dependence. However, if one starts from an apparent magnitude–limited survey, then the mass-thresholded samples one can create are smaller than the luminosity-thresholded samples because of the range of galaxy colors.

While we do not want to impose strong theoretical priors on our HOD parameterization, we have decided to incorporate one of the most generic and useful results that comes from the galaxy formation papers cited above: the distinction between central and satellite galaxies. In hydrodynamic simulations and semianalytic models, most halos contain a galaxy near the center of mass, moving close to the center-of-mass velocity, which is usually more massive and older than any other galaxies in the halo (Berlind et al. 2003). Kravtsov et al. (2004) show that the central-satellite distinction naturally explains one of the quantitatively important properties of the HOD, the sub-Poisson width of \(P(N|\langle N \rangle)\) at small mean occupation numbers. For satellite dark
matter subhalos, the fluctuations about the mean occupation are close to Poisson, but the number of central galaxies is by definition either zero or one, thus obeying nearest-integer (Bernoulli) statistics. Fluctuations in the full $P(N|\langle N \rangle)$ are substantially sub-Poisson as long as the central galaxy makes a significant contribution to $\langle N(M) \rangle$. Zheng et al. (2005) show that this argument carries over to galaxies in hydrodynamic simulations and semianalytic models.

Motivated by these results, we model the mean occupation function for galaxies above a luminosity threshold as

$$\langle N(M) \rangle = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{\log M - \log M_{\text{min}}}{\sigma_{\log M}} \right) \right] [1 + S(M)],$$

(10)

where the last set of square brackets represents the sum of one central galaxy and $S(M)$ satellites and the first provides a smooth cutoff near $M_{\text{min}}$. The form of the cutoff profile corresponds to a lognormal distribution of central galaxy luminosity at fixed halo mass (see Zheng et al. 2005). We apply the same cutoff profile to satellite galaxy numbers, although in practice these are already well below unity close to $M_{\text{min}}$. To model $S(M)$ in a flexible way, we take the free parameters to be the values of $\log S(M)$ at $m$ fixed values of $\log M$ and define the continuous function by a cubic spline that passes through these $m$ values. We set $m = 5$ and choose values of $\log S(M)$ at $(\log M/h^{-1} M_{\odot}) = 11, 12, 13, 14,$ and 15 for the spline fit. Halos more massive than $10^{15} h^{-1} M_{\odot}$ are too rare to have much impact on the clustering statistics we consider here. Theoretical predictions of $S(M)$ can be accurately described by a truncated power law parameterized by a slope, amplitude, and low-mass cutoff (Kravtsov et al. 2004; Zheng et al. 2005), but our five-parameter model allows for curvature or inflections that depart from these theoretical predictions. In addition to the $S(M)$ values, the adjustable parameters of equation (10) are $\sigma_{\log M}$ and $M_{\text{min}}$. We always assume that the uncertainty in the galaxy mean density is negligible in comparison to the uncertainty in the clustering measurements. Therefore, once we have specified $S(M)$ and $\sigma_{\log M}$, we choose the value of $M_{\text{min}}$ that yields the correct mean number density, rather than treating it as a free parameter. (We could pick any of the parameters for this treatment, but the mean density is most sensitive to $M_{\text{min}}$.)

For the clustering measures and analytic approximations used in this paper, the only property of $P(N|\langle N \rangle)$ that we need to specify is the second factorial moment $\langle N(M) \rangle [\langle N(M) \rangle - 1]$, which determines the mean number of pairs per halo. The number of central galaxies is zero or one with a relative probability determined by the mean $\langle N_{\text{cen}}(M) \rangle$. For satellite galaxies, we introduce a width parameter $\omega$ defined by

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

(11)

The value $\omega = 1$ corresponds to a Poisson distribution of satellite numbers $N_{\text{sat}}(M)$ at fixed halo mass, consistent with theoretical predictions. The parameter $\omega$ allows the satellite probability to be broader or narrower than Poisson. The full probability distribution becomes narrow at low $\langle N(M) \rangle$ in any case because of central galaxies. The value of $\omega$ determines how quickly the distribution approaches a Poisson-like width, thus influencing the number of galaxy pairs in halos with mean occupation numbers $\sim 1$ to several.

In the hydrodynamic simulation examined by Berlind et al. (2003) satellite galaxies have a radial profile similar to that of the dark matter. However, the physical processes that determine the satellite profile are complex (see, e.g., Nagai & Kravtsov 2005; Zentner et al. 2005), so there is no reason to expect the profiles to be identical. Here we assume that satellite galaxies in a halo still follow an NFW profile, but we allow the galaxy concentration parameter $c_s$ to differ from the underlying dark matter concentration parameter $c$. We characterize the relative spatial distributions of satellite galaxies and dark matter by $\Delta \log c_s \equiv \log c_s - \log c_0$, where a positive (negative) $\Delta \log c_s$ means that galaxies are more (less) centrally concentrated than dark matter. We assume that central galaxies always reside at the halo center of mass.

We also allow the velocity dispersion of satellites to differ from the velocity dispersion of dark matter by an average factor $\alpha_v$, which we assume to be independent of mass. Central galaxies are assumed to move at the halo center-of-mass velocity. Simulations suggest that $\alpha_v \approx 1$, but the results are not entirely consistent on this point (e.g., Berlind et al. 2003; Kravtsov et al. 2004), and in any event we want this dynamical question to be settled by observations rather than by theory. Tinker et al. (2006) discuss the influence of $\alpha_v$ on redshift-space distortions in detail.

Altogether we have 10 free parameters in our HOD prescription: seven in $\langle N(M) \rangle$ ($M_{\text{min}}, \sigma_{\log M}$, five spline points), one ($\omega$) for the second moment of $P(N|\langle N_{\text{sat}}(M) \rangle)$, one ($\Delta \log c_s$) for the spatial bias within halos, and one ($\alpha_v$) for the velocity bias within halos. One of these (which in practice we take to be $M_{\text{min}}$) is determined by matching the mean galaxy space density once the values of other parameters are specified. The remaining nine are varied to yield the best match to the observables described in §3. The parameters $\omega$, $\Delta \log c_s$, and $\alpha_v$ could in principle depend on halo mass, but the variation would have to be quite strong over a fairly narrow mass range to have an effect that could not be adequately described by an appropriate average value.

The main theoretical preconception that is built into our parameterization is the existence of central galaxies. The central location of these galaxies makes a small but not negligible difference to the two-point correlation function (see Berlind & Weinberg 2002), although this could probably be compensated by $\Delta \log c_0$. However, since $\langle N_{\text{sat}}(M) \rangle$ in practice turns out to be small near the cutoff mass $M_{\text{min}}$, the central galaxy parameterization necessarily leads to sub-Poisson $P(N_{\text{cen}}(M))$ fluctuations at low $\langle N(M) \rangle$. Since this aspect of galaxy formation physics is well grounded, we think it is reasonable to adopt it, and it allows us to represent a wider range of physically realistic HODs with our finite parameter set. However, even if we do not make the central-satellite distinction and instead use a functional form like that of Scranton (2003) to describe the width of $P(N_{\text{cen}}(M))$ as a function of halo mass, we reach similar conclusions about the ability to break degeneracies between cosmology and galaxy bias. In essence, the observations drive us to sub-Poisson fluctuations at low halo masses whether or not we impose this trend as a theoretical expectation (see Benson et al. 2000; Peacock & Smith 2000; Berlind & Weinberg 2002).

3. ANALYTIC CALCULATION OF OBSERVABLE QUANTITIES

The observable quantities considered in this paper fall into three general categories. First are quantities related to the real-space clustering of galaxies: the rms galaxy number density contrast, the galaxy two-point correlation function, the reduced galaxy bispectrum, and the group multiplicity function. Second are dynamical quantities that depend on galaxy peculiar velocities: the large-scale redshift-space distortion parameter $\beta$, the galaxy pairwise velocity dispersion, and the estimated virial mass of galaxy groups as a function of group richness. Finally, we include the galaxy-matter two-point cross-correlation function, which can be
measured by galaxy-galaxy lensing. Most of these quantities are continuous functions of scale or group multiplicity. To obtain a discrete set of observables, we sample these functions at a discrete set of separations or multiplicities, listed in §4. We space our samplings widely so that we can treat the statistical errors in each measurement as approximately uncorrelated.

In addition to these observables, we impose the requirement that the HOD model reproduce the correct value of the sample’s mean space density, implying

$$\bar{n}_g = \int_0^\infty dM \frac{dn}{dM} \langle N(M) \rangle. \quad (12)$$

In effect, we treat $\bar{n}_g$ as an observable with negligible statistical error.

In this paper we limit ourselves to observables for which we have some reasonable analytic approximation. In many cases, these approximations are not accurate enough for application to high-precision clustering measurements from large surveys like the SDSS or 2dFGRS, or at least their accuracy has not been extensively tested with numerical simulations. However, we use the same approximations to calculate observed quantities for our central cosmological model and HOD and predicted quantities for altered models. Provided that our approximations realistically capture the variations of the observables with variations in the cosmological and HOD parameters, they are adequate for our present purposes of forecasting the ability of clustering measurements to break the degeneracy between cosmology and galaxy bias. Observational applications of these techniques will require further testing and development of analytic approximations. They could also benefit from analytic approximations for other statistics, or from numerical studies that provide accurate interpolation formulae in the relevant ranges of cosmological and HOD parameters.

3.1. Spatial Clustering

At large scales, the galaxy two-point correlation function $\xi_{gg}(r)$ is biased relative to the matter correlation function by (the square of) a number-weighted average of the halo bias factor,

$$b_g = \frac{1}{\bar{n}_g} \int_0^\infty dM \frac{dn}{dM} \langle N(M) \rangle b_h(M). \quad (13)$$

We use equations (4) and (5) for $dn/dM$ and equation (6) for $b_h(M)$. The fractional rms fluctuation $\sigma_g(R)$ of galaxy number counts in spheres of radius $R$ can be expressed as an integral over $\xi_{gg}(r)$, so at large $R$ it is related to the rms mass fluctuation $\sigma(R)$ by

$$\sigma_g(R) = b_g \sigma(R). \quad (14)$$

At $R = 8 h^{-1}$ Mpc, the smaller of the two scales we use for $\sigma_g(R)$, the approximation given by equation (14) is good but not perfect (see Tinker et al. 2006).

More generally, $\xi_{gg}(r)$ can be decomposed into one-halo and two-halo terms, $\xi_{1h}^{1h}(r)$ and $\xi_{2h}^{2h}(r)$, which respectively represent pairs of galaxies residing in the same halo and in separate halos. The one-halo term dominates at small separations and the two-halo term at large separations, with the transition occurring near the virial diameter of large halos. Noting that the total number of galaxy pairs [$\times 1 + \xi_{gg}(r)$] is simply the sum of the number of pairs from single halos [$\times 1 + \xi_{1h}(r)$] and that from different halos [$\times 1 + \xi_{2h}(r)$], we have $\xi_{gg}(r) = [1 + \xi_{1h}^{1h}(r)] + \xi_{2h}^{2h}(r)$.

The one-halo term $\xi_{gg}^{1h}(r)$ can be exactly computed in real space through (Berlind & Weinberg 2002)

$$1 + \xi_{gg}^{1h}(r) = \frac{1}{2\pi r^2 \bar{n}_g^2} \int_0^\infty dM \frac{dn}{dM} \left( \frac{N(M)[N(M) - 1]}{2} \right) \times \frac{1}{2R_{vir}(M)} F^r \left( \frac{r}{2R_{vir}} \right), \quad (15)$$

where $\langle N(M)[N(M) - 1]/2 \rangle$ is the average number of pairs in a halo of mass $M$ and $F^r(r/2R_{vir})$ is the cumulative radial distribution of galaxy pairs, i.e., the average fraction of galaxy pairs in a halo of mass $M$ (virial radius $R_{vir}$) that have separation less than $r$. The differential function $F^r(r)$ is determined by the profile of the galaxy distribution within the halo. We assume that there is always a galaxy located at the center of the halo, and others are regarded as satellite galaxies. With this assumption, $F^r(r)$ is the pair number–weighted average of the central-satellite pair distribution $F^r_{cs}(x)$ and the satellite-satellite pair distribution $F^r_{ss}(x)$ (see, e.g., Berlind & Weinberg 2002; Yang et al. 2003),

$$\frac{\langle N(M)[N(M) - 1]/2 \rangle}{2} F^r(x) = \langle N_{sat}(M) \rangle F^r_{cs}(x) + \frac{\langle N_{sat}(M) \rangle[N_{sat}(M) - 1]/2}{2} F^r_{ss}(x). \quad (16)$$

The central-satellite galaxy pair distribution $F^r_{cs}(x)$ is just the normalized radial distribution of galaxies within halos, and the satellite-satellite galaxy pair distribution $F^r_{ss}(x)$ can be derived through the convolution of the galaxy distribution profile with itself (Sheth & Diaferio 2001). In the case of an NFW profile, as used in this paper, $F^r_{ss}(x)$ can be analytically expressed (see Appendix A; also see Sheth & Diaferio 2001).

