MARKED STATISTICS AND THE ENVIRONMENTAL DEPENDENCE OF GALAXY FORMATION

by

Ramin A. Skibba

B.S. in Physics, University of Notre Dame, 2000

B.A. in Philosophy, University of Notre Dame, 2000

M.S. in Physics & Astronomy, University of Pittsburgh, 2002

Submitted to the Graduate Faculty of

the Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2006

UNIVERSITY OF PITTSBURGH

ARTS AND SCIENCES

This dissertation was presented

by

Ramin A. Skibba

It was defended on

24 July 2006

and approved by

Ravi K. Sheth

Andrew J. Connolly

Bhuvnesh Jain

David Turnshek

Vittorio Paolone

Dissertation Advisors: Ravi K. Sheth,

Andrew J. Connolly

MARKED STATISTICS AND THE ENVIRONMENTAL DEPENDENCE OF GALAXY FORMATION

Ramin A. Skibba, PhD

University of Pittsburgh, 2006

Many have used the two-point correlation function to study the clustering of galaxies as a function of their properties, such as luminosity, color, and stellar mass. We explore the technique of 'marked' correlations, in which clustering is measured with galaxies weighted by a particular property or 'mark'. Marked clustering statistics identify and quantify how galaxy properties are correlated with their environment. We present marked correlation analyses in the framework of the dark matter halo model, in which all environmental correlations are due to the correlation of the masses and formation histories of halos with their environment. We perform marked correlation analyses of galaxy luminosity, color, stellar mass, metallicity, and star formation rate in the Sloan Digital Sky Survey and the Millennium Run Simulation. We also analyze luminosity-marked correlations of galaxies in groups and clusters. Our measurements show that luminous, red, massive, metal-rich, and passively star forming galaxies tend to be located in denser environments than fainter, bluer, less massive, metal-poor, and actively star forming galaxies. Our marked correlation measurements also show how these environmental correlations vary as a function of scale. Our halo-model analyses show that the environmental dependence of luminosity and stellar mass of SDSS galaxies is primarily driven by the environmental dependence of halo mass.

Keywords: methods: analytical - methods: statistical - galaxies: formation - galaxies: evolution - galaxies: clustering - galaxies: halos - dark matter - large scale structure of the universe.

TABLE OF CONTENTS

LIST OF FIGURES

PREFACE

I could not have accomplished this Ph.D. thesis without the help and friendly support I received from so many people along the way. I would first like to thank my adviser and committee, and my fellow grad students and postdocs: Ummi Abbas, Ryan Scranton, Simon Krughoff, Brigitte Koenig, Cameron McBride, Ching-Wa Yip, Sam Schmidt, Lorenzo Rimoldini, and Graziano Rossi. For help of a less research-related kind, I thank Kiplin Perkins, Dan Nestor, Doug Hackworth, Lynne Kuemmel, Erin Venter, Arie Baratt, Adam Drake, and Shane Corning. I also want to thank my activist friends, especially Sarah Rogers, Nick Marritz, Anthony Faino, Lee Szymkiewicz, Nicole Freund, Matt Richards, and Dean Gerber, and my sociology friends, especially Ashley Currier, Maria Jose Alvarez, and Kai Heidemann. I would also like to thank Joe Serkoch, Lauren DeMichiei, and Beau DeMonte. Of course I also thank my parents and my brother, my college friends, and my two oldest friends. Finally, I would like to thank the Thomas Merton Center, for all the support they give to so many groups struggling for peace and social justice, many of which I participated in before I began my thesis, the Shadow Lounge, where I heard a variety of inspiring bands and musicians, Kiva Han and Te Cafe, which I frequented while writing my thesis, and the Squirrel Hill Cafe, which I frequented prior to, during, and subsequent to writing my thesis.

1.0 INTRODUCTION

> I am convinced there is only *one* way to eliminate these grave evils, namely through the establishment of a socialist economy, accompanied by an educational system which would be oriented toward social goals. ...The education of the individual, in addition to promoting his own innate abilities, would attempt to develop in him a sense of responsibility for his fellow men in place of the glorification of power and success in our present society.
> –Albert Einstein, "Why Socialism?", May 1949

1.1 THE BASIC IDEA

Being ambitiously curious astrophysicists, we want to understand the processes through which all kinds of galaxies form and evolve, what dark matter halos are and how they form and evolve, and how galaxies and halos interact and are both shaped by their large-scale environments. Our basic picture is that in the beginning, or rather, when the universe was still quite young, primordial density fluctuations were set in place. Matter gradually gravitates towards the initial density peaks and collapses to form structure. When an amorphous aggregation of matter has accreted enough dark matter and the particles have reached equilibrium, the object has become a 'halo'. With 'cold' dark matter, structure forms hierarchically, with small dark matter halos merging to form larger ones, which merge to form still larger halos.

To give an example of what dark matter halos look like, we show a slice from a dark matter simulation, the Millennium Run Simulation, in Figure 1.1 (Springel et al. 2005). The dark matter density field is shown, color-coded by density. The figure zooms in on a very massive halo, which typically host large galaxy clusters. At very large scales, the dark matter density field is nearly homogeneous and isotropic. In the zoomed-in panels one can note a range of densities, with knots, filaments, and voids of various size. The most

massive halos tend to reside in the densest environments, in the largest knots and filaments, while less massive halos tend to ride in less dense regions, in the voids. All halos contain substructure, which are the remnants of usually smaller dark matter halos that merged in to them at earlier times.

All galaxies form simultaneously within dark matter halos. Galaxies are much smaller and more concentrated than the halos that host them. The least massive halos associated with galaxies host dwarf galaxies. More massive halos typically contain spiral or elliptical galaxies, and the rare most massive halos contain galaxy groups and clusters. Halo formation is entirely determined by gravitational forces and the cosmological expansion of the universe. Galaxy formation is much more complex, as many additional physical processes are involved, such as gas infall, star formation, supernovae feedback, disk dynamics, galaxy mergers, black hole feedback, ram-pressure stripping, and others. It is a very interesting and important question as to what extent various aspects of galaxy formation and evolution are determined by the large-scale environments of the galaxies' host halos, and in what ways they are shaped by group- and cluster-specific processes. These are the types of issues we are examining.

1.2 THE DARK MATTER HALO MODEL

Because dark matter particles are non-relativistic and weakly interacting, dark matter structures form hierarchically from the initial density perturbations. As a result, there is a correlation between the formation and abundances of halos and large-scale structure. In other words, dark matter halo properties, especially their masses, are correlated with their environments: less massive halos tend to form in underdense regions and more massive halos tend to form in overdense regions.

Observational studies of galaxy surveys have shown that galaxy properties are similarly correlated with their environments: luminous galaxies tend to be more clustered than faint ones, higher stellar mass galaxies tend to be more clustered more than less massive ones, redder galaxies tend to be clustered more than bluer ones, passively star-forming galaxies tend to be clustered more than actively star-forming ones, elliptical galaxies tend to be

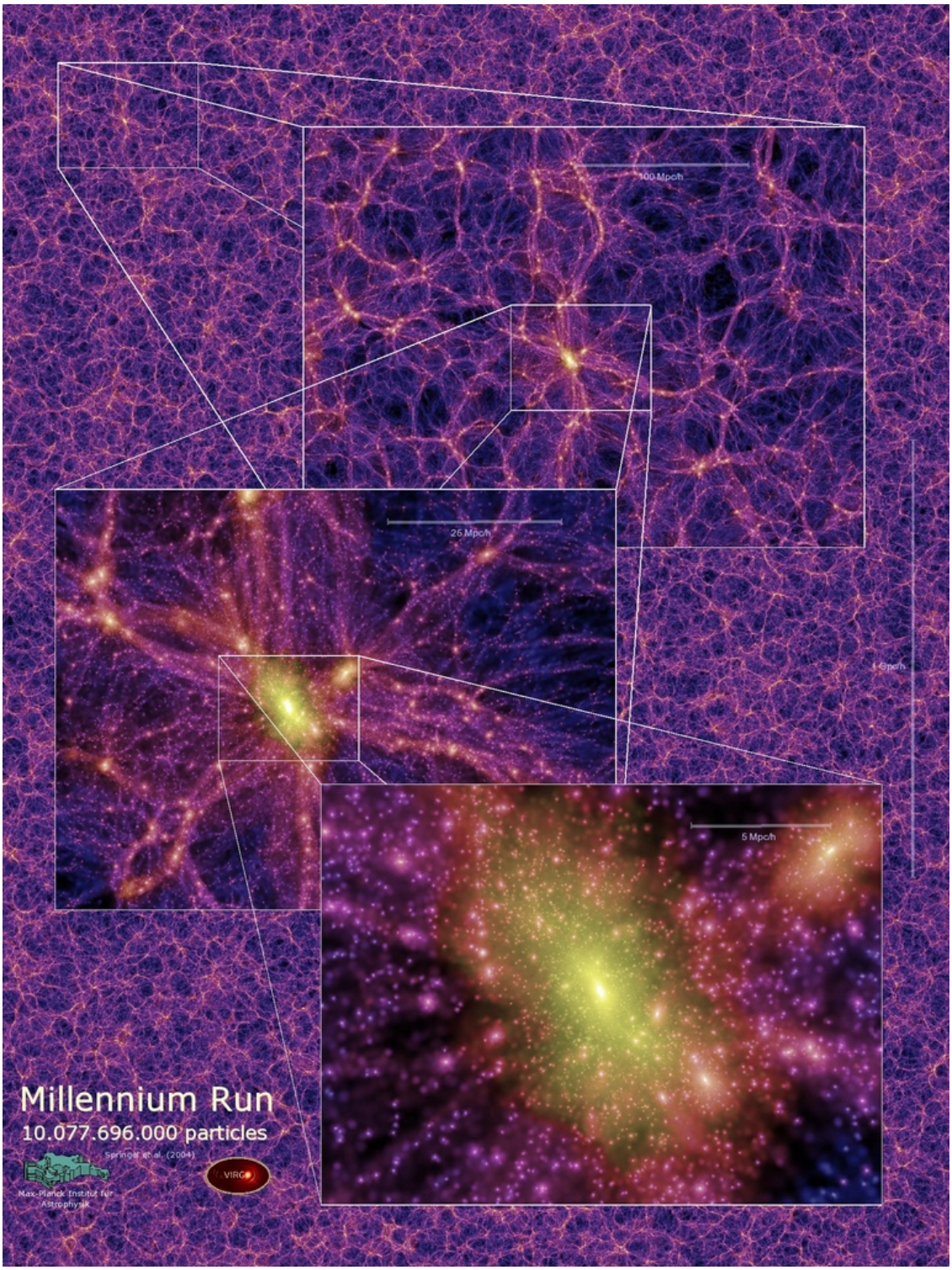

Figure 1.1: The dark matter density field at various scales in the Millennium Simulation (Springel et al. 2005). Each image shows the dark matter density field in a slice of thickness $15\,h^{-1}$Mpc, color-coded by density and local dark matter velocity dispersion. The panels zoom in by factors of four, centered on one of the many massive halos in the simulation.

clustered more than spirals, etc. Based on such evidence, galaxy formation models assume that the properties of a galaxy are determined entirely by the mass and formation history of the halo within which it formed. This would entail that the correlation between halo properties and environment induces correlations between galaxy properties and environment. We want to figure to what extent this is the case, and we want to use this information to learn about galaxy formation processes.

We are doing so by using a novel and versatile set of tools known as marked statistics. We are modeling marked correlations analytically with the halo model and comparing these predictions to measurements we are making with data from galaxy surveys and simulations. Such comparisons allow us to test our assumptions and refine our models, and in the process we are learning about poorly understood aspects of galaxy formation and evolution.

Before proceeding, it is worth acknowledging that it is logically possible that cold dark matter does not exist, at least in the massive quantities we expect. It may turn out that the consensus within the astrophysics community is wrong. Alternative theories involving a modification to the large-scale Newtonian gravitational force law are currently being developed, and they can explain some contentious observations as well as by cold dark matter theories. There are a variety of observations that clearly cannot be explained by baryonic matter and Newtonian gravity alone, such as those of spiral galaxy rotation curves, elliptical galaxy rotation curves, and the discrepancy between the dynamical masses of galaxy clusters and their masses observed with gravitational lensing. It follows that the assumptions of at least one of these is wrong. Either matter consists of more than just baryonic matter, or Newtonian gravitational forces are modified in a particular way on large scales. Both sets of theories have outstanding issues, such such as the lack of a steep central density cusp in some galaxies, which is problematic for dark matter theories, and the mass discrepancy of clusters, which is problematic for MOND theories. With more precise data, observations of galaxy and halo shapes, and observations of the amplitude of power spectrum peaks of the cosmic microwave background radiation (CMB) could potentially distinguish between the theories (Sellwood & Kosowsky 2002).

In any case, even if the theories of modified gravity turn out to be true, the basic halo model framework would still be valid. The most massive and luminous galaxy structures

clearly form in particular types of environments that are distinctly different than those of less massive and fainter galaxies. This is evidently not a coincidence, regardless of whether large-scale structure is made of dark matter or governed by modified gravitational forces.

1.3 THE CASE FOR STUDYING MARKED GALAXY CLUSTERING

We are accumulating an increasing wealth of data in galaxy surveys such as the Sloan Digital Sky Survey and in semi-analytic simulations such as the Millennium Simulation. The large number of galaxies, with multi-band photometry, enables us to employ sensitive statistical tools to probe poorly understood aspects of galaxy formation. Marked statistics, which measure the clustering of galaxies weighted by particular galaxy properties, are ideal for this task.

In most analyses, a galaxy catalog is cut into subsamples based on the mark, such as luminosity or morphological type, and the clustering of the subsample is studied by treating each galaxy in it equally. The galaxy clustering is quantified by n-point correlation functions, but the galaxies are effectively marked as 'ones', and the information about their properties–the galaxy marks–are discarded. Moreover, the choices of critical thresholds for the data cuts are somewhat arbitrary, especially in the case of marks with a continuous range of values, like luminosity and color. With the new large surveys and simulations, many of these marks can be measured with sufficient precision such that *marked* correlation functions and other marked statistics can be measured with little uncertainty. There is much we can learn by exploiting the information provided by weighting each galaxy by its mark.

Marked statistics are powerful tools we can use to quantify the environmental dependence of galaxy properties for a continuous range of density, obviating the difficulties of splitting galaxies into 'field' and 'cluster' subsamples. Since objects that are more highly clustered than others exist in denser environments, we know from past studies that luminous galaxies, elliptical galaxies, red galaxies, and passively star-forming galaxies are typically found in denser regions than faint, spiral, blue, actively star-forming galaxies. We can learn much more from the marked clustering of galaxies, using their luminosities, colors, morphologies,

and star formation rates as marks. We can robustly quantify the environmental dependence of such galaxy properties, and we can consequently better understand their interconnected roles in the formation of galaxies.

Our marked statistics research consists of developing observationally-motivated analytic models, and then contrasting the predicted marked correlations with measurements we make from various datasets. This in turn allows us to improve our theories and our understanding of the physical processes behind the formation of small-scale and large-scale structure. By using the halo model framework of marked statistics, we are better understanding how many galaxy and halo properties are correlated with their environments, which yield information about and constraints on galaxy formation processes.

There are many interesting applications. Marked statistics research allows us to constrain halo occupation distribution models, which describe the statistical relationship between halo mass and its probability to form galaxies, thus characterizing the mass dependence of the efficiency of galaxy formation.. Marked and unmarked clustering measurements can also be used to place constraints on cosmological parameters such as Ω_m and σ_8, as well as $\Delta_{\rm vir}$, the overdensity required for dark matter to collapse and form halos. We can use marked statistics to constrain and test assumptions of semi-analytic models and other simulations of galaxy formation. We can also use them to understand the limitations and biases of particular datasets. With redshift-space marked correlations and with the help of mock galaxy catalogs, we can learn more about redshift distortions and improve our models of their effects on galaxy clustering. Marked statistics also teach us about the distinct populations of 'central' and 'satellite galaxies' and their different properties and formation histories. We can also improve our analytic models in new ways, such as with marked correlation measurements of galaxy clusters, which teach us about the position-dependence of satellite galaxy properties within halos. I am certain many other useful applications will be found as well.

1.4 THE GOALS AND FORMAT OF THIS THESIS

The main purpose of my thesis is to show how we have used and further developed the halo model framework of marked statistics, and to explain what we have learned by applying them to a variety of interesting data sources.

My thesis will proceed as follows. In the next chapter, I will discuss how dark matter accretes from density perturbations, and how this accretion ultimately leads to the apparently universal density profiles we observe today. In Chapter 3, I will explain how we used the halo mass function from simulations with observed luminosity functions and halo occupation distributions to determine the relationship between halo mass and galaxy luminosity. Chapter 4 consists of a description of the halo model components that allow us to model galaxy clustering and marked clustering with two-point correlation functions for luminosity thresholds and ranges. Chapter 5 explains how we have modeled luminosity-marked correlations in real space and redshift space and how we have applied them to Sloan Digital Sky Survey galaxy samples and SDSS-based mock catalogs. We also show some color-marked correlations measurements with this data. In Chapter 6, I explain how we applied luminosity-marked correlations to Millennium Run Simulation galaxies, and compare stellar mass- and SFR-marked correlation functions of the halo model, SDSS, and Millennium Run. I briefly show a simple test of the effect of halo formation time and assembly bias on the halo model of the galaxy correlation function in Chapter 7. In Chapter 8, which is a work-in-progress, I show our applications of luminosity-marked correlations to SDSS catalogs of galaxy groups and clusters. In Chapter 9, I explain how we will measure luminosity- and color-marked angular correlation functions of galaxies at higher redshifts, in the GOODS and GEMS surveys, and I will sketch how we will model these marked correlations. Finally, I end with a chapter discussing our results and some ideas for future research on related topics.