At large scales, the two-halo term becomes $\xi_{gg}^{2h}(r) = b_g^2 \xi_{mm}(r)$, where $b_g$ is the galaxy number–weighted halo bias factor of equation (13). At smaller scales, one must account for the convolution of galaxy profiles over the finite size of halos, scale dependence of the halo bias factor, and the fact that separate halos by definition do not interpenetrate. Because convolutions become multiplications in Fourier space, it is convenient to calculate the two-halo component of the galaxy power spectrum $P_{gg}^{2h}(k)$ and Fourier transform it to get $\xi_{gg}^{2h}(r)$ (Seljak 2000; Scoccimarro et al. 2001). Our full method of calculating the two-halo term is described in Appendix B of Tinker et al. (2005). Our adopted treatment of halo exclusion is referred to there as the “spherical” exclusion method, and we use the Sheth et al. (2001) formula for $b_g(M)$ instead of the modified formula proposed by Tinker et al. (2005). We have used this method in earlier papers (Zehavi et al. 2004, 2005; Zheng 2004).

Galaxies within the same dark matter halo constitute a galaxy group. The cumulative multiplicity function of galaxy groups, which is the number density of galaxy groups that have $N$ or more members, is

$$n_{grp}(\geq N) = \sum_{i=N}^\infty \int_0^\infty dM \frac{dn}{dM} P(i|M). \quad (17)$$

In this work we make the simplifying approximation that $P(i|M)$ is a nearest-integer distribution, where $N$ is one of the two integers bracketing $\langle N(M) \rangle$, and the relative probability is determined by having the right mean. The space density of galaxy groups with multiplicity $\geq \langle N(M) \rangle$ is approximately equal to the space density of halos with mass $M$, so given a cosmological model that specifies $n(M)$, one can effectively “read off” the high end of the
mean occupation function (Peacock & Smith 2000; Berlind & Weinberg 2002; Marconni & Hudson 2002; Kochanek et al. 2003). A full observational implementation should take into account both the width of $P(N|\langle N\rangle)$ and the scatter between estimated and true multiplicities introduced by the group finding algorithm. We consider galaxy groups with five or more members, so the approximation given by equation (17) is fairly good.

### 3.2. Dynamical Measurements

In redshift space, the peculiar velocities of galaxies distort the galaxy power spectrum and correlation function, producing anisotropy in which the line of sight is a preferred direction. By measuring this distortion on large scales, where it is caused by coherent flows into overdense regions and out from underdense regions, one can infer a combination $\beta \equiv \Omega_m^{0.6}/b_0$ of the density parameter and the large-scale galaxy bias factor (Kaiser 1987). We adopt $\beta$ as the first of our dynamically sensitive observables. The task of inferring $\beta$ from clustering measurements in the context of HOD models is discussed in detail in Tinker et al. (2006).

The next dynamical measurement we consider is the pairwise radial velocity dispersion of galaxies, defined as $\sigma_{1h}^2(r) = \langle v_1^2\rangle - \langle v_1\rangle^2$, where $v_1 = (r_1 - r_2)/(r_1 + r_2)$ is the radial (inward) velocity of a galaxy pair, $r_1$ and $r_2$ are velocities and positions of the two galaxies, and the average is over all galaxy pairs with separations near $r = r_1 - r_2$. The pairwise dispersion can be inferred by modeling the redshift-space correlation function (e.g., Bean et al. 1983; Davis & Peebles 1983).

Similar to the galaxy correlation function, on small scales, $\sigma_1^2(r)$ is dominated by a one-halo term, which can be computed through (Berlind & Weinberg 2002)

$$\sigma_{1h}^2(r) = \frac{1}{2\pi r^2 n_g^2 \xi_{gg}(r)} \int_0^{\infty} dm \frac{dn}{dm} \langle N(M)|N(M) - 1]\right \rangle \frac{1}{2} \right.,$$

$$\times \frac{1}{2R_{vir}} F'(\frac{r}{2R_{vir}}) \sigma_{gg}^2(M),$$

where $\sigma_{gg}^2(M)$ should be understood as the pair number–weighted average dispersion of the relative radial velocities between pairs of galaxies in the same halo. If we assume that the velocity distribution of satellite galaxies is isotropic and isothermal and that the central galaxy is at rest with respect to the center of mass of the parent halo, then the pair-weighted average dispersion can be separated into central-satellite and satellite-satellite terms as in equation (16),

$$\frac{\langle N(M)|N(M) - 1]\right \rangle}{2} F' \sigma_{gg}^2(M) = \langle N_{sat}(M)\rangle F'_u \alpha_u^2 \sigma_{gg}^2(M)$$

$$+ \langle N_{sat}(M)|N_{sat}(M) - 1]\right \rangle \frac{2}{2} \right.,$$

$$\times F'_u 2 \alpha_u^2 \sigma_{gg}^2(M).$$

Here $\sigma_{gg}^2(M) = GM/(2R_{vir})$ is the 1D velocity dispersion of dark matter within the halo, $\alpha_u$ is the velocity bias factor, and other quantities have the same meaning as in equation (16). Note that there is a factor of 2 for satellite-satellite velocity dispersions, since both members of the pair are moving with respect to the halo center of mass.

On large scales, the two galaxies of each pair come from different halos, so the relative motions of halos also contribute to the galaxy pairwise dispersion. We ignore the halo velocity correlation ($\Psi$ in eq. [9]), which is an acceptable simplification (Sheth & Diaferio 2001). Following the argument of Berlind & Weinberg (2002) but taking into account the separation of central and satellite galaxies, the galaxy pairwise radial velocity dispersion on large scales can be expressed as

$$\sigma_{2h}^2(r) = \frac{1}{n_g^2} \int_0^{\infty} dm_1 \frac{dn_1}{dm_1} \int_0^{\infty} dm_2 \frac{dn_2}{dm_2} [1 + \xi_{hh}(r; M_1, M_2)]$$

$$\times \left[ \frac{\langle N(M_1)\rangle \langle N(M_2)\rangle \sigma_{1D}^2(r; M_1, M_2)}{\langle N_{sat}(M_1)\rangle \langle N_{sat}(M_2)\rangle \sigma_{m}^2(M_1) M_1} + \langle N(M_1)\rangle \langle N_{sat}(M_2)\rangle \alpha_u^2 \sigma_{gg}^2(M_2) \right],$$

where $\xi_{hh}(r; M_1, M_2)$ is the two-point correlation function between halos of mass $M_1$ and $M_2$ at separation $r$. The three terms inside the last set of square brackets represent contributions by the pairwise velocity dispersion of $M_1$ and $M_2$ halos, by the dispersion of satellite galaxies in $M_1$ halos, and by the dispersion of satellite galaxies in $M_2$ halos. The halo-halo velocity dispersion is computed on all $\langle N(M_1)\rangle \langle N(M_2)\rangle$ galaxy pairs. However, the velocity dispersion of satellite galaxies in, say, $M_1$ halos only contributes to the dispersion of $\langle N_{sat}(M_2)\rangle \langle N(M_2)\rangle$ pairs involving those satellites. The halo-halo velocity dispersion can be computed through equation (9), and on large scales we have $\xi_{hh}(r; M_1, M_2) = b_1(M_1)b_2(M_2)\xi_{mm}$ and $\xi_{gg}(r) = \beta^2\xi_{mm}$. With these substitutions, equation (20) reduces to

$$\sigma_{2h}^2(r) = \frac{2}{1 + \xi_{gg}(r)} \left[ \frac{\sigma_{1D}^2}{\langle \xi_{gg}(r) \rangle} \right]$$

$$+ \frac{\sigma_{m}^2}{1 + \xi_{gg}(r)} \left[ \frac{\sigma_{m}^2}{\langle \xi_{gg}(r) \rangle} \right],$$

where the number- and bias-weighted averages $\langle \xi_{gg}(r) \rangle$ and $\langle \xi_{gg}(r) \rangle$ are defined as

$$\langle f \rangle_{N} \equiv \frac{1}{n_g} \int_0^{\infty} dm \frac{dn}{dm} \langle N(M)\rangle f(M),$$

$$\langle f \rangle_{N,b} \equiv \frac{1}{n_g b_g} \int_0^{\infty} dm \frac{dn}{dm} \langle N(M)\rangle b(M) f(M),$$

and $\langle \xi_{gg}(r) \rangle$ and $\langle \xi_{gg}(r) \rangle$ are defined by replacing $\langle N(M)\rangle$ with $\langle N_{sat}(M)\rangle$ in equations (22) and (23), respectively. Since we do not have a good approximation for $\sigma_{m}^2(r)$ at intermediate scales, we only consider pairwise dispersion measurements at small scales, where equation (18) applies, or on large scales, where equation (21) applies. Ultimately, the constraints obtainable from anisotropies in redshift space on small, intermediate, and large scales can be inferred using the methods of Tinker et al. (2006) and Tinker (2007).

We also consider dynamical measurements of the average virial masses of galaxy groups of fixed multiplicity, $\langle M_{vir}(N)\rangle$. These can be inferred from galaxy positions and velocities using the kind of estimators developed by Heisler et al. (1985). These estimators are affected by velocity bias, with the estimated mass scaling as $\alpha_u^2$. For a given halo mass function and HOD, we therefore calculate this quantity as

$$\langle M_{vir}(N)\rangle = \alpha_u^2 \int_0^{\infty} dm \frac{dn}{dm} P(N|M) M$$

$$\int_0^{\infty} dm \frac{dn}{dm} P(N|M).$$

As in our multiplicity function calculation, we approximate $P(N|\langle N\rangle)$ by a nearest-integer distribution. With equation (24), we implicitly assume that the mass estimators are applied only
to satellite galaxies, or else that the reduced dispersion due to the central galaxy is properly taken into account. With large enough galaxy numbers, more sophisticated mass estimators can circumvent the effects of velocity bias (e.g., Carlborg et al. 1997). In addition, average group masses can be estimated using gravitational lensing or X-ray properties instead of galaxy velocities. We do not consider these kinds of $\alpha_m$-free mass estimates in our analysis, but it is clear that they would add power to our overall set of constraints if they could be implemented robustly.

3.3. Galaxy-Mass Clustering

The galaxy-mass cross-correlation can be probed by galaxy-galaxy lensing (e.g., Fischer et al. 2000; Hoekstra et al. 2001; Sheldon et al. 2004; Seljak et al. 2005a). In practice, what is usually inferred from galaxy-galaxy lensing at low redshift is the product of the mean matter density $\Omega_m$ and the galaxy-mass two-point cross-correlation function $\xi_{gm}(r)$, so we adopt values of $\Omega_m \xi_{gm}(r)$ as the observables in our analysis.

The galaxy-mass two-point cross-correlation function can also be decomposed into one-halo and two-halo terms. For simplicity, we only consider small-scale cross-correlation, where the one-halo term dominates. Similar to the one-halo term of the galaxy-mass two-point correlation function, the one-halo term $\xi_{gm}^{1h}(r)$ of the galaxy-mass two-point cross-correlation function is computed in real space by counting galaxy and mass particle pairs,

$$1 + \xi_{gm}^{1h}(r) = \frac{1}{4\pi r^2 n_c \Omega_m \rho_c} \times \int_0^\infty \frac{dn}{dM} \frac{dN}{dM} \frac{P_3^g(r)}{2 R_{vir}(M)} \left( \frac{r}{2 R_{vir}} \right),$$

where $\rho_c$ is the critical density. The function $F^{gm}_{\xi}(x)$ is the cumulative radial distribution of galaxy and mass particle pairs (not galaxy-galaxy pairs as in eq. [15]), which is determined by the distribution profiles of galaxies and dark matter within the halo. When the effect of the central galaxy is taken into account, we have (similar to eq. [16])

$$\langle N(M) \rangle F^{gm}_{\xi}(x) = F^{gm}_{\xi}(x) + \langle N_{sat}(M) \rangle F^{gm}_{\xi}(x),$$

where $F^{gm}_{\xi}$ and $F^{gm}_{\xi}$ are the cumulative distribution of pairs made of the central galaxy (satellite galaxies) and mass particles. The function $F^{gm}_{\xi}(x)$ is simply the radial distribution of matter, which can be directly obtained from the NFW profile. The function for the satellite-mass pairs, $F^{gm}_{\xi}(x)$, can be derived by convolving satellite and dark matter profiles. Under our assumption that the satellite distribution inside a halo also follows an NFW profile, $F^{gm}_{\xi}(x)$ is the convolution of two NFW profiles with different concentration parameters, which has an analytic expression given in Appendix A. More general discussions of galaxy-galaxy lensing in the halo occupation framework are given by Mandelbaum et al. (2005), who focus on halo virial masses and satellite galaxy fractions, and Yoo et al. (2006), who focus on the derivable constraints on $\Omega_m$ and $\sigma_8$.

3.4. The Galaxy Bispectrum

At the final step of our analysis, we consider the large-scale behavior of the galaxy bispectrum, the Fourier transform of the three-point correlation function. Although this is a measure of real-space clustering, like those in § 3.1, we treat it separately because its information content is rather different and because it is more difficult to measure observationally. The specific observable that we consider is the reduced bispectrum

$$Q(k_1, k_2, k_3) = \frac{B(k_1, k_2, k_3)}{P(k_1)P(k_2)P(k_3) + P(k_2)P(k_3)P(k_1)}.$$

where $B(k_1, k_2, k_3)$ is the bispectrum, $P(k)$ is the power spectrum, and the three wavevectors form a triangle $(k_1 + k_2 + k_3 = 0)$. In the weakly nonlinear regime, the local relation between the galaxy density contrast $\delta_g$ and the mass density contrast $\delta$ can be written as a Taylor expansion $\delta_g = \sum b_{g, \ell} \delta^K_n$. The reduced galaxy bispectrum $Q_g$ is related to the reduced matter bispectrum $Q_m$ through

$$Q_g(k_1, k_2, k_3) = \frac{Q_m(k_1, k_2, k_3)}{b_{g,1} + b_{g,2}^2},$$

which shows contributions from both gravitational clustering and galaxy bias (Fry 1994; see also Fry & Gaztanaga 1993; Juszkiewicz et al. 1995). On large scales, $b_{g,1}$ and $b_{g,2}$ are simply the galaxy number-weighted halo bias factors $b_{h,1}$ and $b_{h,2}$, respectively, where $b_{h,1}$ and $b_{h,2}$ are predicted by nonlinear perturbation theory (Scoccimarro et al. 2001). The reduced matter bispectrum is calculated using the linear matter power spectrum $P(k)$ and the second-order perturbative bispectrum $B_{m PT}^{PT}$ (Fry 1984; Scoccimarro et al. 2001),

$$B_{m PT} = 2 F_2(k_1, k_2) P(k_1) P(k_2) + \text{cyc.},$$

where $F_2(k_1, k_2) = \frac{7}{4} + \frac{7}{4} \cos \theta_{12} k_1 k_2 + \frac{7}{12} \cos^2 \theta_{12}$, with $k_1 \cdot k_2 = k_1 k_2 \cos \theta_{12}$. The weak dependence on $\Omega_m$ (Kamionkowski & Buchalter 1999) is neglected here. The key diagnostic power of $Q_g$ on large scales lies in the distinctive triangle shape dependence of $Q_g$ predicted by gravitational perturbation theory. Positive linear bias ($b_{g,1} > 1$) suppresses $Q_g$, and nonlinear bias ($b_{g,2} > 0$) can restore the amplitude but not the shape dependence (Fry 1994; Verde et al. 2002).

4. BREAKING THE DEGENERACY BETWEEN BIAS AND COSMOLOGY

Motivated by CMB anisotropy measurements (e.g., Netterfield et al. 2002; Pryke et al. 2002; Spergel et al. 2003), the abundance of galaxy clusters (e.g., White et al. 1993; Eke et al. 1996), high-redshift supernova observations (e.g., Riess et al. 1998, 2004; Perlmutter et al. 1999), and measurements of the 3D galaxy power spectrum and the Ly$\alpha$ forest power spectrum (e.g., Croft et al. 2002; Percival et al. 2002b; Tegmark et al. 2004b; Seljak et al. 2005b), we assume a central cosmology with $\Omega_m = 0.3$, $\sigma_8 = 0.9$, $n_s = 1.0$, and $\Gamma = 0.2$. We model a galaxy population of mean space density $\bar{n}_g = 0.01 h^3$ Mpc$^{-3}$, similar to that of galaxies with $r$-band absolute magnitude brighter than $-19.5 + 5 \log h$ in the SDSS. We therefore use the Zehavi et al. (2005) results for $M_r \leq -19.5 + 5 \log h$ galaxies as a guide for our central model’s HOD parameters. We assume a satellite occupation $S(M) = M/(1.4 \times 10^{13} h^{-1} M_\odot)$ in equation (10), i.e., the number of satellites is proportional to the halo mass well above the cutoff $M_{min}$. We set the central galaxy cutoff parameter $\sigma_{logM}$ to 0.2, consistent with predictions from galaxy formation models (Zheng et al. 2005). With these choices, the cutoff mass required to match $\bar{n}_g = 0.01 h^3$ Mpc$^{-3}$ is $M_{min} = 5.55 \times 10^{11} h^{-1} M_\odot$. The probability distribution of satellite galaxies is taken to be a Poisson distribution, i.e., $\omega = 1$ in equation (11). We assume that satellite
galaxies have the same radial profile and velocity dispersion as dark matter within halos, i.e., $\Delta \log c_0 = 0$ and $\alpha_v = 1$.

For the central cosmology and the central HOD, we use the analytic formulae presented in §3 to calculate a set of observables, which we thereafter treat as observational measurements. These observables are the rms galaxy overdensity $\sigma_G^2$ on scales of 8 and 15 $h^{-1}$ Mpc; the group multiplicity function $n_{grp}(N)$ for $N = 5, 10, 20, 40, 80$; the galaxy two-point correlation function $\xi_{gg}(r)$ at $r = 0.1, 0.3, 0.5, 1, 2, 5, 10$ $h^{-1}$ Mpc; the $\beta$ parameter; the galaxy pairwise velocity dispersion $\sigma_v(r)$ at $r = 0.25, 0.5, 1, 3, 5, 10$ $h^{-1}$ Mpc; the average virial mass of galaxy groups $\langle M_{grp}(N) \rangle$ for $N = 10, 20, 40, 80$; the galaxy-mass cross-correlation function $\Omega_m \xi_{mg}(r)$ at $r = 0.25$ and $0.5$ $h^{-1}$ Mpc; and values of the reduced galaxy bispectrum on large scales, for wavevector triangles having two sides fixed at $k_2 = 2k_1 = 0.05$ $h$ Mpc$^{-1}$ and subtending angles $\theta = 0, \pi/4, \pi/2, 3\pi/4,$ and $\pi$. We assume that each of the 30 “measurements” has a 1 $\sigma$ fractional uncertainty of 10% and that the measurement errors are uncorrelated. On its own, the 10% error assumption is probably pessimistic; even the largest scale observables can probably be measured at least this well with the full SDSS, and some of the smaller scale quantities have already been measured more precisely (we investigate the effect of reducing errors for selected observables in §4.5). However, the assumption that the measurement errors are uncorrelated is optimistic, even though we space the observables for any given statistic (e.g., multiple scales of the correlation function) fairly widely. Unless there are strong positive correlations among many observables, the constraints on the HOD and cosmology should not degrade much with respect to our forecasts, but a full application to observations must include the covariances among observables, calculated from the data or from mock catalogs.

Each time we change cosmological parameters, and thus the statistics of the halo population, we search for HOD parameters that yield the best possible match to the observables of the central model. Specifically, we minimize

$$\Delta \chi^2 = \sum_i \frac{(F_i^p - F_i^o)^2}{\sigma_{Fi}},$$

where $F_i$ represents the $i$th observable with an observational error $\sigma_{Fi}$, the superscript “$p$” indicates the predicted value for a given cosmology and HOD model, and the superscript “$o$” indicates the observed value calculated for the central cosmological model and HOD parameters. To isolate the information content of different clustering statistics, we start by considering only the spatial clustering observables described in §3.1, then add new observables one at a time. We use a Gauss-Newton scheme to carry out the $\chi^2$ minimization (see Appendix B for details). We characterize the acceptability of models in terms of $\Delta \chi^2$, appropriate for “flat” priors that treat all parameter values equally.

### 4.1. Constraining HOD Parameters for a Fixed Cosmology

Before turning to cosmological parameter constraints, we first ask how well the HOD can be determined if we assume that the cosmological parameters are known perfectly from independent data. Observational constraints on the HOD for an assumed cosmology have already been derived from the 2dfGRS, the SDSS, and other data sets, but these studies have all used restricted forms of the HOD and limited subsets of the galaxy clustering observables, such as the correlation function or the multiplicity function (e.g., Jing & Börner 1998; Peacock & Smith 2000; Marinoni & Hudson 2002; Kochanek et al. 2003; Lin et al. 2004; van den Bosch et al. 2003b; Magliocchetti & Porciani 2003; Zehavi et al. 2004, 2005; Collister & Lahav 2005). Here we adopt our much more flexible HOD parameterization and examine the constraints that could be obtained with 10% measurements of all the observables discussed above. Like Yan et al. (2003) and van den Bosch et al. (2005), we explore the HOD parameter space using a Monte Carlo Markov Chain (MCMC) technique (e.g., Gilks et al. 1996). We adopt the central cosmological model throughout, and we start the chain from the central HOD parameters, which by definition yield $\Delta \chi^2 = 0$. At any point of the chain, we generate a new set of HOD parameters by taking a random walk in the HOD parameter space with the step size for each parameter drawn from a Gaussian distribution. We accept the new HOD model with a probability of 1 if $\chi^2_{\text{new}} \leq \chi^2_{\text{old}}$ and $\exp [-\chi^2_{\text{new}} - \chi^2_{\text{old}}/2]$ if $\chi^2_{\text{new}} > \chi^2_{\text{old}}$, where $\chi^2_{\text{new}}$ and $\chi^2_{\text{old}}$ are values of $\chi^2$ for the new model and for the previous model, respectively. Flat priors are adopted for HOD parameters $S(M)$, $\sigma_v(M)$, and $\omega$ in logarithmic space and $\Delta \log c_0$ and $\alpha_v$ in linear space.

The constraints on $\langle N(M) \rangle$ and other HOD parameters are illustrated in Figure 2. Figure 2a plots the envelope of $\langle N(M) \rangle$ curves determined from HOD models that have $\Delta \chi^2 < 1$, thus showing the 1 $\sigma$ uncertainty in $\langle N(M) \rangle$. With our adopted set of observables, the shape of $\langle N(M) \rangle$ near the cutoff mass is poorly constrained. The mean galaxy density imposes a strong constraint on the cutoff scale, but either hard or soft cutoffs can produce the same $b_g$. The halo bias factor is nearly independent of mass in the low-mass regime (see, e.g., Jing 1998; Sheth et al. 2001; SELjak & Warren 2004), and these halos contain only a single galaxy, so variations in the cutoff profile do not affect either the two-halo or one-halo terms of $\xi_{gg}(r)$. Observables that we have not considered, like the void probability function or the Tully-Fisher relation, might provide stronger constraints on the location and form of the $\langle N(M) \rangle$ cutoff.

At higher halo masses, our observables provide much greater leverage, and $\langle N(M) \rangle$ is tightly constrained over the range $\langle N \rangle \sim 1.2 - 100$. The constraints loosen at still higher halo masses, since these halos are too rare to have much impact on the correlation function, and we only consider group virial masses up to $N = 80$ and the multiplicity function up to $N = 40$. Figures 2b–2d show the marginalized likelihoods of the parameters $\omega$, $\Delta \log c_0$, and $\alpha_v$, with dotted vertical lines marking the 1 $\sigma$ range. Under our assumption of 10% observational errors, the distribution width parameter $\omega$ can be constrained at the level of 10%–15% (1 $\sigma$). For the difference between galaxy and dark matter concentrations, the 1 $\sigma$ constraint is 0.10–0.15 dex, although with more small-scale points in $\xi_{gg}(r)$ this quantity could be pinned down more tightly. The 1 $\sigma$ uncertainty in the velocity bias $\alpha_v$ is only a few percent, largely because the spatial clustering observables tightly constrain the true halo mass at fixed $N$, and the average virial masses then determine the value of $\alpha_v$.