The research I present in this thesis was done with my adviser, Ravi Sheth. Chapter 5 was also done with Andrew Connolly and Ryan Scranton, and Chapter 9, for which research has just begun, was also done with Andrew Connolly.

Throughout this thesis we adopt a flat ΛCDM cosmology with Ω_m=0.3, Ω_Λ=0.7, σ_8=0.9, $h \equiv H_0/100$ km s^{-1} Mpc^{-1} = 1, and $\Delta_{\rm vir}$=200, except when specified otherwise. None of

our results are qualitatively affected by the exact choice of cosmology, however. That is, the precise choices of cosmological parameters are irrelevant to the reasoning behind our analyses.

2.0 DARK MATTER ACCRETION & HALO DENSITY PROFILES

2.1 INTRODUCTION

Following the pioneering work of Gott (1975) and Gunn (1977), secondary infall models of the structure of spherically symmetric dark matter halos have been studied extensively (Fillmore & Goldreich 1984; Bertschinger 1985; Hoffman & Shaham 1985). In these models, bound objects form from initially overdense regions. Because they are overdense, these regions expand slightly slower than the background universe. They eventually reach a maximum size, after which they turn-around and collapse. Energy conservation arguments show that the turn-around time and radius depend on the initial density run around the center of the perturbation. Work has concentrated on obtaining solutions to this problem which are self-similar: therefore, the analytic results referred to above assume that the initial density run was a power-law in distance from the center: as a result, the mass of the bound object increases as a power-law in time, and the final density profile of the object, which we will call a dark matter halo, is also a power-law.

Numerical simulations of halo formation in CDM dominated universes show that the final density run is not a scale-free power-law (Navarro, Frenk & White 1996). Although there is some discussion as to the exact form of the profile (e.g. Moore et al. 1999), all workers agree that the density run has a characteristic scale. Navarro et al. (1996) also argued that it was better to think of the scale as representing a density, and showed that this density correlated reasonably well with the density of the universe at the epoch when the object 'formed', provided one used a well-defined but ad hoc definition of formation. But they did not relate this formation time to the secondary infall models.

Some have worried that the analytic models fail to produce this characteristic scale

because they assume that mass is accreted smoothly, whereas the simulations show that the opposite is the norm. Nusser & Sheth (1999) studied a model in which the mass of the object grew in jumps—the distribution of jumps chosen to mimic that expected in models of hierarchical clustering from Gaussian initial conditions (a similar approach was taken by Wechsler et al. 2002). That substructure is unlikely to be the main factor determining the halo profile was shown by Moore et al. (1999). He presented results from numerical simulations in which small scale power had been removed, so the mass growth rate was substantially smoother—the characteristic scale in the final profile persisted.

If substructure is unimportant, then perhaps it is the assumption of a power-law accretion rate which is flawed. Departures from power-law accretion are expected for two reasons. First, the CDM fluctuation spectrum is not scale free, so the slope of the power-law which describes the mass accretion rate is actually a function of time, and so perhaps this is what sets the scale of the final profile. However, characteristic scales are also seen in the density profiles of halos which form from scale-free initial conditions (Navarro et al. 1996; Eke et al. 2001) so this cannot be the cause. Departures from power-law accretion are also expected if the halos form from peaks in the initial fluctuation field, because the density run around peaks is not a power-law, even if the initial fluctuation field is scale-free (Bardeen et al. 1986). Indeed, the mass growth rate in numerical simulations of halo formation in a ΛCDM universe is substantially better fit by an exponential, than by a power-law in time (Wechsler et al. 2002). The main purpose of this paper is to present a formula for what happens in secondary infall models if one continues to assume that mass is accreted smoothly and spherically, but that the mass accretion decreases exponentially with time.

2.2 A FIRST APPROXIMATION

Energy conservation arguments suggest that if R_i and Δ_i denote the initial size and over-density of the spherical perturbation, then the turnaround size and time are

$$\frac{R_{\rm ta}}{R_i} = \Delta_i^{-1} \qquad \text{and} \qquad \frac{t_{\rm ta}}{t_i} = \frac{(3/4\pi)}{\Delta_i^{3/2}}. \tag{2.1}$$

The power-law accretion model sets $\Delta_i = (R_i/R_0)^{-3-n}$. So it has $(R_{\rm ta}/R_i) = (R_i/R_0)^{3+n}$ and $(t_{\rm ta}/t_i) = (R_i/R_0)^{3(3+n)/2}$. In an Einstein de-Sitter model, the expansion factor scales as $a \propto t^{2/3}$, so that $(a_{\rm ta}/a_i) = (R_i/R_0)^{3+n}$. The mass at turn-around is $M_{\rm ta} = (4\pi/3)\, R_{\rm ta}^3\, \rho_b(t_{\rm ta}) = (4\pi/3)\, R_{\rm ta}^3\, \rho_b(t_i)\, (a_i/a_{\rm ta})^3$ so it is $(4\pi/3)\, R_{\rm ta}^3\, \rho_b(t_i)\, (R_i/R_0)^{-3(3+n)}$. Therefore, at turnaround, the density within the turn-around radius is $3M_{\rm ta}/4\pi R_{\rm ta}^3 = \rho_b(t_i)\, (R_{\rm ta}/R_0)^{\gamma}$, where the fact that $(R_{\rm ta}/R_0) = (R_i/R_0)^{4+n}$ implies $\gamma \equiv 3(3+n)/(4+n)$.

Following Gunn (1977), a simple model for the density run of the final object follows from assuming that after turning around, the final scale associated with the shell is proportional to the turn-around radius: $R_{\rm f} = f R_{\rm ta}$ with the constant of proportionality being independent of $R_{\rm ta}$. In this case, the density run at a later time is denser than at turnaround by a factor of f^{-3}, but the shape of the density run is unchanged. Thus, in this model, the final density run is a power-law. Fillmore and Goldreich (1984) note that when the initial slope is shallow, i.e., if $3 + n \leq 2$, then the assumption that $R_{\rm f} = f R_{\rm ta}$ with f independent of $R_{\rm ta}$ is not accurate.

Next, we consider initial overdensity profiles which are not simple power-laws. As a first example, consider

$$\Delta_i = \left(\frac{R_0}{R_i}\right)^{\epsilon} \left[D_0 - \left(\frac{R_i}{R_0}\right)^{1+\epsilon}\right] \text{ where } D_0 = \left(\frac{3M_0}{4\pi R_0^3\, \rho_i}\right)^{(1+\epsilon)/3}. \tag{2.2}$$

where $0 \leq \epsilon \leq 3$, and we set $\Delta_i = 0$ if the term in square brackets is negative. This means that the perturbation is finite in size, and contains total mass M_0. Inserting this in the secondary infall model formula (2.1) gives $R_{\rm ta}$ as a function of R_i. This relation can be inverted to obtain an analytic expression for R_i as a function of $R_{\rm ta}$. Thus, at turnaround, the mass within the turnaround radius, is

$$M_{\rm ta}(< R_{\rm ta}) = 4\pi R_i^3\, \rho_i = M_0 \left(\frac{R_{\rm ta}/R_0}{1 + R_{\rm ta}/R_0}\right)^{3/(1+\epsilon)}. \tag{2.3}$$

If $2 \leq \epsilon \leq 3$ it should be a reasonable approximation to assume that the final radius of the shell is a fixed fraction of the turnaround radius: $R_{\rm f} = f R_{\rm ta}$. If so, then the expression above

shows how the enclosed mass varies as a function of radius. Differentiating with respect to $R_{\rm f}$ and then dividing by $4\pi R_{\rm f}^2$ gives the shape of the final density profile:

$$\rho(R_{\rm f}) = \frac{3M_0/(4\pi R_0^3)}{(1+\epsilon)} \, \frac{x^{-3\epsilon/(1+\epsilon)}}{(1+x)^{1+3/(1+\epsilon)}}, \qquad \text{where } x = fR_{\rm ta}/R_0. \tag{2.4}$$

This scales as $x^{-3\epsilon/(1+\epsilon)}$ at small x, and falls as x^{-4} on larger scales. When $\epsilon = 1/2$ this profile has the same form as one first studied by Hernquist (1990). However, in this case ϵ is considerably less than the critical value 2 on small scales, so that the assumption $R_{\rm f} = fR_{\rm ta}$ is suspect. Nevertheless, this is a simple example showing how a feature in shape of the initial profile translates to a feature in the final profile.

Suppose instead that the initial density run is $\Delta_i = \Delta_0 \, \exp(-R_i/R_0)$. Then $(R_{\rm ta}/R_i) = \exp(R_i/R_0)/\Delta_0$ and $(a_{\rm ta}/a_i) = \exp(R_i/R_0)/\Delta_0$. The mass at turn-around is $M_{\rm ta} = (4\pi/3)\, R_{\rm ta}^3 \, \rho_b(t_i) \, \exp(3R_i/R_0)/\Delta_0^3$, and the density within the turn-around radius is $\rho_b(t_i) \, \exp(3R_i/R_0)/\Delta_0^3$ which we can also write as $\rho_b(t_i) \, (R_i/R_{\rm ta})^3$. To finish the calculation, we would like to write R_i explicitly as a function of $R_{\rm ta}$. In the case of the pure and broken power-laws above, this was trivial. Here the relation is a little more complicated.

If we set $r = (R_i/R_0)$ and $x = \Delta_0 \, (R_{\rm ta}/R_0)$, then we have a relation of the form $r = x \, \exp(-r)$. Lagrange's theorem on the inversion of series shows that if

$$r = x \, \exp(-r) \qquad \text{then} \qquad r = \sum_{n=1}^{\infty} \frac{(-n)^{n-1}}{n!} x^n \equiv \text{ProductLog}(x).$$

If we assume that the final radius is proportional to $R_{\rm ta}$ as before, then x can be thought of the final radius divided by a characteristic scale. The enclosed mass, written as a function of this scaled distance $x = \Delta_0 \, (R_{\rm ta}/R_0)$, is

$$\frac{M(<x)}{(4\pi/3)\, \rho_b(t_i) \, R_0^3} = \left(\sum_{n=1}^{\infty} \frac{(-n)^{n-1}}{n!} x^n \right)^3. \tag{2.5}$$

The corresponding expression for the enclosed mass within the Navarro et al. (1996) profile is

$$M_{\rm NFW}(<x) \propto \ln(1+x) - \frac{x}{1+x} = \sum_{n>1} (n-1) \, \frac{(-x)^n}{n}, \tag{2.6}$$

so it is also an alternating series in the scaled distance x. The profile advocated by Moore et al. (1999) has $M(<x) \propto (2/3) \, \ln(1+x^{3/2})$ which is also an alternating series.

2.3 FROM ACCRETION RATE TO VELOCITY DISPERSION

As stated in the previous section, the power-law accretion model sets $\Delta_i(R) = (R/R_0)^{-(3+n)}$. Assuming an essentially flat Friedmann universe, then if each spherical shell of mass is kept at its turnaround radius, this implies that the density within the turn-around radius of the entire object is

$$\rho_{\rm PL}(R) = \frac{M_{\rm ta}}{(4\pi/3)\,R_{\rm ta}^3} = \rho_b(t_i)\,\left(\frac{R_{\rm ta}}{R_0}\right)^{\gamma}, \tag{2.7}$$

where $\gamma \equiv 3(3+n)/(4+n)$ and R_0 is the maximal expansion radius. Then

$$\frac{\rho_{\rm PL}(r)}{\rho_{\rm PL}(r_v)} = \left[\left(\frac{R_v}{R}\right)^{3+n}\right]^3 = \left(\frac{r_v}{r}\right)^{\gamma}, \tag{2.8}$$

where $r = R/R_0$ and r_v is the virial radius. In terms of the scaled radius $x = r/r_v$, the density is given by $\rho_{\rm PL}(x) = \rho_v\, x^{-\gamma}$. It follows that the enclosed mass is

$$M_{\rm PL}(<x) = \int_0^x dx\; 4\pi x^2\, \rho_{\rm PL}(x) \propto x^{3-\gamma}. \tag{2.9}$$

The virial radius is the radius of the halo within which virial equilibrium has been reached and within which the mean density is $18\pi^2 \approx 178$ times the critical density. This is obtained by expressing the time evolution of a mass shell in parametric form. The relation between the proper radius and time is: $R/R_i = A\,(1 - e\,cos\theta)$ and $(t+T)/t_i = B\,(\theta - e\,sin\theta)$, where $A^3 = GMB^2$ comes from $\ddot{r} = -GM/r^2$, T is small compared to t_i, and $e = 1$ in the spherical model. Then at the turnaround radius ($\theta = \pi$), the density relative to the background is $1 + \Delta_{\rm ta} = (A/2(R/R_i))^3\,(6(t/t_i)/B)^2 = 9\pi^2/16$. By the time the halo has collapsed to the virial radius ($\theta \to 2\pi$), the density relative to the background is $1 + \Delta_{\rm vir} = 18\pi^2$, which is 32 times larger (Lacey and Cole 1993).

Some have argued that the mean virialization overdensity is not constant but rather varies with redshift: $1 + \Delta_{\rm vir} \cong (18\pi^2 + 82x - 39x^2)/(1+x)$, where $x(z) \equiv \Omega(z) - 1 = \Omega_m(1+z)^3/E(z)^2 - 1$, for $\Omega_R = 0$ (Bryan & Norman 1998). Using the flat ΛCDM model with the WMAP results of $\Omega_m = 0.27 \pm 0.04$ and $\Omega_\Lambda = 0.73 \pm 0.04$ (Spergel et al. 2003), then the mean overdensity $1 + \Delta_{\rm vir}$ varies from $18\pi^2$ at $z \gg 1$ to ≈ 207 at $z = 1$ to ≈ 359 at $z = 0$.

Nonetheless, since many dark matter halos formed at large redshift, the approximation of $1 + \Delta_{\rm vir} \approx 18\pi^2$ is usually sufficient.

Each particle of mass actually spends most, not all, of its orbital time near its apapsis. Particles in the outer shells spend a small fraction of time passing through inner shells. A more accurate collapse model would have to account for this "secondary infall." When calculating the mass within a particular radius, one must account for the effects of these "extra" particles that are just passing through.

Let $m_{\rm add}(r)$ denote the mass due to particles with orbital apapses that are at large radii which spend some of their time within r. A particle on an orbit that carries it out to a maximum distance $r' > r$ from the center spends some fraction of the time within r. If we assume that all orbits are radial, then the fraction of time that a particle on an orbit with maximum radius r' spends within radius r is

$$P(r|r') = \int_0^r \frac{dr}{v(r)} \Bigg/ \int_0^{r'} \frac{dr}{v(r)}. \tag{2.10}$$

The velocity $v(r)$ is obtained by integrating the equation of motion of the shell: $dr^2/d^2t = -GM_{\rm in}(r)/r^2$, where $M_{\rm in}(r)$ is the mass of the particles that have apapses that are less than or equal to r, so this does not include the additional mass $m_{\rm add}(r)$ of the particles with apapses in the outer shells. Then in the case of the power-law model, the result is

$$v^2_{\rm PL}(r) = \frac{GM_{\rm in}(r)}{r} - \frac{GM_{\rm in}(r')}{r'} = \frac{GM_{\rm in}(r')}{r'} \left(\frac{M_{\rm in}(r)}{M_{\rm in}(r')} \frac{r'}{r} - 1 \right). \tag{2.11}$$

Substituting for $M_{\rm in}(r) = M_{\rm PL}(r) \propto r^{3-\gamma}$ yields

$$P_{\rm PL}(u) = \int_0^u \frac{du}{\sqrt{u^{2-\gamma} - 1}} \Bigg/ \int_0^1 \frac{du}{\sqrt{u^{2-\gamma} - 1}}, \tag{2.12}$$

where $u \equiv r/r'$. The contribution to the mass from particles just passing through r is obtained by taking the total mass of particles with apapses of r', multiplying by the fraction of time these particles spend within r, and integrating over all $r' > r$ out to the turnaround radius R_{max}:

$$\frac{m_{\rm add}(r)}{M_{\rm in}(r)} = \frac{4\pi \int_r^{R_{max}} dr' r'^2 \, \rho_{\rm in}(r') \, P(r|r')}{4\pi \int_0^r dr' r'^2 \, \rho_{\rm in}(r')} = (3-\gamma) \int_1^{R_{max}/r} dw \, w^{2-\gamma} \, P(1/w). \tag{2.13}$$

The additional effective mass in the inner regions of the object means that the outer shells actually contract a little more than they would have otherwise. This is a significant effect, so this means that the assumption of a constant collapse factor, independent of turnaround radius, must be modified.