Figure 3 shows correlations between some pairs of HOD parameters, inferred from the MCMC run. If galaxies are more centrally concentrated than dark matter inside halos, then the preferred width of $P(N|M)$ narrows to suppress one-halo pairs in halos with small virial radii (Fig. 3a). A wider $P(N|M)$ correlates, weakly, with a lower velocity bias ($\alpha_v < 1$ for $\omega > 1$; Fig. 3b), since more small-scale pairs involve satellite galaxies, which make larger contributions to the pairwise velocity dispersion. There is a similar but stronger correlation with galaxy concentration (Fig. 3c), since positive $\Delta \log c_0$ allows more small-scale pairs to come from massive halos with large velocity dispersions. Finally, if the mean occupation at high halo masses increases (higher $S_{15}$ in Fig. 3d), then the mean halo mass at a given high
multiplicity decreases, so $\alpha_v$, greater than unity is favored to keep the apparent virial mass $M_{\text{vir}}(N)$ near its original value.

4.2. Changing $\Omega_m$ with $\sigma_8$, $n_s$, and $\Gamma$ Fixed

Before considering general constraints in the $(\Omega_m, \sigma_8, n_s, \Gamma)$ parameter space in § 4.5, we examine constraints along several axes in this space: a pure change in $\Omega_m$, a pure change in $\sigma_8$, “cluster-normalized” changes that preserve the value of $\sigma_8 \Omega^{-0.5}_m$ at fixed $n_s$ and $\Gamma$, and linked changes in $\Omega_m$, $\sigma_8$, and $\Gamma$ (at fixed $n_s$) that preserve the amplitude and slope of the halo mass function at high masses. As discussed by Zheng et al. (2002), the halo population responds to these changes in relatively simple ways, so we gain physical insight into the origin of the more general cosmological constraints by studying the behavior along these axes. These restricted constraints are also relevant to the case where other parameters are determined by independent data.

Figure 4 illustrates the variation of the halo mass function $dn/d\log M$ (Fig. 4a) and the nonlinear matter correlation function $\xi_{\text{non}}(r)$ (Fig. 4b) along these four axes. Open circles show the central model. Dotted lines show the effect of increasing $\Omega_m$ from 0.3 to 0.5, with other parameters fixed. Dashed lines show the effect of increasing $\sigma_8$ from 0.9 to 1.0. Dot-dashed lines show a cluster-normalized change, with $\Omega_m$ increased to 0.5 and $\sigma_8$ decreased to 0.7 to keep $\sigma_8 \Omega^{-0.5}_m$ fixed. Solid lines show the same $(\Omega_m, \sigma_8)$ combination, but with a corresponding change in $\Gamma$ that keeps the slope of $dn/d\log M$ fixed at the cluster mass scale $5 \times 10^{14} h^{-1} M_\odot$ (see Zheng et al. 2002). In all cases we use the Sheth & Tormen (1999) mass function formula (our eq. [4]) and the Smith et al. (2003) prescription for computing the matter correlation function. Note that peculiar velocities of halos on large scales are proportional to $\sigma_8 \Omega^{-0.5}_m$, similar to the scaling of velocity dispersions of cluster mass halos (Zheng et al. 2002).

The constraints for a pure $\Omega_m$ change are easiest to understand, so we examine this case first and in the greatest detail. As shown in Figure 4, a pure $\Omega_m$ change simply shifts the mass function in proportion to $\Omega_m$. The effect on the matter correlation function is very small, arising entirely from changes in the concentrations of halos. Once halo masses are scaled in proportion to $\Omega_m$, the spatial clustering of halos at a given scaled mass is virtually unchanged, although the velocities change in proportion to $\Omega^{-0.6}_m$ (Zheng et al. 2002).

Figure 5 shows $\Delta \chi^2$ for the best-fit HOD model at each value of $\Omega_m$, as different observables are added in succession. Figures 6 and 7 show the parameters of these best-fit HODs. If we
include only the spatial clustering observables $\sigma_g(r)$, $n_{\text{grp}}(\geq N)$, and $\xi_N(r)$, we obtain results virtually identical to those of the central model by simply shifting the mass scale of the mean occupation function in proportion to $\Omega_m$. For these observables alone, we get $\chi^2 = 0$ at all $\Omega_m$ (Fig. 5, horizontal dotted line). Figure 6 plots the best-fit $\langle N(M) \rangle$ as a function of the scaled halo mass $M(0.3/\Omega_m)$. Because the constraints imposed by our observables on the low end of $\langle N(M) \rangle$ are weak, the cutoff profiles of the scaled mass functions vary, but above $\langle N \rangle \sim 1.2$ they are nearly indistinguishable. Thus, the spatial clustering observables effectively fix $\langle N(M) \rangle$ in terms of scaled halo masses, and only small further adjustments can be accommodated when other observables are introduced. As shown in Figure 7, the best-fit models have $\omega = 1$ just like the central model HOD, and they have $\Delta \log c_0 \approx 0$, although there is a slight reduction in halo concentration at low $\Omega_m$ to compensate for the higher concentrations of the halos themselves.

The $\Omega_m$ degeneracy breaks as soon as we include dynamical measures that are sensitive to the absolute mass or velocity scale of halos. For low $\Omega_m$, the cutoff of $\langle N(M) \rangle$ smooths out so that more massive galaxies come from lower mass halos with lower bias factors ($\text{dashed lines in Figs. 6a and 6b}$). The $P(N|M)$ distribution broadens slightly ($\omega > 1$ in Fig. 7a) to compensate for a drop in one-halo pairs, and concentrations drop ($\Delta \log c_0 < 0$ in Fig. 7b) so that these pairs move to larger separations. However, within the constraints imposed by the multiplicity function, the correlation function, and the galaxy number density, there is virtually no room to change the galaxy bias factor in a way that compensates changes in $\Omega_m$. For $\Omega_m$ in the range $0.2-0.4$, the $\Delta \chi^2$ values in Figure 5 are almost equal to those expected for constant $b_g$, given our adopted 10% uncertainty in $\Omega_m$. When we add galaxy pairwise velocity dispersions as observables, the velocity bias parameter $\alpha_v$ adjusts to compensate for changes in $\Omega_m$ (Fig. 7c, long-dashed line). However, while velocity bias can reduce velocity dispersions within halos, it cannot affect the pairwise velocities of the halos themselves, which scale as $\Omega_m^{0.6}$. For high $\Omega_m$, $\alpha_v$ drops by more than the factor $(0.3/\Omega_m)^{1/2}$ that would keep internal dispersions fixed, in an effort to compensate for the higher halo velocities. This change would drive the pairwise dispersion too low on small scales, where the one-halo term dominates, so halo concentrations increase, allowing more of the small-separation pairs to come from massive halos. In response to these higher concentrations, the width parameter $\omega$...
drops, reducing the one-halo contribution to $\xi_{eq}(r)$. Despite the freedom introduced by velocity bias, the combination of pairwise dispersions at small and large scales adds significant discriminatory power on the value of $\Omega_m$. For $\Omega_m = 0.4$, for example, $\Delta \chi^2$ increases from $\sim 3.5$ ($\beta$ alone) to $\sim 5.6$, with the two dynamical measurements producing most of the discrepancy.

When we add the $\langle M_{\text{vir}}(N) \rangle$ observables, the velocity bias is forced close to the value $c_v = (0.3/\Omega_m)^{1/2}$ that keeps these apparent virial masses fixed (Fig. 7c, dot-dashed line, nearly obscured by solid line). Since there is no longer room to improve the large-scale pairwise dispersion by “overcompensating” with $c_v$, the concentration and width parameters return to the values preferred by the spatial clustering observables (note the similarity of dotted and dot-dashed lines in Figs. 7a and 7b, and the difference between these lines and the long-dashed lines). Although $\alpha_v$ can exactly compensate a pure $\Omega_m$ shift for the $\langle M_{\text{vir}}(N) \rangle$ observables, the match to the pairwise dispersion becomes worse as a result, and $\Delta \chi^2$ increases slightly.

The galaxy-galaxy lensing observable $\Omega_m \xi_{\text{gm}}(r)$ provides a diagnostic of host halo masses that is unaffected by velocity bias. If $\langle N(M) \rangle$ stays fixed as a function of scaled mass, then a pure $\Omega_m$ change simply scales $\Omega_m \xi_{\text{gm}}(r)$ in proportion to $\Omega_m$ (eq. [25]). The effect of higher $\Omega_m$ can be partly compensated by reducing halo concentrations (Fig. 7b, solid line), which dilutes the contribution from satellite galaxy and matter particle pairs. This decrease drives an increase in $\omega$ (Fig. 7a) to restore lost one-halo galaxy pairs at small separations (note that $\xi_{\text{gm}}$ itself is independent of $\omega$). However, altering concentrations has limited power to compensate for a shift in the halo mass function, and $\Omega_m \xi_{\text{gm}}(r)$ measurements add substantially to $\Delta \chi^2$. Measurements in the linear or near-linear regime, not considered here, would add still more discriminatory power.

Finally, large-scale bispectrum measurements have little sensitivity to a pure $\Omega_m$ shift (Fig. 5, long-dashed line). This statistic is mainly a diagnostic for the galaxy bias factor $b_g$. Along this model sequence, the amplitude of dark matter clustering does not change, and the galaxy bias factor is therefore determined by the observed spatial clustering, independent of $\Omega_m$. The bispectrum will become important when we consider sequences with changing $\sigma_8$.

Overall, these galaxy clustering observables have great power to constrain any pure shift in $\Omega_m$. A model with $\Omega_m \neq 0.3$ has a likelihood $\exp(-\Delta \chi^2/2)$ relative to the central model, so even a model with $\Delta \chi^2 = 4.6$ is disfavored by a factor of 10. Including all observables, we find $\Delta \chi^2 \sim 5.8$ and 5.0 for $\Omega_m = 0.25$. 

---

**Fig. 4.** Variations of (a) halo mass function and (b) mass correlation function with cosmology. The fitting formula of Sheth & Tormen (1999) for the halo mass function and that of Smith et al. (2003) for the nonlinear mass power spectrum are used for the calculations. In each panel, we show the results for the central cosmological model ($\Omega_m = 0.3, \sigma_8 = 0.9$, and $\Gamma = 0.2$; open circles), a pure change in $\Omega_m$ to 0.5 (dotted line), a pure change in $\sigma_8$ to 1.0 (dashed line), a cluster-normalized model with $\Omega_m = 0.5$ and $\sigma_8 = 0.7$ (dot-dashed line), and a cluster-normalized model in which we also change the shape parameter to $\Gamma = 0.11$ to preserve the halo mass function as closely as possible (solid line).

**Fig. 5.** Values of $\Delta \chi^2$ as a function of $\Omega_m$, relative to the true model with $\Omega_m = 0.3$, in a cosmological model sequence where only $\Omega_m$ is varied. As each set of observables is added in the analysis (in the order indicated in the legend), the $\chi^2$ for each cosmological model is minimized by solving the best-fit HOD parameters. Different line (point) types show a sequence of $\Omega_m$ constraints from including more and more complementary observables. The inset box shows the parts of the lines with $\Delta \chi^2 < 6$. [See the electronic edition of The Journal for a color version of this figure.]
and 0.35, respectively, implying that $\Delta \Omega_m = 0.05$ can be firmly ruled out. These constraints weaken significantly if we omit the weak-lensing observables with their direct sensitivity to mass scales. Still, even with $\beta$ and $\sigma_s(r)$ as the only dynamical observables, the $\Omega_m = 0.2$ and 0.4 models are rejected with $\Delta \chi^2 > 5$.

4.3. Changing $\sigma_s$ with $\Omega_m$, $n_s$, and $\Gamma$ Fixed

We now consider models that differ only in $\sigma_s$, the amplitude of the linear matter power spectrum, with $\Omega_m$ and the shape of $P(k)$ held fixed. The dashed lines in Figure 4 show the effect of raising $\sigma_s$ from 0.9 to 1.0. The amplitude of the matter correlation function increases at all scales, and the space density of high-mass halos ($M \geq 10^{14} h^{-1} M_\odot$) rises, while the space density of galaxy mass halos ($M \sim 10^{12} - 10^{13} h^{-1} M_\odot$) is nearly unchanged. Figure 8 shows $\Delta \chi^2$ for the best-fit HOD models along the changing $\sigma_s$ sequence, and Figure 9 shows the properties of these best-fit models.

Figure 9b shows the response of $\langle N(M) \rangle$ to $\sigma_s$ when the spatial clustering observables $\sigma_s$, $\xi_{gg}$, and $n_{grp}$ are the only constraints. Higher $\sigma_s$ leads to lower $\langle N(M) \rangle$ at high masses, so that $n_{grp}$...
remains fixed despite an increased halo abundance. This change reduces the galaxy bias factor \( b_g \), but on its own it is not enough to compensate for the increased matter clustering amplitude. The \( \langle N(M) \rangle \) cutoff therefore spreads to lower masses, putting more isolated galaxies in low-bias halos while retaining the galaxy number density \( n_g \). The \( P(N|M) \) width parameter rises to \( \omega > 1 \) to restore small-scale pairs lost when the number of halos in the \( \langle N(M) \rangle \sim 1-2 \) range is reduced (Fig. 9c, dotted line).