If the orbital period of the inner shell is much smaller than the collapse time of the outer shells, then the dynamics of the inner shell admits an adiabatic invariant. For radial orbits, this implies that $rM(r) =$ constant. Then the change of enclosed mass due to secondary infall is associated with an adiabatic change in radius. The collapse factor f describes this change in mass: it is given by the ratio of the original enclosed mass $M_{\rm in}(r')$ to the final mass $M(fr') = M_{\rm in}(r) + m_{\rm add}(r)$. Then the collapse factor is given by

$$f(r) = \frac{M_{\rm in}(r')}{M_{\rm in}(r) + m_{\rm add}(r)}. \tag{2.14}$$

Consequently, the radii shrink from $r' \to r = r'\, f(r')$. In the power-law case, the collapse factor is

$$f_{\rm PL}(r) = \frac{1}{1 + (3-\gamma)\int_1^{R_{max}/r} dw\, w^{2-\gamma}\, P(1/w)}. \tag{2.15}$$

As a result of secondary infall, the velocity dispersions are also modified by the collapse factor:

$$v^2 = \frac{GM(r)}{r} = \frac{G\, r' M(r')}{f(r')^2\, r'^2} = \frac{G\, M(r')}{f(r')^2\, r'}. \tag{2.16}$$

Figure 2.1 shows the collapse factor as a function of radius for different values of γ.

It is useful to compare the mass and circular velocity profiles resulting from the accretion rate to those of the final density profile, to test whether it is consistent with an NFW profile (Navarro et al. 1996). Prior to making such comparisons, one must account for the contraction of the radii, as a result of secondary infall. This is accomplished in two steps: firstly, the total mass distribution is calculated by adding the "additional" mass $m_{\rm add}(r)$ due to secondary infall to the "initial" mass $M_{\rm in}(r)$. The sum of the masses is the mass enclosed within the radius $r\, f(r)$. Secondly, the condition must be satisfied that the mass enclosed within the virial radius remains constant. That is, $M_{\rm tot}(r_{\rm vir})$ for the accretion rate is set to be equal to the mass in the final density profile. Then the mass distributions out to the turnaround radius can be compared, and this is done in Section **??**.

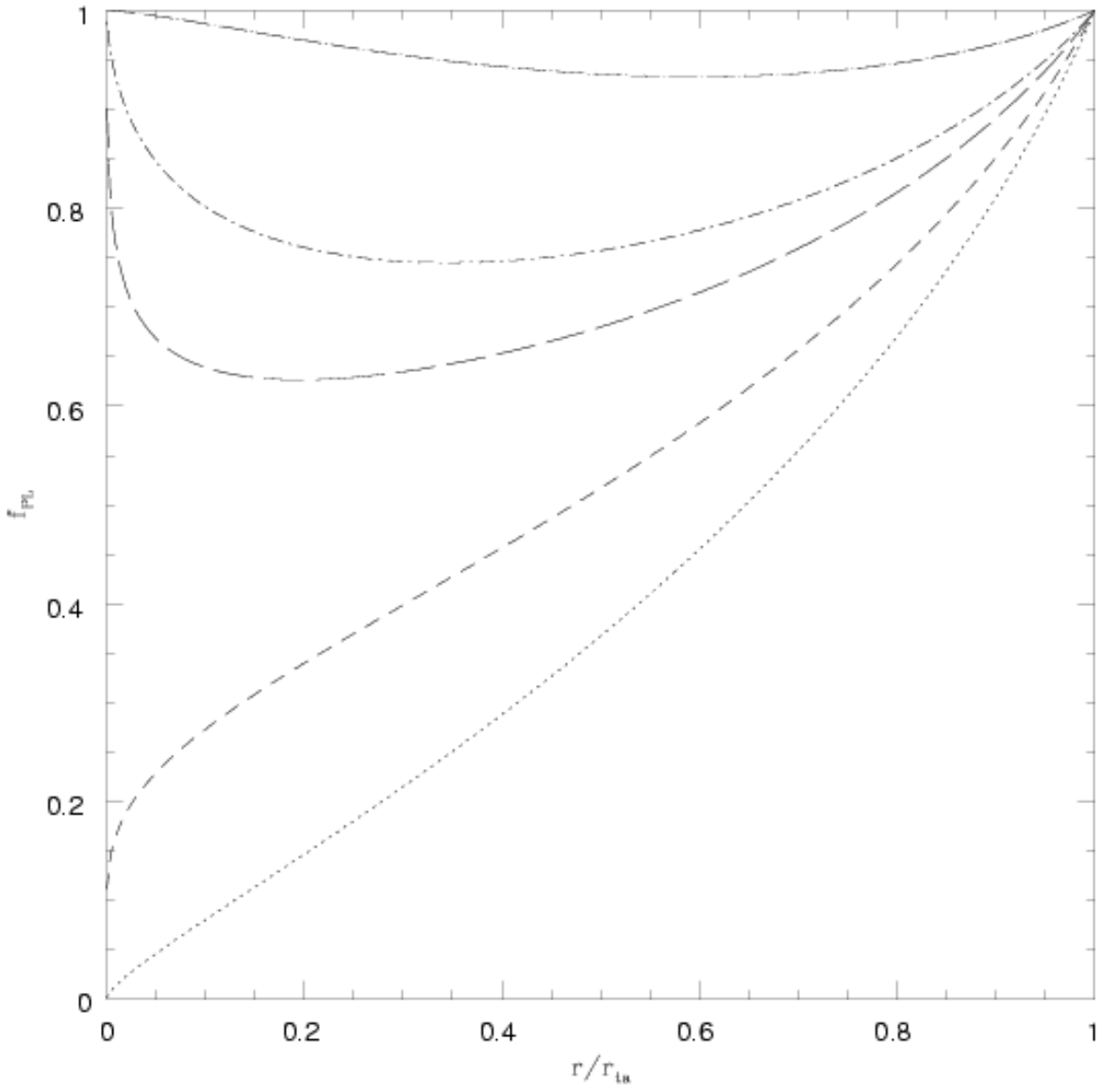

Figure 2.1: Collapse factor profiles $f_{\rm PL}(r/r_{\rm ta})$. The dotted, short-dashed, long-dashed, dot-short dashed, and dot-long dashed curves are, respectively, profiles for $\gamma = 1/2$, 1, 3/2, 7/4, 5/2.

2.4 EXPONENTIAL ACCRETION RATE

We are now going to repeat the calculations of the mass profile, collapse factor, and circular velocity, starting with the initial assumption of an exponential, rather than power-law, accretion rate.

Wechsler et al (2002) used the following exponential mass accretion rate, in terms of expansion factor:

$$\frac{M(a)}{M_0} = \exp\left[-\frac{1}{c'}\left(\frac{a_0}{a} - 1\right)\right], \tag{2.17}$$

where $a = (1+z)^{-1}$ and $c' \equiv c/2c_1$, where c_1 is some typical concentration. In an Einstein-de Sitter model, the expansion factor scales as $a(t) \propto \rho_b(t)^{-1/3} \propto M^{-1/3}r$. Then the accretion rate can be expressed in terms of radius:

$$\frac{M(r)}{M_0} = \exp\left[-\frac{1}{c'}\left(\left(\frac{M(r)}{M_0}\right)^{\frac{1}{3}}\left(\frac{r}{r_0}\right)^{-1} - 1\right)\right]. \tag{2.18}$$

The solution of this equation is

$$M_{\rm EXP}(r) = \mathrm{e}^{1/c'} \exp\left[-3\,\mathrm{ProductLog}\left(\frac{\mathrm{e}^{1/3c'}}{3c'}\frac{1}{r}\right)\right], \tag{2.19}$$

where $y = \mathrm{ProductLog}(x)$ is defined to be the solution to $x = y\,\mathrm{e}^y$, and M and r are now the scaled mass and radius. However, it turns out that a simpler expression is obtained if (2.18) is solved for radius as a function of mass:

$$r_{\rm EXP}(M) = \frac{M^{1/3}}{1 - c'\log(M)}. \tag{2.20}$$

As before, the energy of particles in a particular mass shell can be calculated by integrating the equation of motion. The calculation is simpler if it is done in terms of $r(M)$:

$$E_{\rm EXP}(M) \propto \int \frac{M}{r(M)^2}\left(\frac{dr}{dM}\right)dM = \frac{1}{2}M^{2/3}\left(1 + c'\left(\frac{9}{2} - \log(M)\right)\right). \tag{2.21}$$

The velocity is then $v(M) \propto [E(M') - E(M)]^{1/2}$.

The effect of "secondary infall" can now be accounted for, and the first step is to calculate $P(M|M')$, or equivalently, $P(r|r')$, which is the fraction of time that a particle on an orbit

with maximum radius r' spends within radius r. For the exponential accretion rate, this calculation can only be done numerically:

$$P_{\rm EXP}(M|M') = \int_0^M \frac{(dr/dM)\,dM}{\sqrt{1-E(M)/E(M')}} \Bigg/ \int_0^{M'} \frac{(dr/dM)\,dM}{\sqrt{1-E(M)/E(M')}}. \tag{2.22}$$

Then the ratio of the "additional mass," due to secondary infall, to the "initial mass" is

$$\frac{m_{\rm add}(M)}{M_{\rm EXP}} = \frac{\int_M^{M_{\rm ta}} dM'\,P(M|M')}{M}, \tag{2.23}$$

which can easily be converted into a function of radius. As before, the collapse factor is given by $f_{\rm EXP}(r) = [1 + m_{\rm add}(M_{\rm EXP}(r))\,/\,M_{\rm EXP}(r)]^{-1}$. The circular velocity can also be calculated by using $v^2 = GM(r)/r$. Figure 2.2 shows the collapse factor as a function of radius for different concentrations.

Now the total mass can be calculated by accounting for the contraction of the radii and by requiring that the mass enclosed within the virial radius remains constant. The density can be obtained by taking the derivative of the total mass with respect to r and dividing by $4\pi r^2$. The mass and density distributions are compared to those of an NFW profile in the next section.

2.5 COMPARISONS WITH FINAL DENSITY PROFILE

In this section, we compare the expected density profile in simulations to the mass and density profiles resulting from the models of the mass accretion rate, in order to test their consistency.

Navarro, Frenk, and White (1997) showed that the density profiles of dark matter halos fit well to the following double power-law formula:

$$\frac{\rho_{\rm NFW}(x)}{\rho_{\rm crit}} = \frac{\delta_{\rm char}}{(c\,x)(1+c\,x)^2}, \tag{2.24}$$

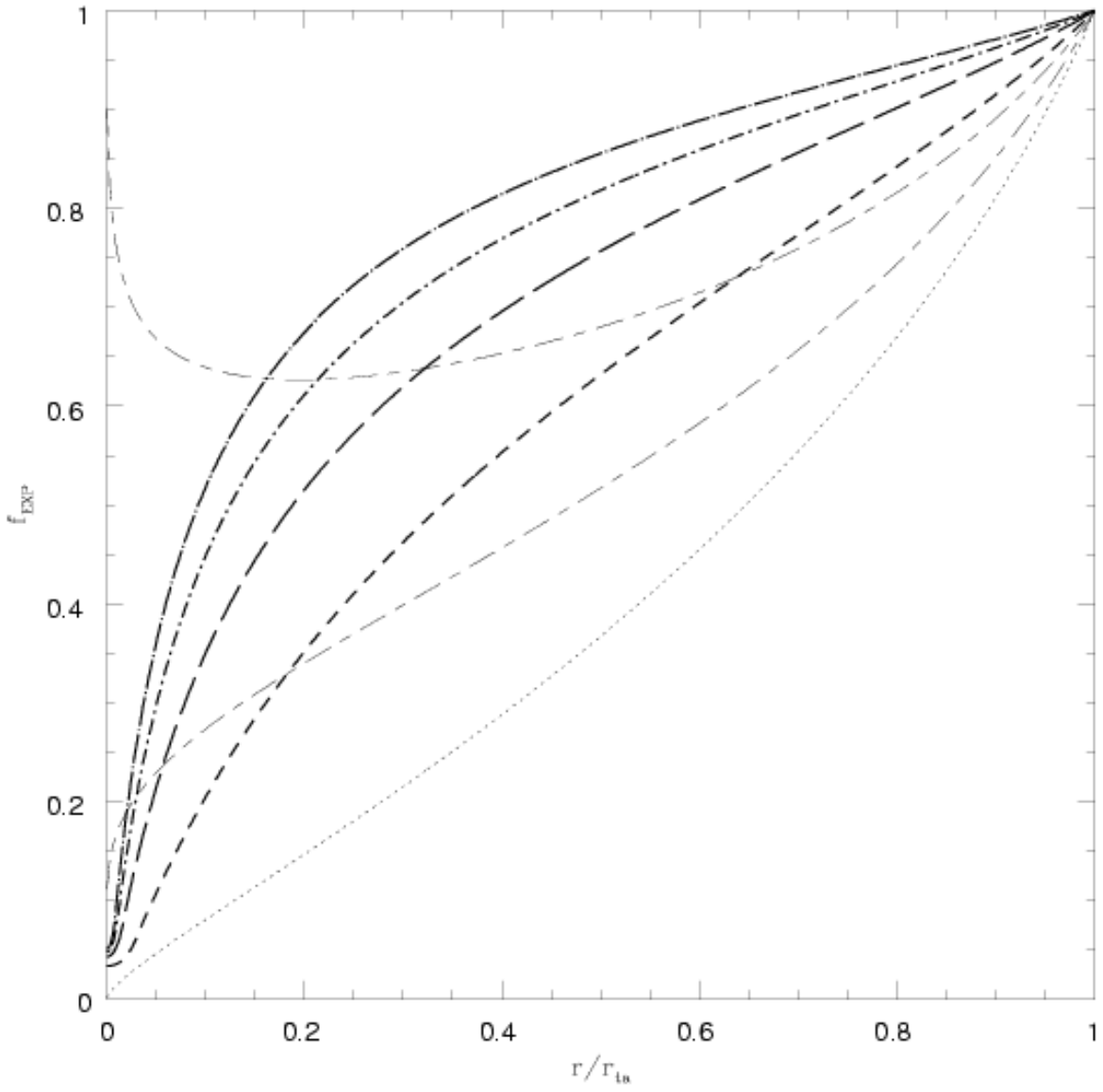

Figure 2.2: Collapse factor profiles $f_{\rm EXP}(r/r_{\rm ta})$. The short dashed, long dashed, dot-short dashed, and dot-long dashed curves are, respectively, profiles for $c = 5$, 10, 15, 20. These can be compared with the power-law profiles for $\gamma = 1/2$, 1, 3/2 (in gray).

where $x = r/r_{\rm vir}$ is the scaled radius, $c = r_{\rm vir}/r_{\rm s}$ is the concentration parameter, and $\rho_{\rm crit} = 3H^2/8\pi G$ is the critical density for closure. $\delta_{\rm char}$ is the characteristic density, which is given by

$$\delta_{\rm char} = \left(\frac{1+\Delta_{\rm vir}}{3}\right)\frac{c^3}{\ln(1+c) - c/(1+c)}, \tag{2.25}$$

where $1 + \Delta_{\rm vir} = 18\pi^2$, according to the calculation in Section 2.3.

The enclosed mass can easily be calculated by integrating $x^2\rho(x)$. One can also solve for the circular velocity profile implied by the NFW profile:

$$v_{\rm NFW}(x)^2 = \frac{GM_{\rm NFW}(x)}{x} = \frac{v_{\rm vir}^2}{x}\frac{\ln(1+cx) - (cx)/(1+cx)}{\ln(1+c) - c/(1+c)}. \tag{2.26}$$

The consistency of the accretion rates can be tested by comparing the mass, density, or circular velocity profiles with those of the expected NFW final density profile. Prior to making such a comparison, one must be aware of the following important caveat: the concentrations used in the NFW profile and the Wechsler et al (2002) exponential accretion rate, in equations (2.19) and (2.24), for example, are not equivalent. The NFW density profile was calculated by fitting to the virialized region of dark matter halos, and the "concentration" of dark matter was defined to be proportional to the virial radius. On the other hand, when the exponential accretion rate was used to calculate the collapse factor, the integrals involving the effects of the infalling mass were calculated out to the turnaround radius.

It is straightforward to determine the relationship between the concentration $c_{\rm ta}$ defined with respect to the turnaround mass and the concentration $c_{\rm vir}$ defined with respect to the virial mass. According to the exponential accretion rate in equation (2.17), we have $M(t)/M(t_{\rm ta}) = \exp[-(1/c'_{\rm ta})(a_{\rm ta}/a - 1)]$. It follows that

$$\begin{aligned}
\frac{M(t)}{M(t_{\rm vir})} &= \exp\left[-\frac{1}{c'_{\rm vir}}\left(\frac{a_{\rm vir}}{a} - 1\right)\right] \\
&= \frac{M(t)}{M(t_{\rm ta})}\left(\frac{M(t_{\rm ta})}{M(t_{\rm vir})}\right) = \exp\left[-\frac{1}{c'_{\rm ta}}\left(\frac{a_{\rm ta}}{a} - 1\right) + \frac{1}{c'_{\rm ta}}\left(\frac{a_{\rm ta}}{a_{\rm vir}} - 1\right)\right] \\
&= \exp\left[-\frac{1}{c'_{\rm ta}}\frac{a_{\rm ta}}{a_{\rm vir}}\left(\frac{a_{\rm vir}}{a} - 1\right)\right]
\end{aligned}$$

This entails that $c_{\rm ta}/c_{\rm vir} = a_{\rm ta}/a_{\rm vir}$, which is consistent with the relation $c(M_{\rm vir}, z) \propto (1+z)^{-1}$ found by Bullock et al. (2001a) and used by Wechsler et al. (2002). In addition, Zhao et al.