For spatial clustering observables alone, these HOD changes effectively compensate for changes in the halo population, and the \( \Delta \chi^2 \) curve is nearly flat in the vicinity of the central model (Fig. 8, dotted line). However, these observables pin down \( \langle N(M) \rangle \) fairly precisely, leaving little room for adjustment when dynamically sensitive observables are added as constraints. The \( \Delta \chi^2 \) from \( \beta = \Omega_m^0 b_g/\sigma_8 \) is close to that expected for our 10% error bar and the scaling \( b_g \propto 1/\sigma_8 \) implied by fixed galaxy spatial clustering. The velocity bias parameter \( \alpha_v \) can partly compensate for changes in halo velocities and internal dispersions, but it cannot fix the small- and large-scale values of \( \sigma_8(r) \) simultaneously (see discussion in § 4.2), so \( \Delta \chi^2 \) rises when \( \sigma_8 \) is added. Adding \( \langle M_{vir}(N) \rangle \) effectively pins down \( \alpha_v \) to the value needed to keep the apparent virial masses fixed, making the overall fit of \( \sigma_8(r) \) worse and further increasing \( \Delta \chi^2 \). The \( \Omega_m \xi_{gm}(r) \) observable is unaffected by velocity bias, and the small adjustments still allowed in \( \langle N(M) \rangle \), \( \Delta \log c_0 \), and \( \omega \) do little to counter the changes in halo masses. Finally, the galaxy bispectrum increases \( \Delta \chi^2 \) through its direct sensitivity to \( b_g \). With all observables considered together, \( \Delta \chi^2 \) exceeds 4.0 for \( \Delta \sigma_8 = \pm 0.1 \), and it rises rapidly for larger changes.

4.4. Cluster-normalized Models

The observed abundance of massive galaxy clusters effectively constrains the amplitude of the halo mass function at a scale \( M \sim 5 \times 10^{14} h^{-1} M_{\odot} \). The halo abundance at this scale is an increasing function of \( \Omega_m \) and \( \sigma_8 \), and the cluster normalization constraint is usually expressed in the form \( \sigma_8 \Omega_m^0 = \text{const.} \), with \( q \sim 0.5 \) (e.g., White et al. 1993; Eke et al. 1996; Pierpaoli et al. 2001). Here, as in Zheng et al. (2002), we consider a sequence of cluster-normalized cosmological models that satisfy \( \sigma_8 = 0.9(\Omega_m^0/0.3)^{0.5} \), with the shape of \( P(k) \) held fixed. The dotted lines in Figure 4 show the effect of simultaneously increasing \( \Omega_m \) to 0.5 and decreasing \( \sigma_8 \) to 0.70. As expected, the halo mass function stays the same at \( M \sim 5 \times 10^{14} h^{-1} M_{\odot} \), but relative to the fiducial model there are more low-mass halos and fewer halos with \( M \gtrsim 10^{15} h^{-1} M_{\odot} \). The matter correlation function is lower at all scales because of the lower value of \( \sigma_8 \).

Figures 10 and 11 show the \( \Delta \chi^2 \) and parameters of the best-fit HOD models along the cluster-normalized sequence, in the same format as Figures 8 and 9. The multiplicity function fixes the high-mass end of \( \langle N(M) \rangle \) in a way that is nearly independent of \( \Omega_m \), although there are slight variations in slope that reflect the changing slopes of the halo mass function (Fig. 11b). The higher abundance of low-mass halos for high \( \Omega_m \) requires an increase in the cutoff mass to keep the galaxy number density constant. The tightened mass range for single-galaxy halos leads to a larger one-halo contribution to \( \xi_{gg}(r) \), which is compensated by decreasing the width parameter \( \omega \) and the concentration parameter \( \Delta \log c_0 \). Although the matter correlation amplitude is reduced for high \( \Omega_m \) and low \( \sigma_8 \), the bias of halos at fixed mass is higher, and these two effects conspire to keep the large-scale amplitude of the galaxy correlation function nearly constant over a substantial range of \( \Omega_m \). For spatial clustering observables alone, \( \Delta \chi^2 \) is nearly flat over the range \( 0.2 \leq \Omega_m \leq 0.5 \).

Because cluster normalization keeps the mass and velocity scale of halos roughly fixed, the addition of dynamically sensitive observables makes relatively little difference to best-fit HOD
parameters, and $\Delta \chi^2$ increases more slowly than it does for pure $\Omega_m$ or pure $\sigma_8$ changes (note the smaller vertical scale of Fig. 10 relative to Figs. 5 and 8). The detailed changes in $\omega$, $\Delta \log c_0$, and $\alpha$, as new observables are added mostly reflect the changes in the cutoff of $\langle N(M) \rangle$ and corresponding changes in the fraction of single-occupancy halos. The $\Delta \chi^2$ curves are asymmetric in $\Omega_m$ because of the nonlinear relation between $\Omega_m$ and $\sigma_8$. For example, cluster normalization requires a 13% reduction in $\sigma_8$ to 0.78 at $\Omega_m = 0.4$ but a 22% increase to $\sigma_8 = 1.10$ at $\Omega_m = 0.2$, and the still larger $\sigma_8$ values required at lower $\Omega_m$ are easily ruled out. The bispectrum constraint is especially useful in distinguishing cluster-normalized models because it responds to $b_g$ independent of $\Omega_m$. Despite the compensating effects of $\sigma_8$ and $\Omega_m$ in cluster-normalized models, the full set of observables yields $\Delta \chi^2 \sim 5.5$ for $\Omega_m = 0.4$ and $\Delta \chi^2 \sim 13.1$ for $\Omega_m = 0.2$.

As shown by Zheng et al. (2002), one can construct cluster-normalized model sequences that match the amplitude and slope of the halo mass function at cluster scales by changing the power spectrum shape parameter in concert with $\Omega_m$ and $\sigma_8$. The solid lines in Figure 4 show a model with $\Omega_m = 0.5$, $\sigma_8 = 0.71$, and $\Gamma = 0.11$. The halo mass function is nearly identical to the central model’s for $M \gtrsim 10^{13.5} \ h^{-1} M_{\odot}$, although it is higher at low halo masses. The shape of the matter correlation function is significantly different because of the different shape of $P(k)$.

Figure 12 shows $\Delta \chi^2$ as a function of $\Omega_m$ along this model sequence. The required values of $\sigma_8$ and $\Gamma$ are shown in Figure 13 of Zheng et al. (2002), and the values of $\Gamma$ are also marked on the top axis of our Figure 12. As speculated by Zheng et al. (2002), matching the slope of the halo mass function by changing the shape of $P(k)$ does more harm than good: $\Delta \chi^2$ values at a given $\Omega_m$ are always higher than they are for the fixed $P(k)$, cluster-normalized models shown in Figure 10. With all observables included, models with $\Omega_m = 0.2$ and 0.4 have $\Delta \chi^2 \sim 17.6$ and $\sim 10.2$, respectively.

### 4.5. More General Cases

The four model sequences considered in Figures 5, 8, 10, and 12 trace four different curves through the $(\Omega_m, \sigma_8, \Gamma)$ space of cosmological parameters. We now examine the constraints and degeneracies in this parameter space in more general terms. The computational expense of finding the best-fit, 10-parameter HOD model for each cosmological model makes a comprehensive
BREAKING DEGENERACIES BETWEEN COSMOLOGY AND BIAS

No. 1, 2007

17

study of the full 3D space difficult. However, the fact that changing the power spectrum shape to match the halo mass function makes models easier to distinguish suggests that the “Γ” dimension is largely decoupled from the Ω_m-σ_8 plane. Furthermore, the galaxy power spectrum shape can be measured directly on large scales (e.g., Tegmark et al. 2004a, 2006; Cole et al. 2005), and we have not incorporated these direct constraints into our observables.

To begin, therefore, we fix Γ = 0.2 and map out Δχ² contours in the Ω_m-σ_8 plane, finding the best-fit HOD parameters that minimize the values of Δχ² at each point on the cosmological parameter grid. We include all the observables discussed in § 3 and incorporated in our earlier constraints. The solid contours in Figure 13a show Δχ² = 2.30 and 6.17, corresponding to 68.3% and 95.4% confidence intervals for two parameters. The most degenerate direction seems to be σ_8Ω_m^{0.6} const with q ~ 0.7–0.8, which deviates from the degeneracy axis of cluster normalization (q ~ 0.5) or large-scale redshift-space distortions (q ~ 0.6). This deviation reflects the steeper scaling favored by the galaxy-galaxy lensing constraint, which has q ~ 1 even on these nonlinear scales (Yoo et al. 2006), and by pairwise velocity dispersions with α_v as an adjustable parameter.

Dotted contours in this panel show Δχ² = 1 and 4, from which one can read projected 1 and 2 σ constraints on the individual parameters σ_8 and Ω_m. The remaining panels show the 1D likelihoods for these parameters and the combination σ_8Ω_m^{0.6}, inferred by integrating the likelihood function \( \exp \left( -\Delta \chi^2/2 \right) \) over each bin of the relevant quantity. Under our assumption of 10% fractional error in each observable, the 1 σ uncertainty in the matter density parameter Ω_m is about 11%. The amplitude σ_8 of the matter power spectrum has a 1 σ constraint that is better than 10%. The 1 σ uncertainty in the best constrained quantity, σ_8Ω_m with q = 0.75, is at the 5% level. If we force q to 0.5, the precision is almost equally good.

What constraints could we infer on Γ from the galaxy clustering measurements considered in our observable set? To address this question in approximate terms, we fix Ω_m and σ_8 to the central model values of 0.3 and 0.9 and vary Γ alone, again finding best-fit HOD parameters and minimum Δχ². We obtain Δχ² = 1 at Γ = 0.183, 0.219 and Δχ² = 4 at Γ = 0.167, 0.240. The 10% (1 σ) constraint on Γ is similar to that found along our matched halo mass function sequence (Δχ² = 1 at Γ = 0.177, 0.223 and Δχ² = 4 at Γ = 0.161, 0.250; Fig. 12), supporting our conjecture that the (Ω_m, σ_8) constraints are mostly orthogonal to the \( P(k) \) shape constraints. With our 30 observables and assumption of 10%, independent fractional errors, therefore, the full χ² surface approximately follows the contours of Figure 13 with an orthogonal Gaussian of roughly 10% 1 σ width in the Γ direction.

As discussed in § 1, traditional approaches to testing cosmological models with large-scale structure focus on the perturbative regime, where it is reasonable to adopt linear theory and linear bias for describing the galaxy power spectrum and redshift-space distortions and second-order perturbation theory and a quadratic bias model for describing the galaxy bispectrum. It is interesting to compare our HOD-based constraints to those that could be derived from the same observables with the perturbative approach. Specifically, we use the σ_8, β, and Q_0 observables adopted previously, with the same 10% uncertainties. We employ a two-parameter quadratic bias model, and we calculate the observables for a given cosmology and bias using σ_8 = gσ, β = 0.5σ_q/β, \( Q_0 = Q_m/b + b/\beta^2 \), where σ and Q_m are the rms fluctuations and reduced bispectrum for the dark matter. We examine an (Ω_m, σ_8) grid with Γ = 0.2 and find the values of b and b_2 that minimize Δχ² for each cosmological model.

The top left panel of Figure 14 compares our HOD-based constraints (repeated from Fig. 13) to those obtained from the perturbative analysis. The principal degeneracy axis for the latter constraints is σ_8Ω_m^{0.6} const, since β constrains Ω_m^{0.6}/b and σ_q constrains bσ_q. The bispectrum constraint on b breaks this degeneracy and produces closed contours. However, the constraints from the full HOD-based analysis of all observables are much tighter than the constraints from the smaller set of observables that can be modeled by the perturbative approach alone. Of course, the perturbative cosmology constraints can be improved by including more observables, such as large-scale weak-lensing measurements or CMB anisotropies, but these observables will also improve the HOD-based constraints. Thus, despite the need for a more complex, multiparameter model of galaxy bias, the ability of the HOD approach to model intermediate and small-scale clustering leads to substantial improvements in cosmological power. At the same time, an HOD analysis of observed galaxy clustering yields valuable tests of theoretical models of galaxy formation.