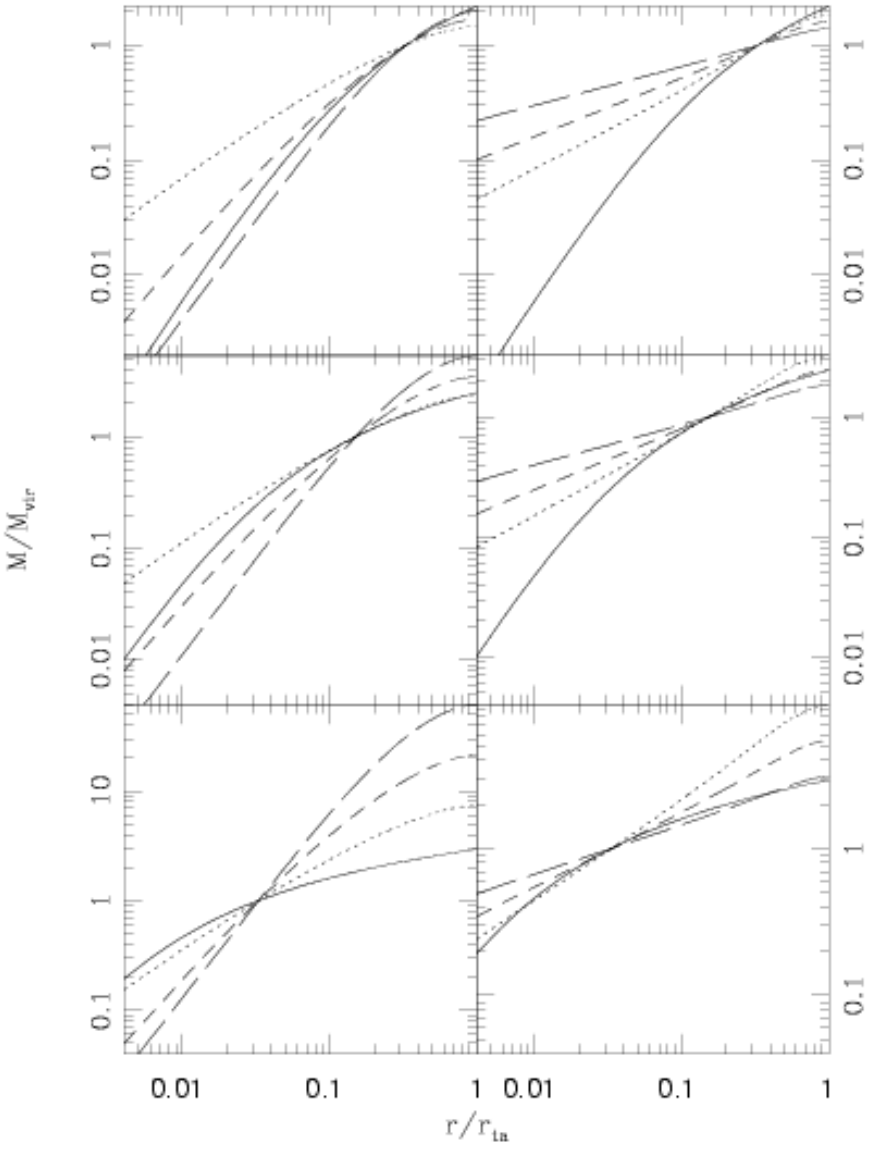

Figure 2.3: Mass profiles $M(r/r_{\rm ta})/M_{\rm vir}$. The solid curves are the NFW profiles for concentrations $c = 5$, 10, 20, in the top, middle, and bottom rows, respectively, and the dotted, short-dashed, and long-dashed curves are the profiles for the power-law with $\gamma = 1/2$, 1, 3/2, on the left side, and $\gamma = 7/3$, 5/2, 8/3, on the right side, respectively.

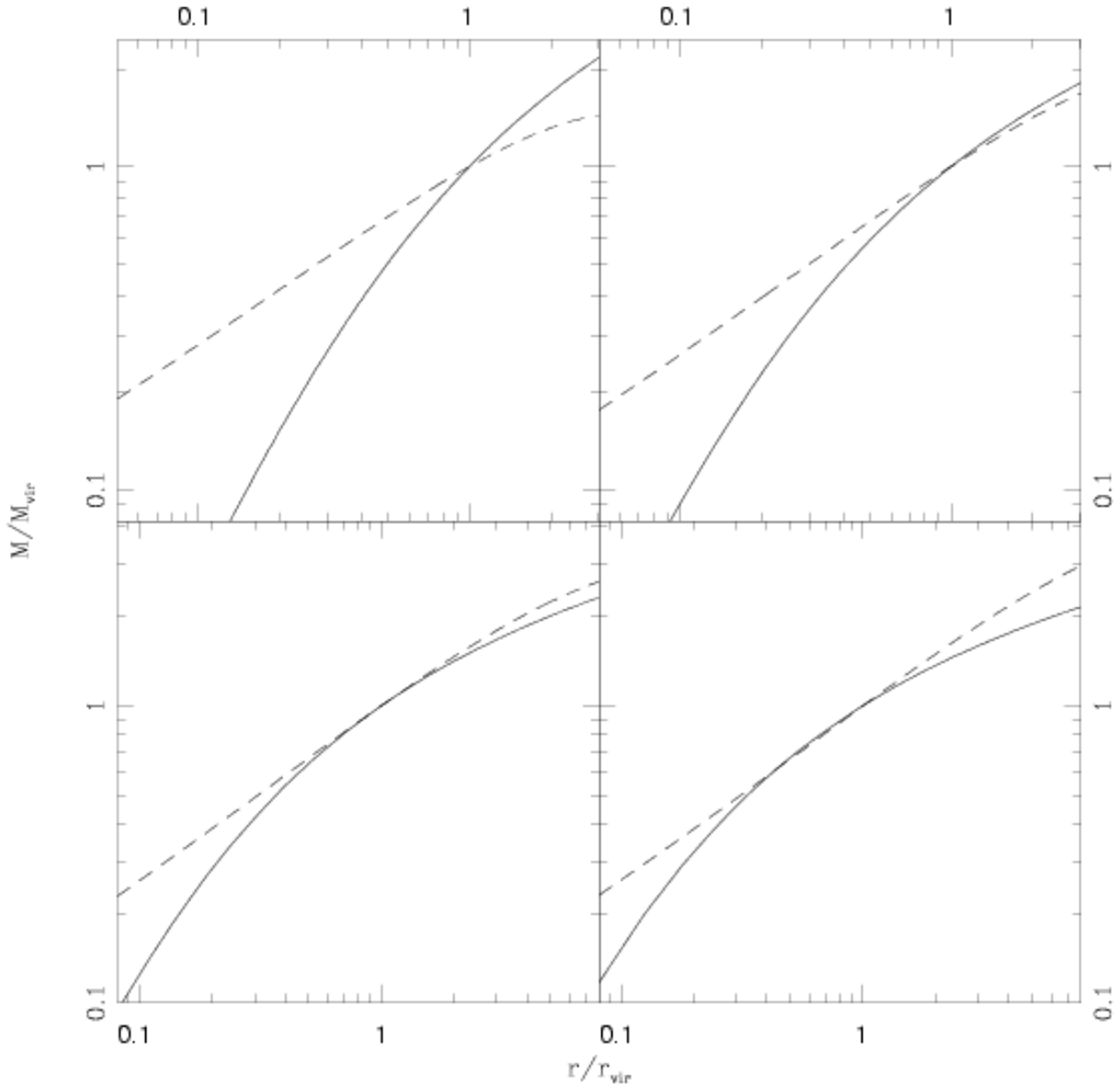

Figure 2.4: Mass profiles $M(r/r_{\rm vir})/M_{\rm vir}$. In each figure, the solid curves are the NFW mass profiles and the dashed curves are the mass profiles of the exponential accretion rate. The profiles are shown for concentrations $c = 5$, 10, 15, 20, in the top left, top right, lower left, and lower right, respectively.

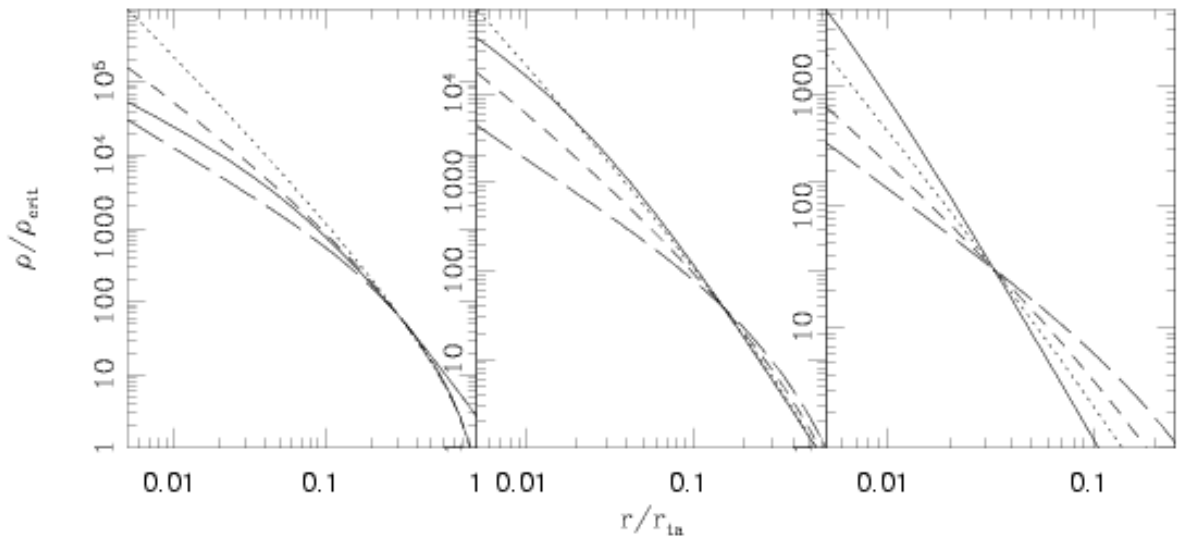

Figure 2.5: Density profiles of $\rho/\rho_{\rm crit}$ as a function of $r/r_{\rm ta}$. The solid curves are the NFW density profiles for concentrations $c = 5$, 10, 20, in the left, middle, and right figures, and the dotted, short dashed, and long dashed curves are the density profiles of the power-law accretion rate for $\gamma = 1/2$, 1, 3/2. As is the case with the mass profiles, the trend for the power-law as $\gamma = 1/2 \to 1 \to 3/2$ is fairly similar to the trend as $\gamma = 8/3 \to 5/2 \to 7/3$.

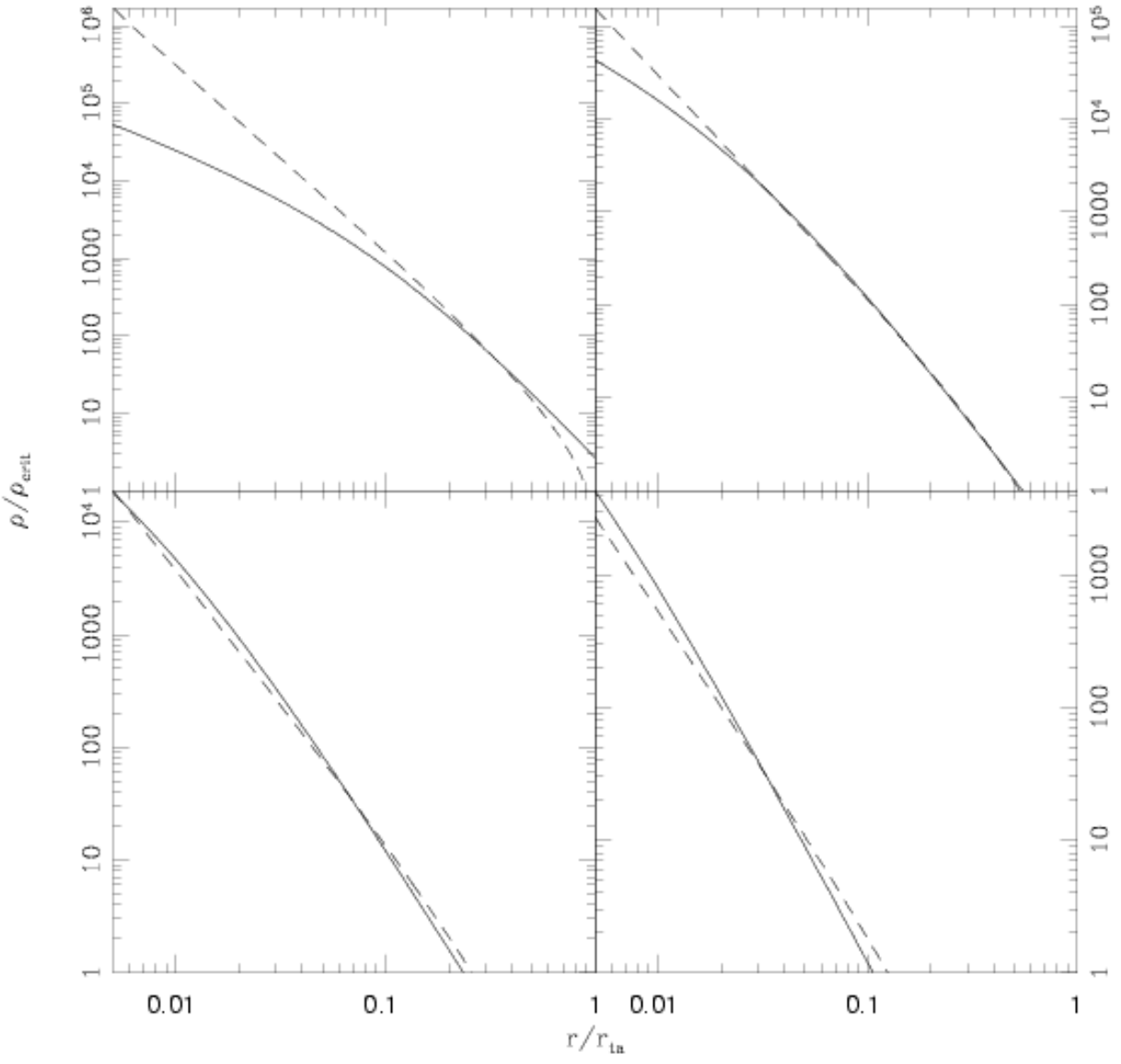

Figure 2.6: Density profiles of $\rho/\rho_{\rm crit}$ as a function of $r/r_{\rm ta}$. The solid curves are the NFW density profiles and the dashed curves are the density profiles of the exponential accretion rate. The profiles are shown for concentrations $c = 5$, 10, 15, 20, in the top left, top right, lower left, and lower right, respectively.

(2003) recently argued that there is a significant anticorrelation between the concentration and halo mass, and they found that, for large c, $c \propto a\, M_{\rm h}^{-(1-\alpha)/3\alpha}$, where $\alpha < 1$. This is consistent with the mass dependence obtained from the relation $c(M,z) = 9(M/M_*(z))^{-0.13}$, where $M_*(z)$ is the characteristic mass scale (Bullock et al. 2001b). Whether this dependence is significant in secondary infall models has yet to be resolved.

In any case, if we use the relation $c_{\rm ta}/c_{\rm vir} \cong a_{\rm ta}/a_{\rm vir}$, we can use this with the fact that the time it takes a dark matter halo to virialize is about twice the turnaround time. In other words, at some time $a_0 \propto T_{\rm vir}^{2/3}$ a halo has a virialized mass, and a larger turnaround mass that will in turn virialize at some later time $a_{\rm ta} \propto T_{\rm ta}^{2/3}$. Since with this notation $T_{\rm ta} \cong 2T_{\rm vir}$, it follows that

$$c_{\rm ta}/c_{\rm vir} \cong 2^{2/3} \quad \text{and} \quad M_{\rm vir}/M_{\rm ta} \cong \exp[-(2^{2/3}-1)/c'_{\rm ta}]. \tag{2.27}$$

These results make it possible to consistently compare the mass, density, and circular velocity profiles due to the exponential accretion rate with those of the expected NFW final density profile.

Figure 2.3 shows the mass profiles of NFW and the power-law accretion rate for various values of γ. Figure 2.4 shows the mass profiles of NFW and the exponential accretion rate for various concentrations. In each case, the enclosed mass is set to be constant at the virial radius. The ratio of the virial radius to the turnaround radius is determined from (2.27): $r_{\rm vir}/r_{\rm ta} \cong 0.331, 0.151, 0.070, 0.033$ for $c = 5, 10, 15, 20$, respectively. For most concentrations, the ratio is generally smaller than the ratio one could crudely estimate from the mean overdensity: $r_{\rm vir}/r_{\rm ta} \approx (\Delta_{\rm vir}/\Delta_{\rm ta})^{-1/3} \approx 1/\sqrt[3]{32}$.

In the power-law accretion case, there tends to be too little mass at small radii and too much at the turnaround radii for low values of γ, and it is almost the opposite for high values of γ. The exponential rate also does not reproduce the inner mass profile well, with too much mass in the inner regions for a wide range of halo concentrations.

It is also useful to compare the NFW density profile with those of the power-law and exponential accretion rates. Figure 2.5 shows comparisons with the power-law and Figure 2.6 shows comparisons with the exponential rate. The density is set to be constant at the virial radius, and this density, $\rho(r_{\rm vir}/r_{\rm ta})/\rho_{\rm crit}$, varies slightly with concentration.

The contrast between the power-law and exponential accretion rates is more evident in these two figures. The power-law poorly compares to the profiles of concentrated (low-mass) halos, resulting in a profile with too much density at small radii and not enough at large radii. The exponential rate agrees well with the density profiles of concentrated halos, while it results in an overly dense profile in the inner regions of diffuse halos.

2.6 MATTER ACCRETION WITH ANGULAR MOMENTUM

We have until now made the approximation that all orbits are radial in spite of the fact that most mass particles have a significant amount of angular momentum, which could significantly affect the resultant mass and density profiles. If one were to account for angular momentum, this would cause some of the mass apparently at smaller radii to be shifted to larger radii. We attempt to account for angular momentum in this section, and then make new comparisons with the NFW density profile.

A few different approaches have been used to account for angular momentum in dark matter halos. Nusser (2001) (following Sikivie et al. 1997) has suggested a procedure in which mass particles are assumed to have purely radial motion until they acquire a particular amount of angular momentum at the turnaround radius, when the gravitational forces cancel the force due to the expansion of the universe. Alternatively, others have attempted to use the dimensionless spin parameter (*e.g.*, Peebles 1969) to show how dark matter particles acquire a generally growing amount of angular momentum through tidal interactions with neighboring objects (Vitvitska et al. 2002, Bullock et al. 2001b). We apply each of the approaches in the two subsections below in order to make more accurate comparisons of the mass and density profiles.

2.6.1 Angular momentum acquired at turnaround

In Nusser's (2001) approach, dark matter particles acquire a particular amount of angular momentum per unit mass when they reach turnaround. The angular momentum is constant

for all shells, and it is conserved. In reality, particles at different locations on a given shell have different values of vector angular momentum, but an average over the distribution can be used (Sikivie et al. 1997). Since the angular momentum is thought to be mainly generated by tidal interaction between the infalling satellites, and since the dark matter particles spend the majority of their orbital periods at the maximum orbital radius, the assumption that the angular momentum of a particle is negligible before it reaches maximum expansion is a fairly accurate one. Angular momentum introduces a relatively small effective repellent force that prevents shells from reaching the halo center. Including angular momentum into the secondary infall model results in galactic halos with an effective core radius, having the effect of depleting the inner halo and of making the halo contribution to the rotation curve go to zero at the center, as expected.

Nusser (2001) begins with the equation of motion governing the evolution of a shell with a given angular momentum per unit mass, L:

$$\frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = -\frac{GM(<r,t)}{r^2} + \frac{L^2}{r^3}. \tag{2.28}$$

The angular momentum of a mass shell is assigned at the shell's turnaround radius according to

$$L^2 = \mathcal{L}^2\, G\, M_{\rm ta}\, r_{\rm ta}\,, \tag{2.29}$$

where $\mathcal{L}$ is assumed to be a constant value (in order to obtain self-similar solutions), between 0 and 1, for all shells. For small values of $\mathcal{L}$, particles move on nearly radial orbits, and particles with large $\mathcal{L}$ move on nearly circular orbits. It is important to note that, in the limit $\mathcal{L} \to 0$, the motion of a particle with zero angular momentum is not recovered.

The angular momentum term can be included with the radial energy when the velocity is calculated, which is used to determine $P(r|r')$, the fraction of time a particle on an orbit with maximum radius r' spends within r. The total energy of a particle, which is conserved, can be expressed as

$$E = \frac{1}{2}\left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2 + \frac{1}{2}\left(r\frac{\mathrm{d}\theta}{\mathrm{d}t}\right)^2 + \Phi(r), \tag{2.30}$$

where $L \equiv r^2\dot{\theta}$ and $\Phi(r)$ is the gravitational potential. Then the radial velocity is given by

$$v(r)^2 = \left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2 = 2[E - \Phi(r)] - \frac{L^2}{r^2}. \tag{2.31}$$

However, what is required in $P(r|r')$ is the difference between velocities:

$$v(r)^2 = \left(\left.\frac{\mathrm{d}r}{\mathrm{d}t}\right|_{r'}\right)^2 - \left(\left.\frac{\mathrm{d}r}{\mathrm{d}t}\right|_{r}\right)^2 = -2\Big(\Phi(r') - \Phi(r)\Big) - L^2\left(\frac{1}{r'^2} - \frac{1}{r^2}\right). \tag{2.32}$$

The potential is obtained by integrating $-\mathrm{d}\Phi(r)/\mathrm{d}r = -GM(<r)/r^2$, so in effect, $v(r)^2 = v_{L=0}(r)^2 - L^2/2r^2$, where $v_{L=0}(r)$ is the velocity calculated in Sections 2.3 and 2.4. Moreover, it is a ratio of velocities that will become important: $v(r)^2/v(r_{\rm ta})^2 = v_{L=0}(r)^2/v_{L=0}(r_{\rm ta})^2 - \mathcal{L}^2/2(r/r_{\rm ta})^2$. Once this velocity is calculated, with a particular amount of angular momentum, then equations (2.12) and (2.22) can be expressed as

$$P_{\rm PL}(u) = \frac{\displaystyle\int_0^u \frac{du}{\sqrt{2(u^{2-\gamma}-1) - L^2(u^{-2}-1)}}}{\displaystyle\int_0^1 \frac{du}{\sqrt{2(u^{2-\gamma}-1) - L^2(u^{-2}-1)}}} \tag{2.33}$$

$$P_{\rm EXP}(M|M') = \frac{\displaystyle\int_0^M \frac{(dr/dM)\,dM}{\sqrt{2(E(M') - E(M)) \;-\; L^2(r(M')^{-2} - r(M)^{-2})}}}{\displaystyle\int_0^{M'} \frac{(dr/dM)\,dM}{\sqrt{2(E(M') - E(M)) \;-\; L^2(r(M')^{-2} - r(M)^{-2})}}}. \tag{2.34}$$

These equations must be solved numerically, with a particular value of $\mathcal{L}$ for all shells. The results can be used to calculate the mass ratio of secondary infall, the collapse factor, the circular velocity profile, and the mass and density profiles. The modification to the collapse factor is shown in Figure 2.7, where it can be seen that including angular momentum this way slightly reduces the effect of "secondary infall".

By accounting for the effect of angular momentum in this way, we can again compare the mass and density profiles predicted by the power-law and exponential accretion rates with those of the NFW profile. Figure 2.8 shows comparisons with the power-law mass profile and Figure 2.9 shows comparisons with the mass profile of exponential accretion, for different amounts of angular momentum.

Some tentative conclusions can be drawn from the comparisons of the mass profiles. Firstly, power-law accretion tends to predict too little enclosed mass at smaller radii and too much at larger radii, whereas the opposite is the case for exponential accretion. Furthermore,

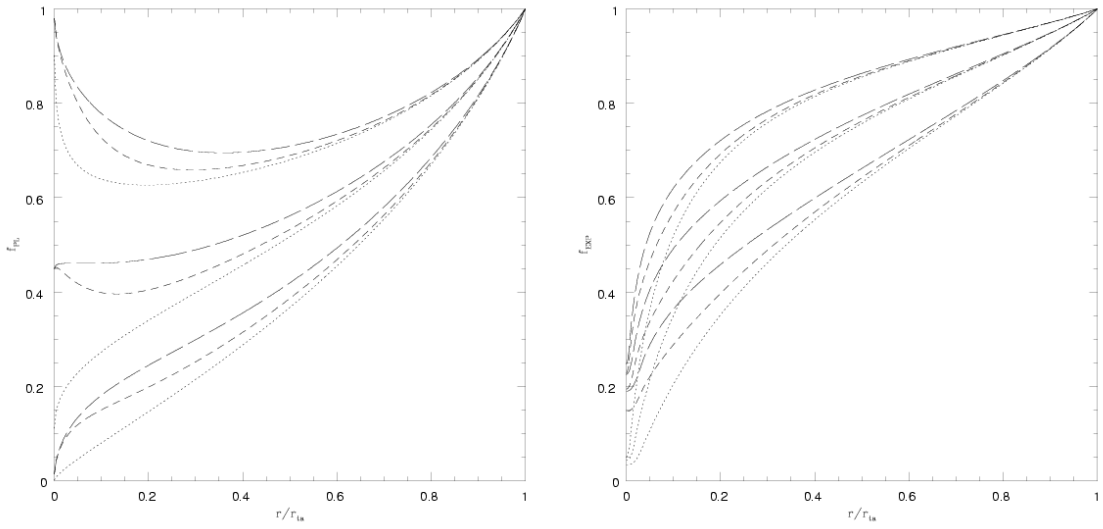

Figure 2.7: Collapse factor profiles $f_{\rm PL}(r/r_{\rm ta})$ and $f_{\rm EXP}(r/r_{\rm ta})$ for power-law ($\gamma = 1/2,\ 1,\ 3/2$) and exponential ($c = 5,\ 10,\ 20$) accretion. The dotted curves are the profiles calculated in Sections 2.3 and 2.4, and the short dashed and long dashed curves are the profiles subsequent to the inclusion of angular momentum, with $\mathcal{L} = 0.1,\ 0.3$.

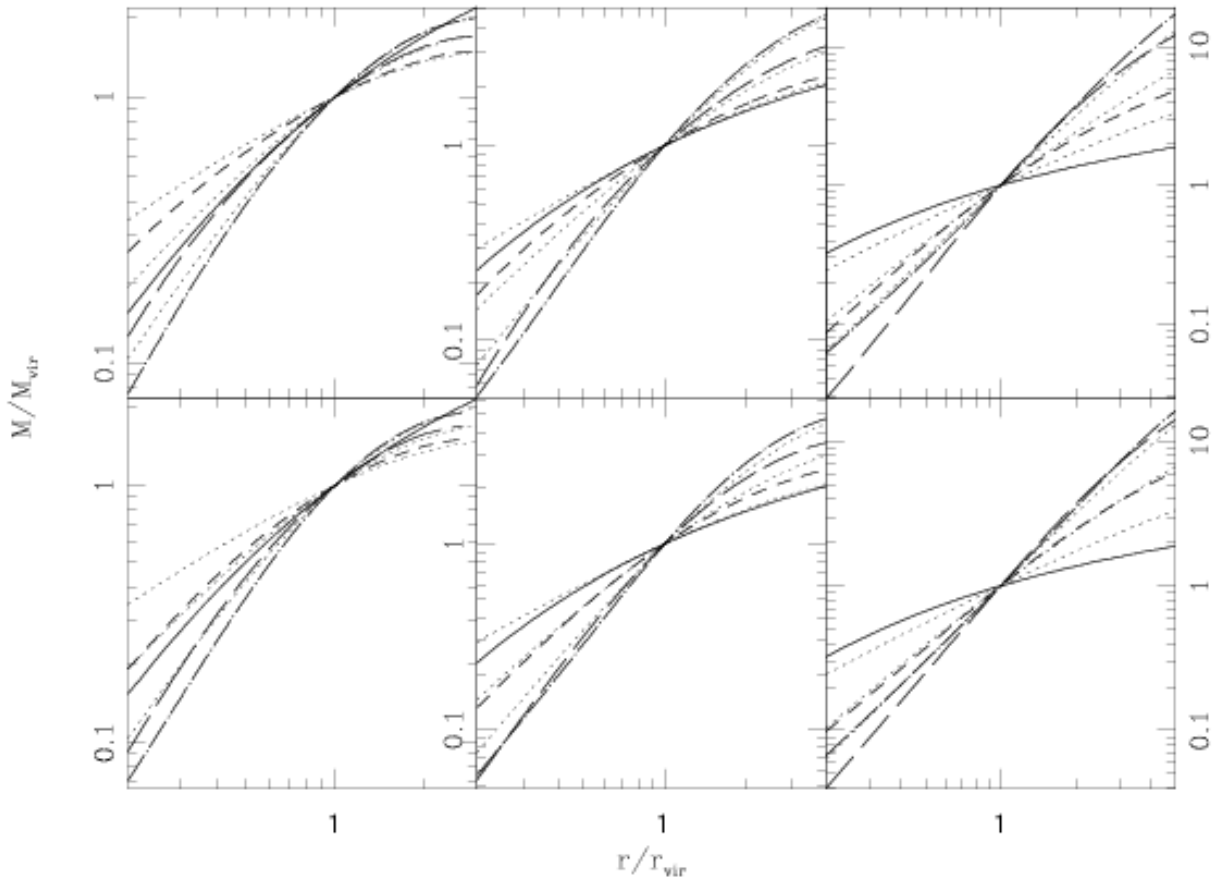

Figure 2.8: Mass profiles $M(r/r_{\rm ta})$. The solid curves are the NFW mass profiles for concentrations $c = 5$, 10, 20. The dotted curves are the mass profiles of power-law accretion prior to the inclusion of angular momentum, and the short dashed, long dashed, and dot-long dashed curves are the profiles subsequent to the inclusion of angular momentum, with $\gamma = 1/2$, 1, 3/2, respectively, and with values of $\mathcal{L} = 0.1$ and 0.3 in the top and bottom rows. In each case, the mass enclosed by the virial radius is set to be the same for the curve with angular momentum and the NFW profiles.

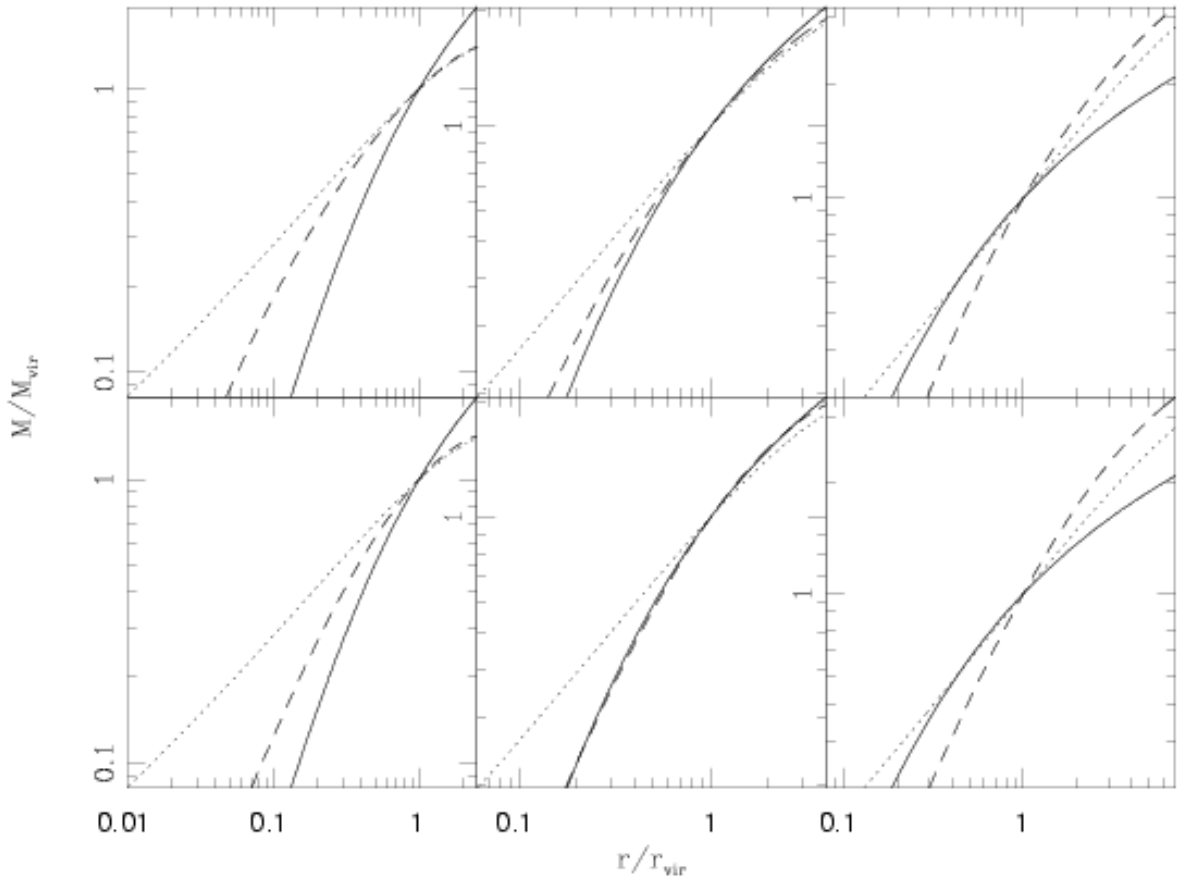

Figure 2.9: Mass profiles $M(r/r_{\rm ta})/M_{\rm vir}$. In each plot, the solid curves are the NFW mass profiles, and the dotted and dashed curves are the mass profiles of exponential accretion, prior and subsequent to the inclusion of angular momentum, with values of $\mathcal{L} = 0.1$, 0.3 in the top and bottom rows, for concentrations $c = 5$, 10, 20. In each case, the mass enclosed by the virial radius is set to be the same for the curve with angular momentum and the NFW profile.

the discrepancy of the mass profile of power-law accretion, as compared to the NFW profile, is exacerbated when angular momentum is included. In contrast, the inclusion of angular momentum with exponential accretion results in its mass profile more closely matching that of the NFW profile.