The obtainable constraints on cosmological parameters depend, of course, on the precision of the clustering measurements. Our assumption of 10% fractional uncertainty on each observable is probably conservative, and current galaxy clustering data have already yielded higher precision for some of these observables, such as the two-point correlation function. To study the impact of improving measurement precision, we change the assumed error bars of different sets of observables by turns and recalculate Δχ² contours. This investigation also highlights the relative cosmological sensitivity of the different observables. Each panel of Figure 14 (except for the top left panel discussed previously) shows changes in the (Ω_m, σ_8) constraints caused by improving the measurement precision of one set of observables. The two dotted contours are the 1 and 2 σ confidence levels (Δχ² = 2.30, 6.17) obtained by assuming 10% fractional errors for all sets of observables; they match the solid contours in Figure 13a. The solid contours are obtained by assuming 5%
errors for the set of observables indicated in the upper right corner of the panel, keeping 10% for other sets.

With the galaxy bias factor constrained by other observables, a better measurement of galaxy density contrasts $\sigma_g$ leads to a better constraint on $\sigma_8$, shrinking contours at both ends. The tighter $\sigma_8$ constraint in turn produces a slightly tighter $\Omega_m$ constraint, but there is essentially no improvement in the $\sigma_8 \Omega_m$ constraint (i.e., the width of the contours in the narrow direction is unchanged). Reducing the observational error bar on $Q_g$ produces a similar improvement for a different reason: the more precise measurements of $Q_g$ allow a better constraint on the galaxy bias factor $b_g$, so they help to break the degeneracy between $b_g$ and the amplitude of the dark matter power spectrum.

Improving the pairwise velocity dispersion measurements produces modest improvements in both the $\sigma_8$ constraint and the individual constraints on $\sigma_8$ and $\Omega_m$. The former improvement comes mainly from the large-scale velocity dispersion, which depends on halo velocities and thus scales roughly with $\sigma_8 \Omega_m$. The individual parameter improvements come mainly from the interplay of small-scale $\sigma_v$ with the apparent virial mass $\langle M_{\text{vir}}(N) \rangle$. Better $\sigma_v$ measurements yield tighter constraints on the velocity bias parameter, and the small-scale $\sigma_v$ and $\langle M_{\text{vir}}(N) \rangle$ measurements then constrain the halo mass scale. Improving the $\langle M_{\text{vir}}(N) \rangle$ measurements on their own yields better $\sigma_8$ and $\Omega_m$ constraints but does not change the $\sigma_8 \Omega_m$ constraint.

Improving the $\xi_{gg}(r)$ measurements leads to a significantly better determination of $\sigma_8$ for two reasons: it tightens the constraints on $P(N|M)$ and thus on the large-scale galaxy bias factor, and it improves the measurement of the large-scale galaxy clustering amplitude itself. These two effects are analogous to those discussed earlier for $Q_g$ and $\sigma_8$, respectively. Although $\xi_{gg}(r)$ is not directly sensitive to $\Omega_m$, the tighter $\sigma_8$ constraint produces a tighter $\Omega_m$ constraint when combined with dynamical measurements. Improving $n_{\text{grp}}(\geq N)$ measurements produces surprisingly little improvement in the cosmological parameter constraints. If we instead degrade the fractional errors on $n_{\text{grp}}(\geq N)$ to 50%, the constraints (shown by light solid contours) get noticeably but not dramatically worse. While it is always possible to reproduce $n_{\text{grp}}(\geq N)$ exactly by choosing the appropriate $\langle N(M) \rangle$, this choice pins down the mass associated with a given multiplicity, and it largely pins down the galaxy bias factor. It appears, however, that the $\xi_{gg}(r)$ measurements already determine the high end of $\langle N(M) \rangle$ well enough for these purposes, so that moderate improvements in $n_{\text{grp}}(\geq N)$ do not add much power. If we degrade the presumed measurement errors on $\xi_{gg}(r)$ to 25%, then the impact of improving the $n_{\text{grp}}(\geq N)$ measurements is larger, but it is
form a test in which we fix the cutoff profile width $\sigma_{\log M}$ to the central model value. The resulting cosmological constraints are virtually identical to those in Figure 13, where $\sigma_{\log M}$ is not fixed. However, if the value of the absolute cutoff mass scale can also be pinned down, e.g., from the Tully-Fisher relation or other dynamical measurements in the single-galaxy regime, then cosmological constraints improve substantially. We have investigated the idealized case in which $M_{\text{min}}$ and $\sigma_{\log M}$ are fixed to the values of the central model. We maintain the galaxy number density by adjusting the amplitude of the satellite galaxy occupation function. The strong constraint on the halo mass scale from the value of $M_{\text{min}}$ produces a tight constraint on $\Omega_m (0.300 \pm 0.008)$, which in turn leads to a much tighter constraint on $\sigma_8 (0.90 \pm 0.04)$.

Finally, we consider the constraints achievable with a brighter, lower density galaxy sample. Different samples of galaxies have different HODs, but they should probe the same underlying cosmology. All the investigations we have presented so far are based on the luminosity-threshold sample with a number density $\bar{n}_g = 0.01 \, h^3 \, \text{Mpc}^{-3}$, roughly corresponding to the $M_r < -19.5 + 5 \log h$ galaxies in the SDSS. We increase the mass scale of our central model HOD to construct a galaxy sample with number density 10 times smaller ($\bar{n}_g = 0.001 \, h^3 \, \text{Mpc}^{-3}$), which approximates the SDSS $M_r < -21 + 5 \log h$ sample (Zehavi et al. 2005). With the same assumption of 10% fractional error on each of the 30 observables, we find that the cosmological constraints from these two galaxy samples are quite similar, with the degeneracy direction $\sigma_8 \Omega_m^2$ almost exactly linear on $q$ for the lower density sample. Thus, measurements for these two largely independent galaxy samples should allow an important consistency check on cosmological conclusions and, if the results are combined, a $\sim \sqrt{2}$ reduction in statistical errors.

### 4.6. The “Influence Matrix”

Perhaps the most important implication of Figure 14 is that the cosmological constraints do not rely on one or two clustering statistics but instead emerge from the interlocking web of measurements. Figure 15 demonstrates this point in a different way. It shows the “influence matrix,” which we define in terms of the partial derivatives $\partial \ln F / \partial \ln a_i$ evaluated at the central HOD and cosmological model. Each predicted observable $F_i$ is a function of cosmological and HOD parameters, and each $a_i$ represents one of these parameters, so the influence matrix encodes the response of the observables to the individual model parameters. When evaluating the partial derivatives, we adjust $M_{\text{min}}$ so that the galaxy number density stays fixed, but we keep the remaining HOD and cosmological parameters (other than $a_i$) fixed. The influence matrix is closely related to the Fisher matrix,

$$
\frac{\partial^2 \ln L}{\partial a_i \partial a_j} = \left( \frac{\partial F}{\partial a_i} \right)^T \mathbf{C}^{-1} \left( \frac{\partial F}{\partial a_j} \right),
$$

where $F$ is the vector of the predicted observables and $\mathbf{C}$ is the covariance matrix of the errors. However, while the Fisher matrix depends strongly on the assumed error covariance matrix $\mathbf{C}$, the influence matrix is independent of this assumption, and it reveals the dependence of observables on model parameters in a more transparent manner.

Figure 15 shows the influence matrix calculated at the central model. The area of each circle is proportional to the element of the matrix from the corresponding observable (vertical axis) and parameter (horizontal axis). The HOD parameter $M_{\text{min}}$, determined by matching the galaxy number density, is presented here...
as an observable, since its value could in principle be probed by dynamical or weak-lensing measurements in the single-galaxy regime. If two columns have similar elements, then the corresponding parameters are largely degenerate, since an increase in one can be compensated by a decrease in the other.

The first three columns of Figure 15 show how clustering observables vary with cosmological parameters. Most elements of $\partial \ln F/\partial \ln \Omega_m$ and $\partial \ln F/\partial \ln \sigma_8$ are positive. An increase in $\Omega_m$ shifts the halo mass function to higher mass scales (Fig. 4), and with the fixed satellite HOD, this shift increases the group multiplicity function $n_{\text{grp}}$. To keep a fixed galaxy number density, $M_{\text{min}}$ increases by a larger factor than $\Omega_m$, and the shift of galaxies to halos with larger $M/M_*$ leads to stronger galaxy bias for spatial clustering statistics $\sigma_8$ and $\xi_{gg}$. An increase in $\sigma_8$ boosts the matter clustering directly. Raising either parameter increases halo masses and velocities, and thus the values of dynamical observables. The similar form of these vectors implies significant degeneracy between these two parameters, with an overall trend that roughly follows $\partial \ln F/\partial \ln \Omega_m = 0.75 \partial \ln F/\partial \ln \sigma_8$, but the degeneracy is not complete. The power spectrum shape parameter $\Gamma$ is largely decoupled from $\Omega_m$ and $\sigma_8$. The much stronger sensitivity of $M_{\text{min}}$ to $\Omega_m$ than to any other cosmological or HOD parameter explains why a tight observational constraint on $M_{\text{min}}$ can produce such a tight constraint on $\Omega_m$ ($\S$ 4.5). The sensitivity of spatial clustering statistics to $\Omega_m$, which contrasts with the flat $\Delta \chi^2$ shown by the dotted line in Figure 5, arises because we here keep all mass scales in the HOD fixed except for $M_{\text{min}}$. If we instead scale all masses $\propto \Omega_m$, then the derivatives of $\xi_{gg}$, $n_{\text{grp}}$, and $\sigma_8$ with respect to $\Omega_m$ essentially vanish, as expected, while the derivatives of $\langle M_{\text{vir}}(N) \rangle$ become one instead of zero.

The rightmost nine columns in Figure 15 show how clustering observables vary with the HOD parameters, with cosmological parameters held fixed. A degeneracy between a cosmological parameter and HOD parameters exists to the extent that the “influence vector” $\partial \ln F/\partial \ln a_C$ can be approximated by a linear combination of the influence vectors $\partial \ln F/\partial \ln a_{H,1}$, where $a_C$ is the cosmological parameter and $a_{H,1}$ are the HOD parameters. As expected from our earlier results, a substantial degeneracy exists if we restrict $F$ to the spatial clustering observables $\xi_{gg}$, $n_{\text{grp}}$, and $\sigma_8$. For example, an increase in $\sigma_8$ can be largely compensated by decreases in the parameters $S_{14}$ and $S_{15}$ that describe the high end of $\langle N(M) \rangle$, and changes in $\omega$ and $\Delta \log c_0$ can take up some amount of residual difference. However, with the full set of observables, the structures of $\partial \ln F/\partial \ln \Omega_m$, $\partial \ln F/\partial \ln \sigma_8$, and

Fig. 13.—Constraints on $\Omega_m$ and $\sigma_8$ (with $\Gamma$ fixed) from galaxy clustering data. (a) Contour plot. The two dotted contours correspond to $\Delta \chi^2 = 1$ and 4, and the two solid ones are for $\Delta \chi^2 = 2.30$ and 6.17 (i.e., about the 68.3% and 95.4% confidence levels for two parameters). The two dashed lines represent $(\sigma/0.9)(\Omega_m/0.3)^q = 1$ with $q = 0.5$ (thick) and $q = 0.75$ (thin). The other three panels show marginalized likelihoods on $\sigma_8$, $\Omega_m$, and $\sigma_1\Omega_m^2$, respectively. Vertical dotted lines in each of these three panels mark the central 68.3% of the distribution (i.e., the 1$\sigma$ range). In (d), thin and thick lines are for $q = 0.5$ and 0.75, respectively.
\[ \partial \ln F / \partial \ln \Gamma \] are quite different from any of the influence vectors of individual HOD parameters.

The visual structure of Figure 15 suggests that there is no linear combination of the HOD influence vectors that accurately approximates the cosmological parameter influence vectors. We verify this suggestion by solving for the combination coefficients that minimize the squared difference between the \( \Omega_m \) influence vector and the corresponding linear combination of the HOD influence vectors. (We do not include \( M_{\rm min} \) as an observable here, since it probably cannot be measured as precisely as the other quantities.) Figure 16a shows the HOD influence vectors multiplied by these linear combination coefficients. The \( S_{14} \) and \( S_{15} \) parameters dominate the best-fit linear combination, with \( \alpha_1 \) compensating for the increase in \( \langle M_{\rm vir} \rangle \) that would otherwise occur. Figure 16c shows that this best-fit linear combination (open circles in Fig. 16c) is a better approximation to the \( \Omega_m \) influence vector, with clear improvement in \( \sigma_v \) and \( \Omega_m \zeta_{\rm gm} \). Figure 16b shows that the contribution from the \( \sigma_8 \) influence vector is much larger than that from any of the HOD influence vectors. Nonetheless, the changes in \( S_{14} \) and \( S_{15} \) significantly improve the agreement with the multiplicity function and small-scale correlation function, relative to a pure \( \sigma_8 \) change with no freedom in the HOD. As argued earlier, the shape parameter \( \Gamma \) is largely decoupled from the strong cosmological degeneracy between \( \Omega_m \) and \( \sigma_8 \).