Density profile comparisons suggest similar conclusions. Figure 2.10 shows comparisons of the density profiles of each accretion rate with the NFW profiles. It is clear that the power-law accretion rate, except for the case of small values of γ compared to NFW profiles with low concentration, generally predicts too little mass at the inner radii and too much mass at the outer radii. Including angular momentum only exacerbates this problem. The exponential accretion rate, on the other hand, tends to yield more accurate predictions, at least within the virial radius of halos.

2.6.2 Angular momentum from spin parameter

Other astrophysicists have employed a somewhat different approach to account for angular momentum, involving the dimensionless spin parameter (Vitvitska et al. 2002, Bullock et al. 2001b). They follow the work of Fall & Efstathiou (1980), who studied the formation of disk galaxies within extended dark matter halos. Rather than assuming an angular momentum profile, they implicitly deduced it from the constraint that the final surface density profile of the disk is exponential. They deduced that the spin parameter, λ, is related to the total angular momentum, total energy, and mass of the system, in the following way (*e.g.*, Peebles 1969):

$$\lambda \equiv \frac{J\,|E|^{1/2}}{G\,M^{5/2}}. \tag{2.35}$$

The value of the spin parameter roughly corresponds to the ratio of the angular momentum of an object to that needed for rotational support. The distribution of spin parameters in N-body simulations is well-fitted by a log-normal distribution, with a median value of $\lambda \sim 0.05$, and typical values are between $0.02 - 0.11$.

According to Bullock et al. (2001b), in principle, given the spin parameter λ, the value of the global angular momentum of a halo, J/M (equivalent to Nusser's angular momentum per unit mass L), can be determined by using an assumed energy content for the halo. Since

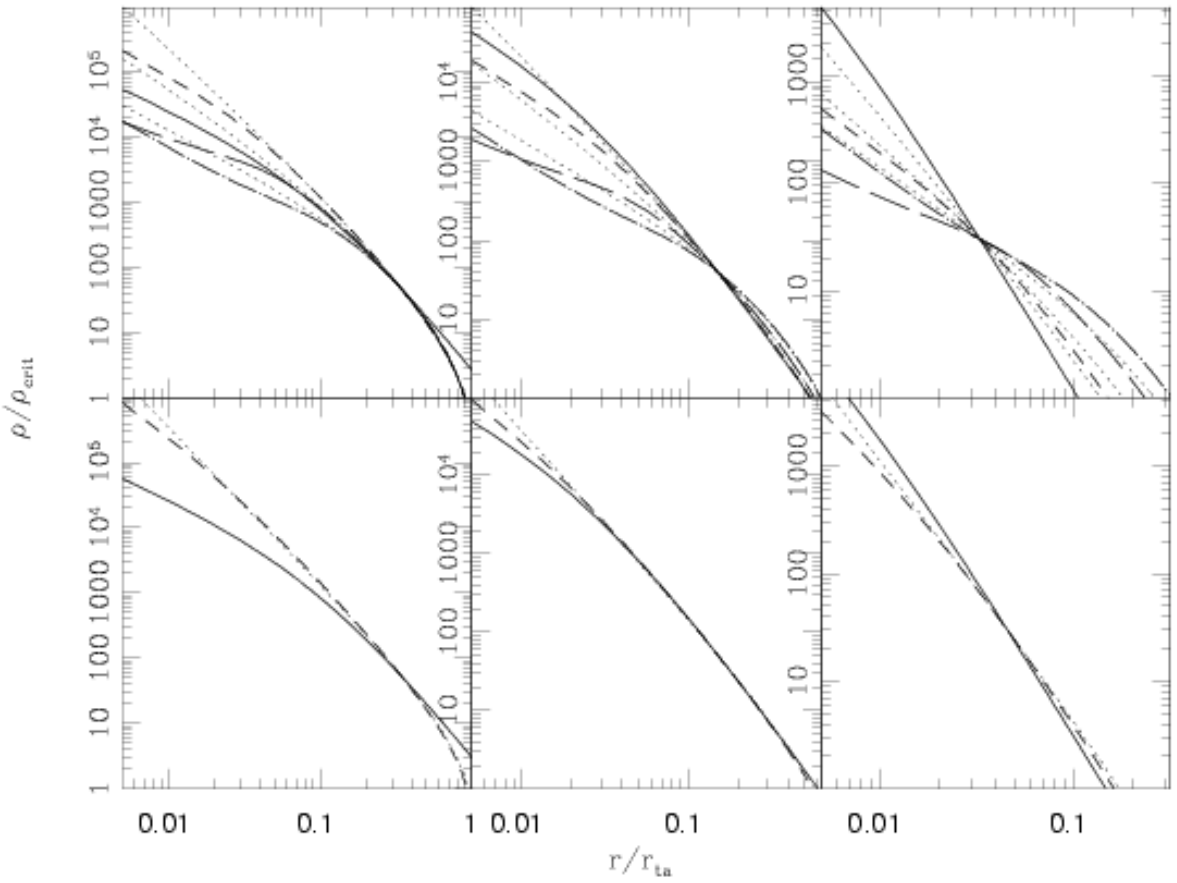

Figure 2.10: Density profiles of $\rho(r/r_{\rm ta})/\rho_{\rm crit}$. The top row shows the profiles of power-law accretion, prior (dotted curves) and subsequent (short dashed, long dashed, and dot-long dashed curves) to the inclusion of angular momentum, of amount $\mathcal{L} = 0.1$, for $\gamma = 1/2$, 1, 3/2. The bottom row shows the profiles of exponential accretion, prior (dotted curves) and subsequent (dashed curves) to the inclusion of angular momentum, for concentrations $c = 5$, 10, 20. In each case, the predicted profiles are compared to the corresponding NFW density profiles (solid curves) for $c = 5$, 10, 20.

this is not a straightforward procedure in practice, it is more useful to define a modified spin parameter λ' to characterize the global angular momentum:

$$\lambda' \equiv \frac{J}{\sqrt{2}\, M(<r)\, v\, r} = \frac{J}{\sqrt{2\, G}\, M(<r)^{3/2}\, r^{1/2}}, \tag{2.36}$$

where v is the circular velocity at radius r. The distribution of λ' is very similar to the distribution of λ, and the median value is $\lambda' \sim 0.04$.

Expressing the global angular momentum per unit mass in terms of the modified spin parameter, we have

$$L_\lambda = \frac{J}{M} = \sqrt{2\, G}\, \lambda'\, M_{\rm vir}^{1/2}\, r_{\rm vir}^{1/2}. \tag{2.37}$$

The fraction $P(M|M')$ can be calculated by using L_λ in equations (2.33) and (2.34). The collapse factor, circular velocity profile, and mass and density profiles can then also be calculated.

Since the values of λ' have been found to be fairly small, this entails relatively small values of L, and one would expect smaller values of $L(r)$ at the inner radii compared to the corresponding amount of angular momentum from Nusser's (2001) method in equation (2.28). Comparisons of the mass and density profiles show that this is indeed the case. In Figure 2.11,the mass profiles calculated from equation (2.36) with $\lambda' = 0.04$ and 0.1 are compared to the mass profile calculated from Nusser's method with $\mathcal{L} = 0.1$ and with the NFW profile. (Note that, since the calculations are scaled in terms of the virial and turnaround radii, in the case of power-law accretion, a value of $r_{\rm vir}/r_{\rm ta} \approx 1/\sqrt[3]{32}$ is assumed in (2.37). In the case of exponential accretion, this ratio is determined according to the concentration.)

Differences between the profiles are most significant at radii $r \leq 0.1$. One can clearly observe that even the mass profile with the large spin parameter $\lambda' = 0.1$, which is an amount of spin that only occurs in a small fraction of halos, predicts a little too much mass at the inner radii, and slightly more than the profile with $\mathcal{L} = 0.1$. Similar results are also evident in the density profiles plotted in Figure 2.12.

In this spin parameter approach, the angular momentum is weakly dependent on the mass distribution, the virial radius, and the concentration of the halo. The concentration and halo mass, in turn, are weakly anticorrelated, and they both increase at least proportionally with

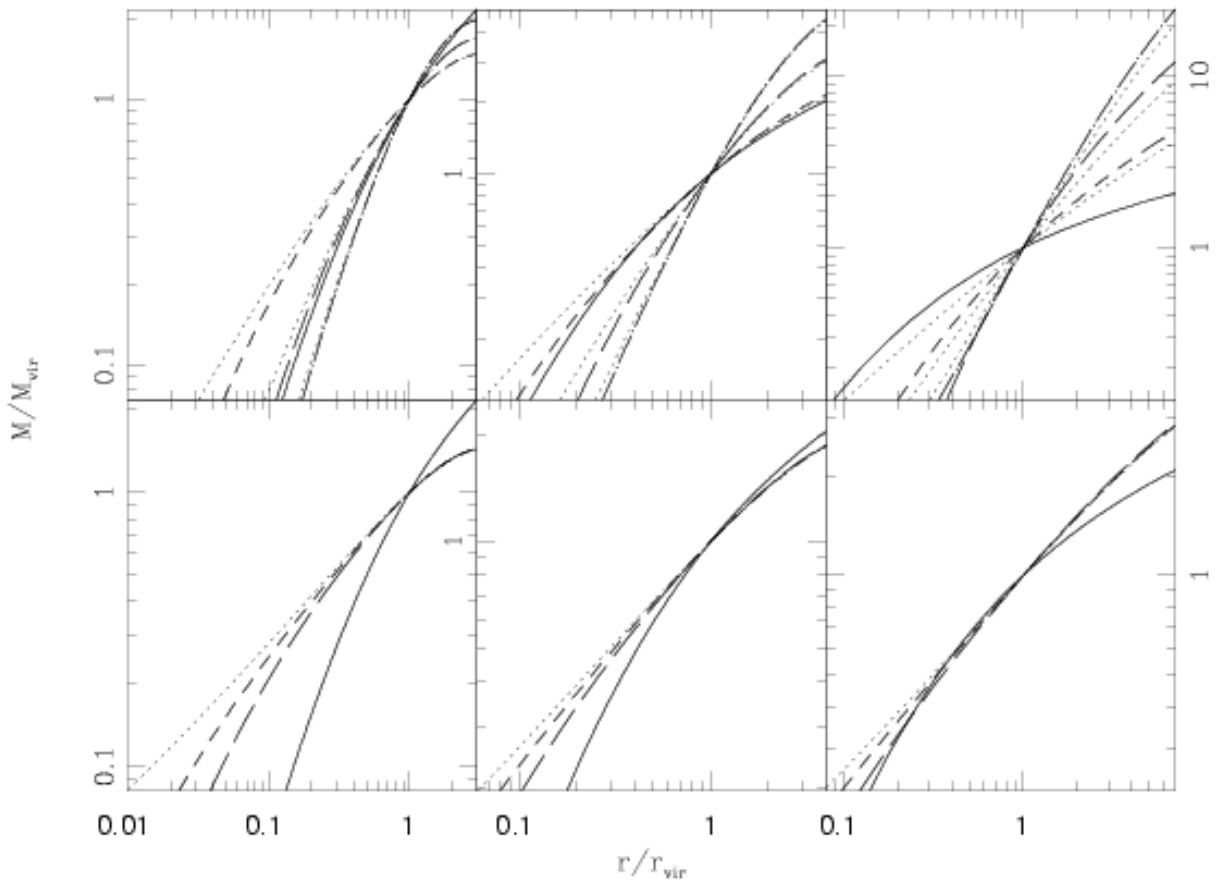

Figure 2.11: Mass profiles $M(r/r_{\rm vir})$. The top row shows the profiles of power-law accretion, prior (dotted curves) and subsequent (short dashed, long dashed, and dot-long dashed curves) to the inclusion of angular momentum, of amount $\lambda' = 0.04$, for $\gamma = 1/2,\ 1,\ 3/2$. The bottom row shows the profiles of exponential accretion, prior (dotted curves) and subsequent (dashed curves) to the inclusion of angular momentum, of amount $\lambda' = 0.04$ and 0.1, for concentrations $c = 5,\ 10,\ 20$. In each case, the predicted profiles are compared to the corresponding NFW mass profiles (solid curves) for $c = 5,\ 10,\ 20$.

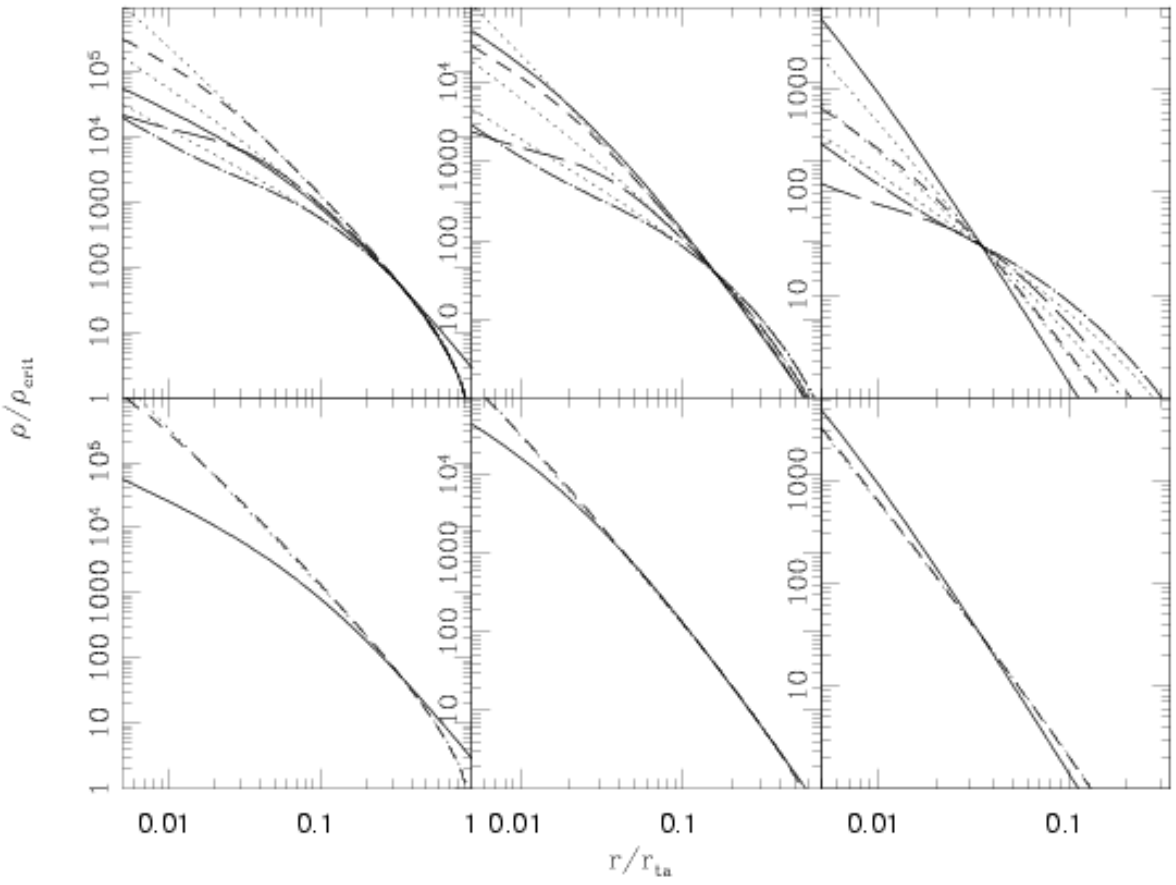

Figure 2.12: Density profiles of $\rho(r/r_{\rm ta})/\rho_{\rm crit}$. The top row shows the profiles of power-law accretion, prior (dotted curves) and subsequent (short dashed, long dashed, and dot-long dashed curves) to the inclusion of angular momentum, for $\gamma = 1/2$, 1, 3/2. The bottom row shows the profiles of exponential accretion, prior (dotted curves) and subsequent (dashed curves) to the inclusion of angular momentum, for concentrations $c = 5$, 10, 20. In each case, the predicted profiles are compared to the corresponding NFW density profiles (solid curves) for $c = 5$, 10, 20. The density profiles are shifted only by a small amount as a result of angular momentum, so a value of $\lambda' = 0.1$ was used.

the expansion factor a (Wechsler et al. 2002, Zhao et al. 2003). The angular momentum, like other properties of halo structure, is dependent on the halo's mass accretion history, especially if there is particularly clumpy accretion. The spin parameter is sensitive to the sequence of minor and major mergers, and especially to recent major mergers, which can cause it to increase almost by 25% (Vitvitska et al. 2002).

By analyzing the changes in angular momentum in terms of merger histories, one can begin to explain how dark matter particles acquire angular momentum through tidal interactions with neighboring infalling satellites of dark matter clumps. On average, major mergers dominate the halo formation at high redshift, and minor mergers dominate the slow accretion phase (Zhao et al. 2003). The spin parameter jumps up and down during the earlier fast accretion phase, usually with a net increase, and then it changes very little, sometimes with a slight decrease, during the later slow accretion phase. So as the halo mass rapidly increases at high redshift, the angular momentum increases rapidly as well, and then it gradually levels off at $z \leq 1$ (Vitvitska et al. 2002).