Figure 15 also shows that clustering observables are insensitive to some HOD parameters, most notably \( S_{11} \) and \( \sigma_{\log M} \), which describe the low-mass end of the mean occupation function. We reached similar conclusions from our analysis of \( \langle N(M) \rangle \) constraints in §4.1 and the effect of a \( \sigma_{\log M} \) prior in §4.5. The insensitivity of clustering observables to these parameters means that it is difficult to constrain them with clustering data. However, this insensitivity also means that poor knowledge of these parameters does not introduce uncertainty in the cosmological conclusions.
We have also investigated the influence matrix for a simpler HOD parameterization in which the mean satellite occupation function is a truncated power law (Kravtsov et al. 2004; Zheng et al. 2005). The influence vector of the power-law normalization is similar to that of $S_{14}$ in Figure 15, while the influence vector of the power-law slope is similar in overall form but with a different mass and length scale dependence, reflecting its differential impact on high- and low-mass halos. The influence of the low-mass cutoff is, again, weak.

5. ENVIRONMENTAL VARIATION OF THE HOD

The halo occupation function $P(N|M)$ is, by definition, a distribution weighted by halo number, the probability that a randomly selected halo of mass $M$ contains $N$ galaxies. Our clustering calculations implicitly assume that this distribution and other parameters of the HOD like $S_{14}$ are independent of the larger scale environment. A systematic dependence of a halo’s galaxy occupation on the overdensity $\delta$ of its surroundings would alter the clustering predicted for a given halo population and (number averaged) $P(N|M)$. Ignoring this dependence could therefore lead one to infer an incorrect HOD or incorrect cosmological parameters when fitting observed galaxy clustering.

The assumption of an environment-independent HOD is rooted partly in the Bond et al. (1991) excursion set derivation of the extended Press-Schechter formalism. This derivation predicts that the statistical features of a halo’s assembly history depend only on its present mass, a point emphasized by White (1996). In agreement with this prediction, Lemson & Kauffmann (1999) and Sheth & Tormen (2004) found that $N$-body halos of fixed mass in different environments have similar properties and formation histories, although Sheth & Tormen (2004) found a subtle trend for close pairs of halos to have higher formation redshifts. The resolution of the simulations used in these studies effectively limited them to halo masses $M \gtrsim 10^{13} M_\odot$. Two recent studies (Gao et al. 2005; Harker et al. 2006), based on the higher resolution, Millennium Run $N$-body simulation (Springel et al. 2005), show that the correlation between halo formation time and large-scale environment becomes much stronger for halo
masses $M \lesssim 10^{12.5} \, M_\odot$. To the extent that galaxy properties depend on halo formation time and halo mass, this correlation will in turn produce an environmental dependence of $P(N|M)$.

There are several reasons for thinking that the quantitative impact of such a dependence is small, at least for galaxy samples defined by simple thresholds in luminosity or baryonic mass. Empirically, Blanton et al. (2006) show that the correlation of SDSS galaxy properties with galaxy overdensity on a scale of $8 \, h^{-1} \, \text{Mpc}$ can be entirely explained by the correlation with overdensity on the $1 \, h^{-1} \, \text{Mpc}$ scale characteristic of large halos; at fixed local overdensity, they detect no residual correlation with the larger scale environment. On the theoretical side, Berlind et al. (2003) show that the HOD (more precisely, the mean occupation function) of galaxies above a baryonic mass threshold in the Weinberg et al. (2004) cosmological hydrodynamic simulation is independent of environment within the statistical uncertainties imposed by the $(50 \, h^{-1} \, \text{Mpc})^3$ simulation volume. Yoo et al. (2006), analyzing the same simulation, show that explicitly eliminating any environmental dependence by randomly shuffling the galaxy populations among halos of similar mass changes the galaxy-galaxy and galaxy-matter correlation functions by $<5\%$ on scales $0.1 \, h^{-1} \, \text{Mpc} < r < 5 \, h^{-1} \, \text{Mpc}$, again within the statistical uncertainties. In a similar experiment using semianalytic galaxy populations in the Millennium Run simulation, Croton et al. (2007) find that the correlation functions of luminosity bin samples change by a few percent; because of the large simulation volume, these shifts are measured with high precision.

Since we hope in the long term to achieve few percent precision on cosmological parameter measurements, the Millennium Run results imply that environmental dependence of the HOD must eventually be considered as part of the program outlined in this paper. Allowing arbitrary dependence of $P(N|M)$ on large-scale overdensity $\delta$ would probably make the problem intractable and degenerate, but given the small magnitude of the anticipated effects, a more restricted parameterization should suffice. For example, since the age dependence of halo clustering is found mainly in the mass range of individual galaxy halos, it may well be enough to allow $M_{\min}$ to vary with $\delta$, or to treat the effective bias parameter of low-mass halos as a parameter to be fitted instead of using the standard $N$-body result. Since much of the signal in...
most clustering statistics comes from multigalaxy halos, addition of such parameters may not have much impact on the achievable constraints on cosmological parameters. Semianalytic mock catalogs will be valuable in guiding the choice of parameterizations and in testing whether fitting methods that adopt a particular parameterization (or that ignore environmental variations altogether) yield incorrect estimates of HOD or cosmological parameters.

For cosmological applications, it will be desirable to work with galaxy samples defined by properties that are sensitive to halo mass and relatively insensitive to formation history, such as stellar mass or luminosity in a band dominated by old stellar populations. Color-selected or morphologically selected samples will be more susceptible to environmental dependence of halo formation history at fixed mass. Conversely, samples with high luminosity or mass thresholds may be the least susceptible, since the correlation of formation time with environment is weak for high-mass halos. While the breakdown of the Bond et al. (1991) prediction is an annoyance for the cosmological applications that are the focus of this paper, it could be an asset for efforts to understand the physical processes that determine galaxy morphology and spectral properties. For example, if fitting the clustering of a selected class of galaxies requires environmental dependence of the HOD, then the form and magnitude of that dependence will provide strong clues to the aspects of halo formation that determine whether galaxies are members of that class. Conversely, if the color and morphology dependence of observed clustering can be entirely explained without environmental variations of the HOD, it will imply that aspects of halo formation that correlate with environment do not play a strong role in determining these properties of galaxies.

6. SUMMARY AND DISCUSSION

Our analysis shows that HOD modeling can substantially increase the cosmological power of galaxy clustering measurements, by breaking degeneracies between the clustering of dark matter and the bias of galaxies with respect to mass. Changing the shape or amplitude of the matter power spectrum or the value of \( \Omega_m \) alters the mass function, spatial clustering, and velocity statistics of the dark halo population in well-understood ways (Zheng et al. 2002). Our experiments here, which verify the qualitative arguments of Berlind & Weinberg (2002) and Zheng et al. (2002), show that changes to the galaxy HOD cannot mask these changes in the underlying dark halo population. With our highly flexible parameterization of the HOD, the set of observables considered here yields 1 \( \sigma \) uncertainties of \( \sim 10\% \) in \( \sigma_8, \Omega_m \), and \( \Gamma \) and \( \sim 5\% \) uncertainty in the combination \( \sigma_8 \Omega_m^q \) with \( q \sim 0.75 \). We expect these forecasts to be conservative, as we have not included observables for which we did not have ready analytic approximations, and our assumption of 10\% measurement errors is pessimistic in at least some cases.

The physical origin of these cosmological constraints is straightforward to understand for simple changes in \( \Omega_m, \sigma_8, \) or \( \sigma_8 \Omega_m^q \), as discussed in \$4.2–4.4. The general theme of these discussions is that, for a given cosmological model, the spatial clustering of galaxies largely determines the number of galaxies in halos of a given spatial abundance. Dynamically sensitive statistics then reveal the halo mass scale, which depends on \( \sigma_8 \) and \( \Omega_m \). We allow an arbitrary bias \( \alpha_t \) between the galaxy and dark matter velocity dispersions within halos, but this freedom does not eliminate the constraining power of dynamical observables because the space velocities of the halos themselves do not change. The main parameter degeneracy is approximately \( \sigma_8 \Omega_m^{0.75} \), because fixing this combination roughly fixes the halo velocity scale and the abundance of halos at the mass scale of rich galaxy groups. However, the changing shape of the mass function, the differing sensitivities of different velocity measures, and the different \( (\sigma_8, \Omega_m) \) dependence of galaxy-galaxy lensing all serve to break this degeneracy. Figures 14 and 15 demonstrate that the cosmological constraints emerge from the full web of clustering observables and are not dominated by one or two statistics on their own.

Constraints on the galaxy HOD will themselves provide valuable tests of galaxy formation models. The cutoff regime of \( \langle N(M) \rangle \) is difficult to pin down with the clustering statistics considered here, but for a known cosmological model the relation between average satellite number and halo mass is well determined, and the satellite distribution width \( \omega \), concentration \( c_\text{con} \), and velocity bias \( \alpha_v \) are measured to \( \sim 10\%, 30\% \), and \( 3\% \), respectively (Fig. 2). All of these quantities depend in detail on the physics that governs the evolution of satellites in larger halos (see, e.g., Taylor & Babul 2004, 2005a, 2005b; Zentner et al. 2005), while the relative mass of halos that host central and satellite galaxies depends on the efficiency with which halos feed baryonic mass to their central objects (see, e.g., discussions by Berlind et al. 2003; Zheng et al. 2005). These galaxy formation constraints will be especially powerful when derived as a function of luminosity, stellar mass, or other observables.

The two main assumptions built into our modeling are the central-satellite parameterization and environment independence of the HOD. The central-satellite distinction appears well rooted in galaxy formation physics, and it allows us to represent the range of plausible galaxy HODs more completely with a moderate number of parameters. However, we have confirmed with other tests that if we model galaxy bias with a flexible HOD parameterization that does not impose a central-satellite distinction, but instead introduces a characteristic mass for the narrow-to-wide transition of \( P(N|M) \) as in Scranton (2003), then we reach almost identical conclusions about the cosmological constraining power of the clustering observables considered in this paper. Recent numerical results imply an environmental dependence of halo formation times that opens the door to environmental variation of the HOD, especially in the single-galaxy regime. As discussed in \$5, we expect the quantitative impact of such dependence to be small, but potentially significant at the high level of precision we ultimately hope to attain. Investigation of environmental dependence effects and methods of allowing for them in HOD modeling are a high priority for future work.

The cosmological modeling approach advocated here is closely related to the CLF method introduced by Yang et al. (2003) and van den Bosch et al. (2003a), who use clustering data and the global galaxy luminosity function to constrain the dependence of the luminosity function on halo mass. In principle, the CLF and HOD methods are equivalent: they are merely differential and integral forms of one another. One can derive the CLF from a series of HOD fits to galaxy samples with different luminosity thresholds (Zehavi et al. 2005; Tinker et al. 2005). Conversely, one can integrate the CLF to infer \( \langle N(M) \rangle \) for galaxies above a luminosity threshold (Yang et al. 2003; van den Bosch et al. 2003b, 2005). The principal virtue of our HOD-based approach is that by focusing on a single, well-defined class of galaxies, we can parameterize the HOD in a way that seems likely to capture the predictions of any reasonable galaxy formation model. This kind of comprehensive parameterization is more difficult to achieve for the full CLF, and most analyses to date have assumed, for example, that the CLF has a Schechter form in halos of fixed mass. Nonetheless, it is valuable to pursue both HOD and CLF approaches and test for consistency of conclusions.
HOD modeling complements rather than replaces the perturbative approach based on large-scale measures that can be modeled with linear or quadratic bias. HOD modeling is more complex, but it can take advantage of high-precision clustering measurements on small and intermediate scales. HOD modeling can also amplify the power of the perturbative approach, extending its reach further into the nonlinear regime and checking its range of validity at a desired level of precision. For example, Tinker et al. (2006) show that an HOD-based approach to redshift-space distortions can improve recovery of the perturbative parameter $\beta$ that controls large-scale flows. Yoo et al. (2006) show that the linear bias model provides an accurate description of galaxy-galaxy lensing for $r \geq 2 \, \text{h}^{-1} \, \text{Mpc}$, and they show how to accurately model this phenomenon on smaller scales (see also Guzik & Seljak 2002; Mandelbaum et al. 2005). J. Yoo et al. (2007, in preparation) show that the scale-dependent bias factors derived by fitting the projected galaxy correlation function can extend recovery of the shape of the linear matter power spectrum into the mildly nonlinear regime.