It may be the case that the spin parameter approach explains how dark matter halos acquire angular momentum more adequately than the approach involving angular momentum being acquired near turnaround. Using these approaches with an exponential mass accretion history appears to help explain the final density profiles of halos observed in simulations. Nonetheless, more sophisticated models and additional information about interactions among substructure and other physical processes are still required.

2.7 DISCUSSION AND CONCLUSION

We began with an exponential, rather than power-law, halo mass growth rate (Wechsler et al. 2002). We applied this accretion rate to spherical infall models, while continuing to assume that mass is accreted smoothly and spherically. We used these models to calculate the final mass, density, and circular velocity profiles that result from an exponential accretion rate. We have also attempted to include the effects of angular momentum in the processes of halo formation. We have found that, especially for large concentrations and within the virial

radii of halos, the final profiles are fairly consistent with the 'universal' profiles observed by Navarro, Frenk & White (1997). There is also more consistency with NFW profiles than one obtains with a simple power-law accretion rate.

Nonetheless, at least for massive halos with small concentrations, an exponential accretion rate still generally predicts too much mass at the inner radii and too little mass near the turnaround radius, compared to an NFW profile. In follow-up studies of NFW and M99, Navarro et al. (2004) and Diemand et al. (2004) are now in agreement that a steeper inner slope of ≈ -1.2 provides better fits, and this would lessen but not eliminate our discrepancy. A better understanding of the various dynamical and physical processes involved, such as ellipsoidal collapse and triaxial halos (*e.g.*, Jing & Suto 2002, Allgood et al. 2006), and the contraction of halos as a result result of the adiabatic infall of cooling baryons (*e.g.*, O. Gnedin et al. 2004, Sellwood & McGaugh 2005), may help to explain the origin of universal halo density profiles.

Since this work was completed, many interesting developments have occurred in the field, and we will mention some of them here. Hiotelis (2002) included angular momentum in a more sophisticated way than Nusser (2001) did, and was able to obtain double power-law final density profile. Manrique et al. (2003) and Salvador-Solé et al. (2005) used the Wechsler et al. (2002) exponential accretion rate with extended Press-Schechter theory to obtain NFW-like profiles. These final profiles were independent of the merger history, as Moore et al. (1999) showed they should be. Ascasibar et al. (2004) studied dark matter collapse and angular momentum with N-body simulations, and they argued that the density profiles are determined by the height and smoothing scale of the primordial density fluctuations. Austin et al. (2005) and Barnes et al. (2005) found agreement between their extended secondary infall model and N-body simulations about the scale-free nature of the phase-space density, ρ/σ^3, and they suggested that this is the outcome of violent relaxation. This seems to agree with the findings of Lu et al. (2006) and MacMillan et al. (2006), that the initial fast accretion phase of halo assembly, when the halo potential well is built and angular momentum is acquired, naturally results in universal inner density profiles.

Finally, Avila-Reese et al. (2005) showed that, although halo density profiles appear to be universal, their concentrations, central densities, ellipticities, and spin parameters vary

with environment. They argued that these correlations with environment may play a major role in causing the environmental dependence of galaxy properties, and in particular, the environmental dependence of the surface densities, circular velocities, bulge-to-disk ratios, gas fractions, and colors of disk galaxies. We explore these kinds of issues further in the following chapters of this thesis.

3.0 CONNECTING HALO MASS AND GALAXY LUMINOSITY

3.1 INTRODUCTION

In this paper, we determine relationships between halo mass and galaxy luminosity. This will be useful for two important purposes. Firstly, galaxy catalogs are usually defined by a luminosity threshold, and to calculate galaxy n-point correlation functions in the halo model, one needs to convert these minimum luminosities to minimum halo masses. Secondly, for luminosity-*marked* correlation functions, the relation between halo mass and the luminosity mark is required. Our analysis of galaxy clustering with luminosity-marked statistics and the luminosity-dependent halo occupation distribution in the halo model framework is very similar to the conditional luminosity function approach taken by Yang et al. (2003; 2005b) and Cooray (2005b; 2006).

Our approach for determining the relationship between central galaxy luminosity and host halo mass is similar to that of Vale & Ostriker (2004; 2005). They combine the cumulative number density of galaxies from the luminosity function with the cumulative number density of halos and subhalos from the mass function and the subhalo occupation distribution of simulations. Shankar et al. (2006) use the same approach, but they subtract the cluster mass function from the total halo mass function to obtain the mass function of halos containing a single galaxies. We similarly use cumulative number densities, but we use the *observed* luminosity-dependent halo occupation distribution with the mass function to obtain the cumulative number density of halos and subhalos. Our results for the SDSS *ugriz* bands are fairly similar to theirs for the 2dF b_J band. Our results are also similar to, but more tightly constrained than, those of Yang et al. (2003; 2005b).

It is important to point out that the assumption of a one-to-one monotonic relationship

between central galaxy luminosity and halo mass is physically motivated. Astronomers have observed that many galaxy properties, such as X-ray luminosity and temperature, are strongly correlated with cluster richness B_{gc}, which is in turn correlated with halo mass (*e.g.*, Yee & López-Cruz 1999, Yee & Ellingson (2003). Berlind et al. (2005) showed that central galaxies tend to be the most luminous, most massive, and among the reddest galaxies in halos, especially in denser environments. In underdense environments, central galaxies are fainter and bluer than their overdense counterparts and reside in less massive halos. De Lucia & Blaizot (2006) have also recently found from their semi-analytic models that brightest cluster galaxies (BCGs), a major subset of which are central galaxies, are strongly related to their host halo masses as well as their formation and assembly histories. The recent analysis of SDSS BCGs by Bernardi et al. (2006) also showed that, unlike fainter less massive galaxies, BCGs do appear to have distinct formation histories, forming from 'dry mergers' without recent star formation, and with massive black holes.

This chapter is organized as follows. First, we briefly describe the dark matter halo mass function determined from simulations. Next we describe the galaxy luminosity function determined from observations in the Sloan Digital Sky Survey (SDSS). In the fourth section we discuss the halo occupation distribution, the probably distribution of the number of galaxies occupying halos of given mass. In the fifth section, We show our results of the luminosity-mass relation and its dependence on luminosity function and halo occupation distribution parameters. We discuss our results in the final section.

3.2 HALO MASS FUNCTION

In this section we describe the dark matter halo mass function, the abundance of halos as a function of mass and redshift. Let $n(M, z)$ denote the comoving number density of dark matter halos, in which 'halo' refers to bound virialized almost-spherical objects with mean enclosed density approximately 200 times the critical cosmic density, independent of cosmology. Halos formed from regions in the primordial density field that were sufficiently dense that they later gravitationally collapsed. The following is a simple model (Bond et al.

1991) that describes the number density of these regions:

$$\nu\, f(\nu) \;\equiv\; m^2\, \frac{n(M,z)}{\bar{\rho}}\, \frac{\mathrm{d}\ln \mathrm{M}}{\mathrm{d}\ln \nu}\,, \tag{3.1}$$

where $\bar{\rho}$ is the comoving background density and $\nu\, f(\nu)$ is a distribution that is approximated by the first-crossing distribution of barriers crossed by Brownian random walks

$$\nu\, f(\nu) \;=\; \sqrt{\frac{\nu}{2\pi}}\, \exp(-\nu/2)\,. \tag{3.2}$$

Lacey & Cole (1993) developed such an analytical description of the merging rate of halos for barriers of fixed height, and (Sheth & Tormen 2002) expanded this for ellipsoidal collapse moving barriers.

The first-crossing distribution, is a function of redshift and mass: $\nu \equiv \delta^2(z)/\sigma^2(m)$. Here $\delta(z)$ is the critical density required for ellipsoidal collapse, modified by the growth factor for linear density fluctuations. The density fluctuation field is approximately Gaussian, so that the statistical properties of halos of mass M depend on the variance of the field smoothed with a tophat filter of scale $R(M) \propto (3\,M/4\,\pi\,\bar{\rho})^{1/3}$, extrapolated to the present:

$$\sigma^2_{\rm lin}(M) \;=\; \int \frac{\mathrm{dk}}{k}\, \frac{k^3\, P_{\rm lin}(k)}{2\pi^2}\, |W(kR)|^2\,, \tag{3.3}$$

where the window function is $W(x) = (3/x^3)(\sin(x) - x\cos(x))$ and σ^2 is normalized by the measured value of σ_8, the variance of the density field smoothed on a scale of $8\,\mathrm{Mpc}/h$. The CDM power spectrum can be approximated by

$$P_{\rm lin}(k) \;\propto\; k\, T^2(k) \tag{3.4}$$

with the transfer function $T(k)$ of Efstathiou, Bond & White (1992). The value $\nu = 1$ defines a characteristic mass scale such that $\sigma(M_*) = \delta_c(z)$, and it is $\sim 1.1 \times 10^{13} M_\odot/h$ at $z = 0$ in the ΛCDM cosmology. Halos more massive than M_*, which often host galaxy groups and clusters, are much more rare than less massive halos.

Finally, Sheth & Tormen (1999) found that the following modification of the Press-Schechter formula provides a better fit to the number density of halos in simulations of gravitational clustering:

$$\nu(M,z)\, f(\nu(M,z)) \;=\; A\left(1 \,+\, \frac{1}{(a\nu^2)^p}\right)\left(\frac{a\nu^2}{2\pi}\right)^{1/2} \exp(-a\nu^2/2) \tag{3.5}$$

where $a \approx 0.707$, $p \approx 0.3$, and the normalization is $A(p) \approx 0.3222$. This corrected mass function is significantly more accurate than the constant barrier model.

3.3 GALAXY LUMINOSITY FUNCTION

Similar to the halo mass function, we define the galaxy luminosity function $\phi(L,z)$ to be the number density of galaxies as a function of luminosity, or absolute magnitude, and redshift.

In the Sloan Digital Sky Survey (SDSS), the galaxy luminosity function and luminosity density has been measured in the five optical bandpasses shifted to their rest-frame shape at $z = 0.1$, denoted by u, g, r, i, z, with $\lambda_{\rm eff} \;=\; (3216,\, 4240,\, 5595,\, 6792,\, 8111$ Å), respectively (Blanton et al. 2001; 2003a). Absolute magnitudes (model or Petrosian) are constructed from the apparent magnitudes as follows:

$$M_r \;-\; 5\log_{10}h \;=\; m_r \;-\; \mathrm{DM}(z,\Omega_0,\Omega_\Lambda,h) \;-\; K(z)\,, \tag{3.6}$$

where the distance modulus DM depends on cosmology, and we assume ΛCDM parameters with $h = 1$, and $K(z)$ is the K-correction, which accounts for the fact that the observed band corresponds to different rest frame bands at different redshifts (Blanton et al. 2003b). One then obtains luminosities with

$$L_r \;=\; 10^{-0.4\,(M_r - M_{r,\odot})} \tag{3.7}$$

in terms of solar luminosities in the given band.

The general form of luminosity functions has typically been parametrized with a Schechter (1976) function, given by

$$\phi(L)\,\mathrm{d}L \;=\; \phi_*\left(\frac{L}{L_*}\right)^{\alpha}\exp\!\left(-\frac{L}{L_*}\right)\frac{\mathrm{d}L}{L_*} \tag{3.8}$$

The best fit luminosity functions are calculated with a maximum likelihood method that maximizes the joint likelihood of absolute magnitude and redshift. The results of this analysis are shown in Blanton et al. (2001; 2003a), and we often use the latter because of their corrections for evolution in the luminosity function. It is very important to account for the significant evolution in the luminosity and number density of galaxies as a function of redshift, and astronomers have employed different methods and obtained different results with them (Lin et al. 1999, Loveday 2004, Andreon 2004). The evolution of luminosity is quantified with an evolving M_*, and the evolution of number density is quantified with an evolving ϕ_*. Andreon (2004) also allows the faint-end slope α to evolve, but the lack of constraints on this and the lack of consensus in the field motivates us to treat the slope as constant. The following are the r-band LF parameters at $z = 0.1$ for ΛCDM obtained by Blanton et al. (2001; 2003a) and Bell et al. (2003), to give an example of the variety of luminosity functions astronomers have calculated: $\phi_*\,(10^{-2}\,h^3\,\mathrm{Mpc}^{-3}) \;=\; (1.46 \pm 0.12,\ 1.49 \pm 0.04,\ 1.37 \pm 0.07)$; $M_* - 5\log_{10}h \;=\; (-20.83 \pm 0.03,\ -20.44 \pm 0.01,\ -20.57 \pm 0.03)$; $\alpha \;=\; (-1.20 \pm 0.03,\ -1.05 \pm 0.01,\ -1.07 \pm 0.03)$, respectively. Observations in the r-band are considered to be the *most* accurate. In the following Sections, we use the results of Blanton et al. (2003a) and Bell et al. (2003), and those of Baldry et al. (2005) for the u-band.

3.4 CONNECTING HALO MASS AND GALAXY LUMINOSITY

We will connect the luminosities of galaxies and the masses of halos by matching the cumulative number densities of galaxies from the luminosity function and the cumulative number densities of *individual* halos from the mass function and the luminosity-dependent halo occupation distribution (HOD). We will describe the halo occupation distribution first, and

then we will show our resulting luminosity-mass relation and its dependence on LF and HOD parameters.

3.4.1 Halo Occupation Distribution

The luminosity function describes the number densities of individual galaxies as a function of luminosity, while the halo mass function describes the number densities of halos which may host subhalos and hence *multiple* galaxies. Sufficiently massive halos (more massive than $\sim 10^{11} M_{\odot}$) host galaxies, and low-mass halos typically host a single one. High-mass halos tend to host a single luminous 'central' galaxy as well as containing many subhalos that also host fainter 'satellite' galaxies, so that these massive halos can contain dozens, hundreds, or even thousands of satellite galaxies.

We connect the luminosities of central galaxies with the host halo masses using models of the halo occupation distribution, the probability distribution of halos of mass M hosting N galaxies. The first moment of the HOD has been found to be sub-Poisson at masses near the minimum mass (Berlind et al. 2003), and this has been explained by the distinct populations of central and satellite galaxies (Kravtsov et al. 2004). We assume that each halo contains exactly one central galaxy and possibly some satellite galaxies, which do have a Poisson distribution. The halo occupation distribution is then the following:

$$\langle N_{\rm gal}|M\rangle \;=\; \langle N_{\rm cen}|M\rangle \;+\; \langle N_{\rm sat}|M\rangle\,, \tag{3.9}$$

where

$$\langle N_{\rm cen}|M, M_{\rm min}\rangle \;=\; \frac{1}{2}\,{\rm erfc}\left(\frac{\log_{10}(M_{\rm min}(L_{\rm min})/M)}{\sqrt{2}\,\sigma_{\rm cen}}\right) \;\approx\; 1\,. \tag{3.10}$$

It is sometimes necessary to model the effect of scatter in the luminosity-mass relation by using an error function rather than a step function in $N_{\rm cen}$. Since the minimum halo mass and minimum galaxy luminosity are related in some function form as $\langle M_{\rm min}\rangle = f(L_{\rm min})$, if the scatter between them has a Gaussian distribution of width σ, then integrating over this distribution results in an error function.

In practice, we usually study galaxy samples with luminosities brighter than some threshold, which equates to a minimum halo mass (as indicated in equation 3.10). Therefore, the

halo occupation distribution consists of the central and satellite galaxies halos contain that are above the luminosity threshold. Since the luminosity (mass) function decreases with luminosity (mass), the HOD at fixed mass increases with fainter thresholds.

The first moment of the satellite occupation distribution is the mean number of satellite galaxies per halo of given mass:

$$\langle N_{\rm sat}|M, M_{\rm min}\rangle \;=\; \sum_{N{\rm sat}=0}^{\infty} N_{\rm sat}\, P(N_{\rm sat}|M, M_{\rm min}) \tag{3.11}$$

The satellite galaxies' probability distribution is a Poission distribution:

$$P_\nu(N_{\rm sat}|M) \;=\; \frac{\nu^{N_{\rm sat}} {\rm e}^{-\nu}}{N_{\rm sat}!}\,. \tag{3.12}$$

The mean value of the distribution increases with halo mass like a power law, and below a power law at low mass, and we model it as follows:

$$\langle N_{\rm sat}|M, M_{\rm min}\rangle \;=\; \exp\left(-\beta\,\frac{M_{\rm min}(L_{\rm min})}{M}\right)\left(\frac{M}{\mu\, M_{\rm min}(L_{\rm min})}\right)^{\alpha(L_{\rm min})}. \tag{3.13}$$

The minimum mass of halos that host the population of galaxies in a volume-limited catalog is $M_{\rm min}$; the critical mass above which halos typically host one or more satellite galaxies is $M_1 \equiv \mu M_{\rm min}$; and the power-law slope α characterizes the mass dependence of the efficiency of galaxy formation. If the number of satellites drops below the power law at masses near the minimum mass, we include the exponential factor with $\beta > 0$; $\beta = 0$ means no departure from the power law.

The best-fit parameters of the halo occupation distribution, determined from galaxy clustering measurements, are highly dependent on galaxy properties such as luminosity and color (Zehavi et al. 2005, Collister & Lahav 2005, Yang et al. 2005b). As a result, when studying a galaxy catalog with data cuts, and especially when studying a particular galaxy mark for marked statistics, it is crucial to carefully model the dependence of the halo occupation distribution on that galaxy property in an observationally-motivated way. By measuring the luminosity dependence of galaxy clustering in the SDSS, Zehavi et al. (2005) have constrained the luminosity dependence of the main HOD parameters. They found that μ is approximately independent of luminosity at the value of ≈ 23, although the slope of the

mean of the HOD α increases with the luminosity threshold and for thresholds brighter than $M_r < -20.5$ it is significantly greater than unity and increases more rapidly. This is consistent with the findings of Jing & Börner (2004) using 2dF galaxies. We adopt the Zehavi et al. (2005) $\alpha(L_{\rm min})$ in our $L_{\rm cen}(M_h)$ calculations in the next section. We expect the HOD parameterization to evolve with redshift as well, but clustering measurements at high redshifts are not yet precise enough to put any meaningful constraints on this.

3.4.2 Luminosity-Mass Relation

We determine the relationship between halo mass and central galaxy luminosity as follows. The cumulative number density of galaxies brighter than some luminosity is simply an integral over the SDSS luminosity function (Blanton et al. 2003a) above that luminosity. The cumulative number density of halos more massive than some mass is the sum over the product of the mass function and the luminosity-dependent halo occupation distribution (Zehavi et al. 2005), so that we are counting *all* halos hosting galaxies, parent halos as well as subhalos. For each such halo mass M we find the critical luminosity that yields the same number density, and this is the average luminosity of central galaxies hosted by halos of that mass:

$$\int_{L_{\rm cen}}^{\infty} {\rm d}L\, \phi(L,z) \;=\; \int_{M}^{\infty} {\rm d}M'\, n(M',z)\, \langle N_{\rm gal}|M', > M\rangle \tag{3.14}$$

Our procedure is very similar to that of Vale & Ostriker (2004; 2005) and Shankar (2006). It yields a central galaxy luminosity as a function of halo mass, $L_{\rm cen}(M)$, which increases with mass like a broken power-law and has a shallower slope at the high-mass end (see Section 3.4.3). One can also account for scatter in the relation between luminosity and mass by also integrating over a Gaussian function on the right-hand side of the above equation, for example, with

$$\langle N_{\rm gal}|M\rangle \;\rightarrow\; \int {\rm d}L\, \Phi(L|M) \tag{3.15}$$

where the log-space width of the Gaussian comes from the scatter of the observed Tully-Fisher relation. This yields the mean central galaxy luminosity $\langle L_{\rm cen}|M\rangle$.

For the mean satellite luminosity as a function of mass, $\langle L_{\rm sat}|M, L_{\rm min}\rangle$, the amount that it is greater than the minimum luminosity is obtained by integrating over the luminosity-

dependent HOD. This $L_{\rm sat}$ is a function of *parent* halo mass; any estimation of satellite luminosity in terms of subhalo mass would require an accurate modeling of tidal stripping and other dynamical processes, and hence would be very uncertain. Our calculation of the mean satellite luminosity and a comparison of its mass dependence and that of $\langle L_{\rm cen}|M\rangle$ are shown in Figure 5.2. The central and satellite luminosity marks, as well as the stellar mass and metallicity marks explained in Chapter 6, are coupled to each other in the sense that at fixed halo mass luminous (massive) central galaxies have fairly luminous (massive) satellites on average. This naturally produces the 'galactic conformity' phenomenon observed by Weinmann et al. (2006a).

It is important to bear in mind that our assumptions are few and simple. We assumed that sufficiently massive halos host at least one galaxy, and we assumed that the central galaxies are typically the brightest in their host halos, and that their luminosities and host halo masses have a monotonic one-to-one relation. Our results are determined by the halo mass function measured from simulations and the luminosity function and luminosity-dependent halo occupation distribution observed from galaxy surveys.

3.4.3 Results

We first compare the halo mass function (Sheth & Tormen 1999) to the r-band SDSS galaxy luminosity function (Blanton et al. 2003a), both at $z = 0.1$, in Figure 3.1. To compare the two, we have made the same assumptions for the halos as Yang et al. (2003) in their Figure 2, that each halo corresponds to exactly one galaxy and vice versa, and that each halo has exactly the same mass-to-light ratio $M/L = 100\,h\,{\rm M}_\odot/{\rm L}_\odot$. This crude model sets $N_{\rm sat} = 0$ for all M. The difference in shape between $\phi(L)$ and $n(M)$ then entails that the mass-to-light ratio M/L depends on M. In order to match the luminosity function and mass function one can either choose a reasonable M/L around L_* or one can match the cumulative counts of galaxies and halos. Either case results in an unreasonable mass-to-light ratio over the observed range of luminosities, which is why a non-trivial halo occupation distribution is required (equation 3.13).

The observed luminosity function in Figure 3.1 (red curve) is shallower at the faint end

and steeper at the bright end, suggesting that luminosity should increase more rapidly as a function of mass at lower masses and then become shallower at higher masses. From the figure one might expect the relation to change at $L \sim 10^{10} L_{\odot}$, or $M \sim 10^{12} M_{\odot}$, and we will show that this is approximately what in fact occurs when one matches the cumulative number densities of galaxies and halos.

Our central galaxy luminosity-halo mass relation resulting from integrating equation 3.14 is shown in Figure 3.2. It is approximately a double power-law, which breaks at $L_{\rm cen} \sim 10^{10} L_{\odot}$, and it monotonically increases for the whole mass range, although the slope is shallow at large masses. As expected, brighter central galaxies tend to reside in more massive halos.

The black solid curve is our fiducial model, with the r-band SDSS luminosity function observed by Blanton et al. (2003a) and the luminosity-dependent halo occupation distribution of Zehavi et al. (2005). Our result is consistent with that of Zehavi et al. (2005) for $M > 10^{12} M_{\odot}$. Vale & Ostriker (2004) obtain a similar central galaxy luminosity-halo mass relation in the b_J-band, and Shankar et al. (2006) obtain a luminosity-mass relation with a steeper high-mass slope, possibly because of the different luminosity function they assume.

The double power-law shape of the central galaxy luminosity-halo mass relation is obviously related to the shapes of the luminosity and mass functions, and the decreasing slope at higher masses and the mass scale of the break ($\sim 5 \times 10^{12} M_{\odot}$) have physical significance. The mass scale is similar to the critical mass scale at which the dominant mode of star formation shifts from cold gas flows to shock-heating (Birnboim & Dekel 2003, Keres et al. 2005). In addition, feedback effects, especially supernova feedback in low-mass galaxies and AGN feedback in massive galaxies, to varying degrees prevent gas from condensing and forming stars, thus reducing the overall efficiency of star formation (White & Frenk 1991, Benson et al. 2003, Croton et al. 2005). Moreover, "archeological downsizing", in which stars in more massive galaxies tend to form at earlier times and over shorter time spans, is largely a consequence of hierarchical halo structure formation (Neistein et al. 2006). The upshot is that, in low-mass galaxies and halos, whether at recent times or in the earlier progenitors of current massive halos, stars formed relatively rapidly from the continual cold gas supply, resulting in a steep relation between central galaxy luminosity and halo mass. At larger masses, as a result of merging and feedback processes that shut down the gas supply, star

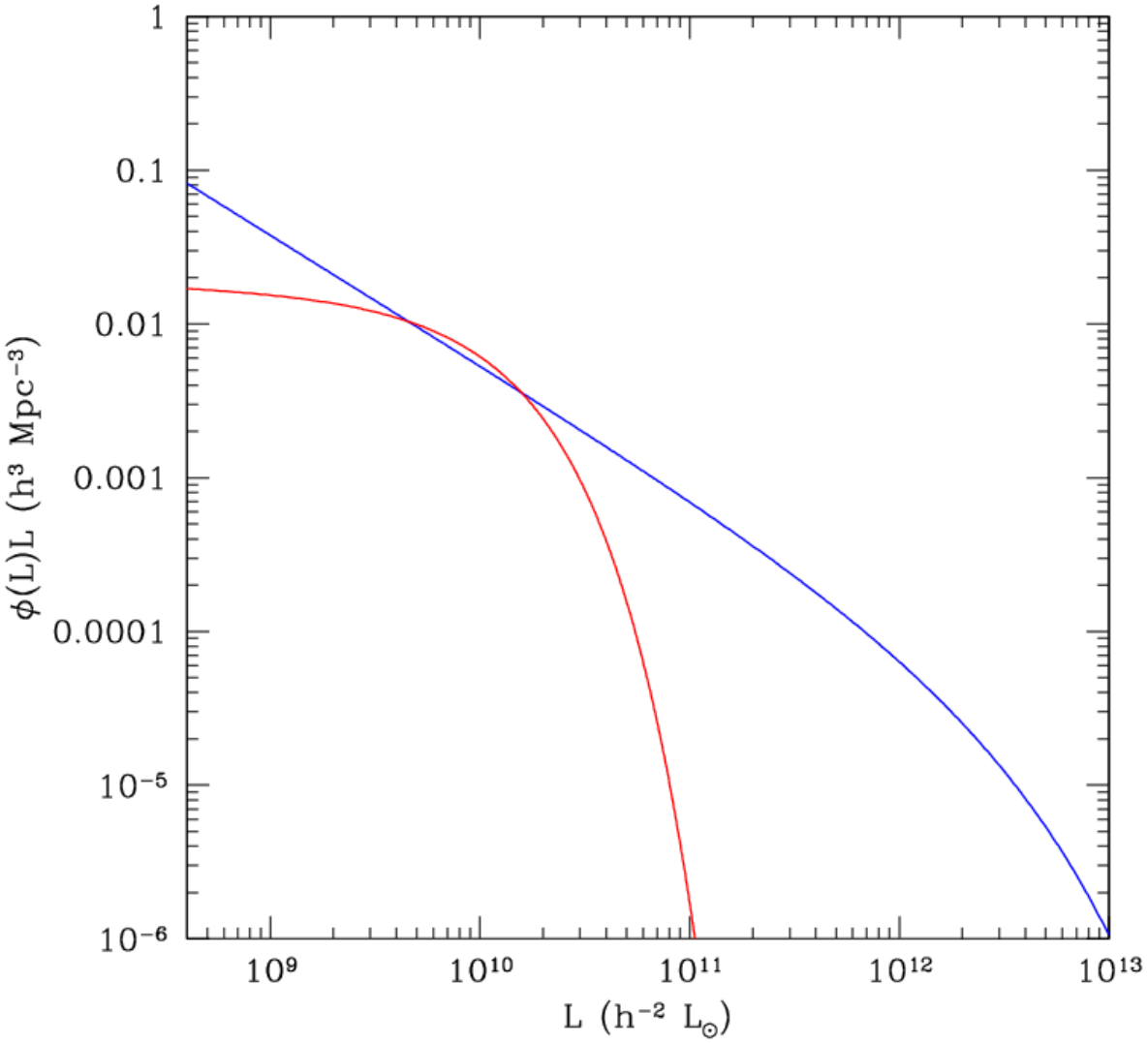

Figure 3.1: Comparison of galaxy luminosity function (red curve) and halo mass function (blue curve). The mass function is shown under the crude assumption that each halo hosts exactly one galaxy and that each halo has the same mass-to-light ratio $M/L = 100\,h\,\mathrm{M}/\mathrm{L}_\odot$, à la Yang et al. 2003 in their Figure 2.

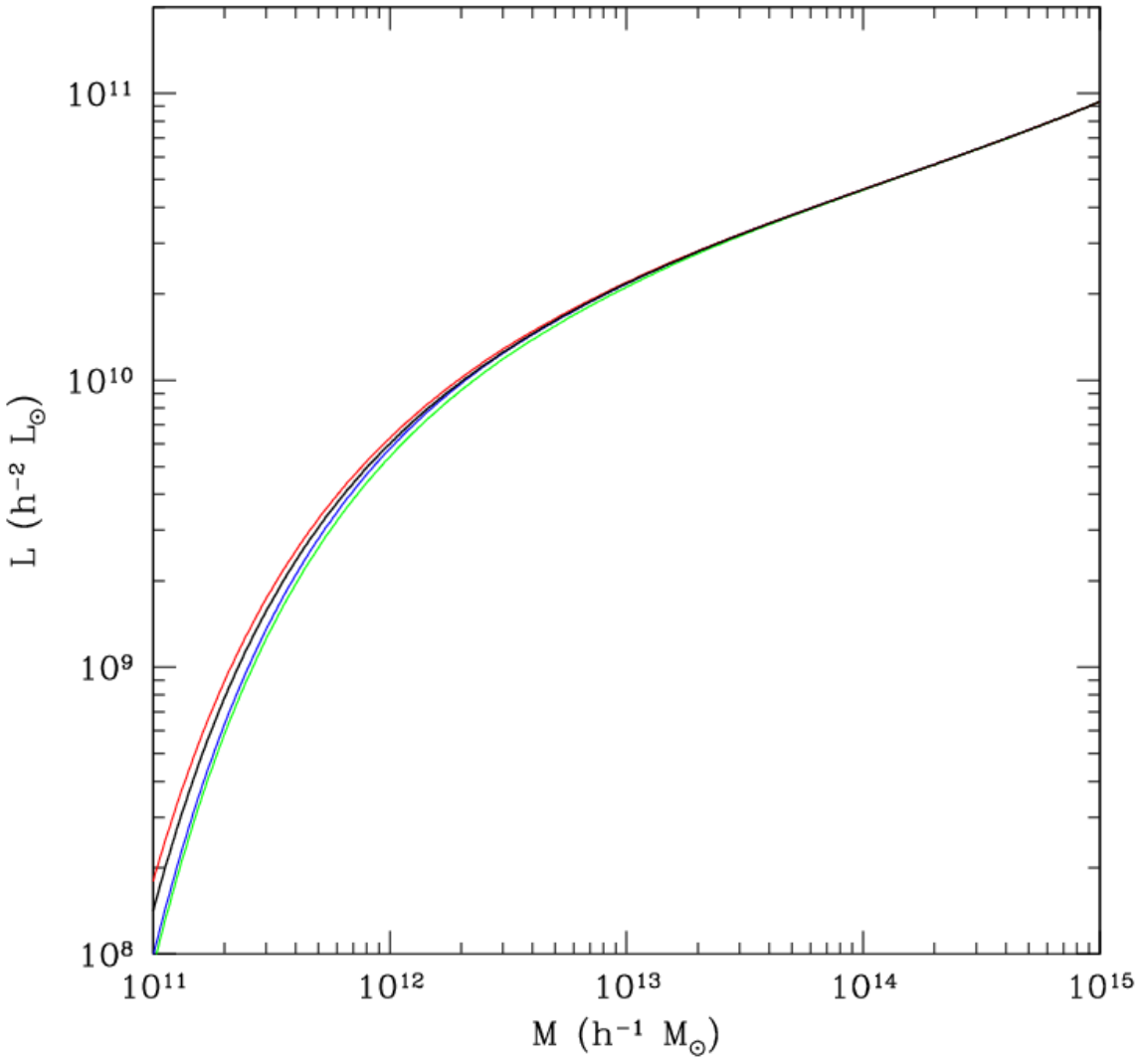

Figure 3.2: Central galaxy luminosity-halo mass relation $L_{\rm cen}(M)$ and its dependence on the halo occupation distribution. Fiducial model (black curve) assumes Zehavi et al. 2005 luminosity-dependent HOD, with $\mu = 23$, α increasing with luminosity, no low-mass exponential cutoff, and a step function in $\langle N_{\rm cen}|M\rangle$. Note the different effects of lower $\mu = 16$ (green curve), higher α (blue curve), nonzero β (red curve), and nonzero σ (not shown). See text for details.

formation is suppressed and increasing halo mass no longer translates into rapidly increasing luminosity and stellar mass, and this corresponds to the shallow slope of $L_{\rm cen}(M)$ at the high-mass end.

We show the dependence of the luminosity-mass relation on the HOD parameters in Figure 3.2, with a lower value of $\mu = 16$ (green curve) and a 20% larger value of the slope α (blue curve). Both of these result in more satellite galaxies in halos of a given mass, and their effect of slightly decreasing $L_{\rm cen}(M)$ is degenerate. We also show the effect of including an exponential low-mass cutoff with $\beta = 4$, which slightly increases $L_{\rm cen}(M)$ at the faint end. Finally, if one includes an error function in $\langle N_{\rm cen}|M\rangle$ of reasonable width ($\sigma = 0.1-0.15$) (not shown, for clarity), it results in slightly decreasing $L_{\rm cen}(M)$ if one integrates equation 3.14 down to lower mass, to include the tail of the error function, and slightly increases it if one does not, and this effect does persist to the high-mass end.

Most important, however, is the fact that the central galaxy luminosity-halo mass relation clearly has a very minimal dependence on the HOD parameters, especially for $M_r < -20$ and $M_h > 10^{12} M_\odot$. The HOD is *much* more important and its parameters are less degenerate in clustering predictions in the halo model, such as predictions of the two-point correlation function and marked correlation functions. The HOD produces the departure from a power-law in $\xi(r)$ at scales of $r \sim 1\,{\rm Mpc}/h$ and is thus somewhat constrained by clustering measurements at those scales (Berlind et al. 2003, Zehavi et al. 2004).

The dependence of the central galaxy luminosity-halo mass relation on the LF parameters is much more important, as shown in Figure 3.3. The $L_{\rm cen}(M)$ relations for the evolution-corrected LFs in the u, g, r, i, and z bands (