CLF and HOD analyses of the 2dFGRS and SDSS redshift surveys have already produced a number of interesting results, even though they have considered only a fraction of the potential clustering observables. These results confirm, at least qualitatively, many of the basic predictions of current galaxy formation models, including the general form of the mean occupation function, the dependence of this function on luminosity, the existence of a minimum mass-to-light ratio in the halos of $\sim L^*$ galaxies, and the large gap between the minimum halo mass for central and satellite galaxies above a luminosity threshold, the sub-Poisson fluctuations of $P(N|M)$ that are a consequence of this gap, and the strong preference of galaxies with older stellar populations for higher mass halos (van den Bosch et al. 2003a, 2003b; Magliocchetti & Porciani 2003; Tinker et al. 2005; Yang et al. 2005; Zehavi et al. 2004, 2005; Collister & Lahav 2005). In combination with CMB data, the cosmological constraints from HOD modeling of the SDSS projected galaxy correlation function are almost as tight as those from the large-scale galaxy power spectrum, and the two analyses are consistent within their statistical uncertainties (Abazajian et al. 2005). For the most part, the cosmological inferences from CLF/HOD modeling of galaxy clustering agree with those from other methods, but matching the mass-to-light ratios of galaxy clusters simultaneously with other clustering data appears to require values of $\Omega_m$ and/or $\sigma_8$ that are substantially lower than the commonly adopted values of 0.3 and 0.9 (van den Bosch et al. 2003a; Tinker et al. 2005). If this conclusion is correct, then the evidence for it should become much stronger as more clustering observables are brought into play and the SDSS data set itself moves to completion.\(^5\)

We have focused in this paper on the cosmological parameter constraints that can be derived from galaxy clustering alone, using external data only to guide the choice of power spectrum shape and motivate the assumption of Gaussian initial conditions. As with perturbative analyses of large-scale structure, the long-term interest lies in combining these constraints with those from the CMB, the Lyman forest, Type Ia supernovae, and other cosmological observables. The complementary sensitivities of these observables lead to much tighter parameter constraints. More importantly, conflicts among them could point the way to physics beyond the simplest versions of $\Lambda$CDM, such as evolving dark energy, a gravitational wave contribution to CMB anisotropy, departures from scale invariance in the primordial power spectrum, nonzero space curvature, cosmologically significant neutrino masses, and so forth. By sharpening the constraints from large-scale structure in the new generation of galaxy redshift surveys, HOD modeling can play a critical role in efforts to test the standard cosmological model and, perhaps, discover its breaking points.

We thank Andreas Berlind, Jeremy Tinker, and Jaicyl Yoo for valuable discussions on these topics. We thank Barth Netterfield for suggesting a cubic spline parameterization of the mean occupation function, Sandy Faber for suggesting the influence matrix investigation, and Andy Gould for advice on the $\chi^2$ minimization method. This work was supported by NSF grants AST 00-98584 and AST 04-07125. Z. Z. acknowledges the support of NASA through Hubble Fellowship grant HF-01181.01-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS 5-26555. Z. Z. was also supported by a Presidential Fellowship from the Graduate School of the Ohio State University at an early stage of the project.

\(^5\) Since the original submission of this paper, the analysis of the three-year WMAP data set (Spergel et al. 2006) has provided strong support for this shift in cosmological parameter values, in excellent agreement with the results of the HOD and CLF modeling and the related “empirical model” method of Vale & Ostriker (2006).

**APPENDIX A**

PAIR DISTRIBUTIONS OF TWO POPULATIONS FOLLOWING NFW PROFILES WITH DIFFERENT CONCENTRATIONS

For the calculation of the one-halo term of the galaxy-galaxy or galaxy-mass two-point correlation function, we need to know the distribution of galaxy pairs or galaxy-mass particle pairs as a function of separation in each halo. The pair distribution is basically the convolution of the density profiles of two populations. For spherically symmetric density distributions $\rho_1$ and $\rho_2$, the fraction of pairs with separation in the range of $r$ to $r + dr$ is

$$\frac{dF}{dr} \, dr \propto \int \rho_1(x_1) \, d^3x_1 \, \rho_2(x_2 + r) \, d^3r$$

$$\propto \int 4\pi x_1^2 \rho_1(x_1) \, dx_1 \int \rho_2(x_2) 2\pi r^2 \, \sin \alpha \, d\alpha \, dr$$

$$\propto r^2 \, dr \int \left[ \frac{x_1}{x_1^2 + r^2 - 2x_1r\beta} \right] \rho_1(x_1) \int_{-1}^{1} \, d\beta \, \rho_2(x_2).$$

\[(A1)\]

where $F(r)$ is the fraction of pairs with separation less than $r$, $x_2 = x_1^2 + r^2 - 2x_1r\beta$, $\beta = -\cos \alpha$, and $\alpha$ is the angle between $x_1$ and $r$. If $\rho_1$ and $\rho_2$ follow the same NFW profile truncated at the virial radius, the analytic expression of $dF/dr$ can be found in Sheth &
Diaferio (2001) [see their eqs. (1) and (A25); note that \(\lambda(r)\) in their equations is proportional to \(r^{-2} dF/dr\)]. In this paper we assume that, like the dark matter, the distribution of galaxies inside halos also follows the NFW profile. Therefore, we use the formula in Sheth & Diaferio (2001) when evaluating the one-halo term of galaxy-galaxy correlations. However, we allow galaxies to follow a different NFW profile than the dark matter in halos of a given mass, which is our way to parameterize the spatial bias inside the halo in this paper. So, for the purpose of calculating the one-halo term of the galaxy-mass two-point correlation function, we need the pair distribution function corresponding to two NFW profiles with different concentration parameters. We evaluate the convolution in equation (A1) for two NFW profiles with concentration parameters \(c_1\) and \(c_2\), both truncated at the virial radius \(R_{\text{vir}}\) of the halo. After tedious algebra, we derive the following analytic expression for the (differential) pair distribution function \(dF(x; c_1, c_2)/dx\), where \(x = r/(2R_{\text{vir}})\).

For \(0 \leq x \leq 0.5\),

\[
\frac{dF(x; c_1, c_2)}{dx} = A(c_1, c_2)[f_1(x; c_1, c_2) + f_2(x; c_1, c_2) + f_3(x; c_1, c_2) + f_4(x; c_1, c_2)]s, \tag{A2}
\]

where

\[
f_1(x; c_1, c_2) = \frac{1}{(c_2 + c_1 + c_1c_2s)^2} \ln[(1 + c_1s)(1 + c_2s)] + \frac{c_1s}{c_2(c_2 + c_1 + c_1c_2s)(1 + c_1s)}, \tag{A3}
\]

\[
f_2(x; c_1, c_2) = \frac{1}{(c_2 - c_1 + c_1c_2s)^2} \ln\left[\frac{(1 + c_1s)(1 + c_2 - c_1s)}{1 + c_1}\right] - \frac{c_1(1 - s)}{c_2(c_2 - c_1 + c_1c_2s)(1 + c_1s)(1 + c_1)}, \tag{A4}
\]

\[
f_3(x; c_1, c_2) = \frac{1}{(c_2 - c_1 - c_1c_2s)^2} \ln\left[\frac{(1 + c_2s)(1 + c_1 - c_1s)}{1 + c_2}\right] + \frac{c_1(1 - s)}{c_2(c_2 - c_1 - c_1c_2s)(1 + c_1 - c_1s)}, \tag{A5}
\]

\[
f_4(x; c_1, c_2) = -\frac{s}{c_2(1 + c_1)(1 + c_2)(1 + c_1 - c_1s)}, \tag{A6}
\]

and \(s = r/R_{\text{vir}} = 2x\). For the purpose of stability, one should compute \(f_2\) using the limit value \([c_1^{-2} - c_2^{-2}(1 + c_1)^{-2}]\) as \(s\) approaches \((c_1 - c_2)/(c_1c_2)\) and similarly \(f_3 = [c_1^{-2}(1 + c_2)^{-2} - c_2^{-2}]\) as \(s\) approaches \((c_2 - c_1)/(c_1c_2)\).

For \(0.5 < x \leq 1\),

\[
\frac{dF(x; c_1, c_2)}{dx} = A(c_1, c_2)f(x; c_1, c_2)s, \tag{A7}
\]

where

\[
f(x; c_1, c_2) = \frac{1}{(c_2 + c_1 + c_1c_2s)^2} \ln\left[\frac{(1 + c_1)(1 + c_2)}{(1 - c_1 + c_1s)(1 - c_2 + c_2s)}\right] + \frac{s - 2}{(1 + c_1)(1 + c_2)(c_2 + c_1 + c_1c_2s)}. \tag{A8}
\]

The distribution is normalized so that \(\int_0^1 dx \frac{dF}{dx} = 1\). The normalization factor \(A(c_1, c_2)\) in equations (A2) and (A7) can be expressed as

\[
A(c_1, c_2) = \sqrt{A_s(c_1)A_s(c_2)}, \tag{A9}
\]

where we have the following fitting formula for \(A_s(c)\):

\[
A_s(c) = A_0 e^{\frac{c - 0.617}{1.651}} \left[1 + (c/c_T)^{16/4}\right]^{\log(\sin(\phi/\pi))} + B_0, \tag{A10}
\]

with \(A_0 = 2.4575\), \(\alpha = -3.099\), \(\beta = 0.617\), \(c_T = 1.651\), \(\mu = 4.706\), \(B_0 = 0.0336\), \(\omega = 2.684\), and \(\phi = 0.4079\). The fitting formula has a fractional error less than 0.2% for \(1 < c < 100\).

It can be shown that the pair distribution derived here (eqs. [A2] and [A7]) reduces to equation (A25) in Sheth & Diaferio (2001) for the case \(c_1 = c_2\). We also test the above formulae by generating random particle distributions following different NFW profiles and counting particle pairs directly. The Monte Carlo result agrees with the analytic formulae perfectly.

**APPENDIX B**

\(\chi^2\) AND MINIMIZATION

We use the Gauss-Newton method to perform the \(\chi^2\) minimization. Here we briefly describe the main procedure. Readers are referred to Gould (2003) for more details.

Assume that we have \(n\) observables \(F_k\) \((k = 1, 2, \ldots, n)\) predicted by a model with \(m\) parameters \(a_i\) \((i = 1, 2, \ldots, m)\). Observations give \(F_{k,\text{obs}}\). The \(\chi^2\) is defined as (Einstein’s summation convention is used)

\[
\chi^2 = \left[ F_k(a) - F_{k,\text{obs}} \right] B_{kk} \left[ F_i(a) - F_{i,\text{obs}} \right], \tag{B1}
\]
where $B_{ij}$ is the inverse of the covariance matrix $\sigma_i^2$ and has only diagonal components $B_{ii} = 1/\sigma_{\epsilon i}^2$ if observational errors are uncorrelated.

If the initial guess of the solution of $a$ is $a_*$, through linearizing $F_k(a)$ around $a_*$ and demanding that the first derivatives of $\chi^2$ are zero, we obtain an equation for the correction to the initial guess (Gould 2003),

$$b \Delta a = d,$$  \hspace{1cm} (B2)

where

$$b_{ij} = \left. \frac{\partial F_k}{\partial a_i} \right|_{a_*} \left. \frac{\partial F_l}{\partial a_j} \right|_{a_*} = \frac{1}{2} \left. \frac{\partial^2 \chi^2}{\partial a_i \partial a_j} \right|_{a_*},$$ \hspace{1cm} (B3)

and

$$d_i = \left. \frac{\partial F_k}{\partial a_i} \right|_{a_*} B_{kl} \left( F_{l, \text{obs}} - F_{l}(a_*) \right).$$ \hspace{1cm} (B4)

The direction of correction for $a_*$ is given by $\Delta a$. We set $a + \epsilon \Delta a$ as the new $a_*$, where $\epsilon$ is a small number, and refine the estimation of the solution of $a$. The above process is repeated until $\chi^2$ reaches a minimum. It can be proved that the covariance matrix of $a$ is just $b^{-1}$ (Gould 2003).

If some parameters have additional constraints, we introduce a cost function to realize the constraints and minimize the sum of $\chi^2$ and the cost function. For example, if parameter $a_1$ cannot exceed $a_2$, the cost function can be defined as $C = [(a_1 - a_2)/\sigma]^2$ for $a_1 > a_2$ and $C = 0$ for $a_1 \leq a_2$, where $\sigma$ is the tolerance. Another example is that a parameter has some priors. If $a_1$ is assigned a prior of a Gaussian distribution with mean $\tilde{a}_1$ and standard deviation $\sigma$, the cost function can be defined as $C = [(a_1 - \tilde{a}_1)/\sigma]^2$. Including a cost function leads to additional terms, $-\partial C/\partial a_i/2$, in $d_i$ (eq. [B4]).
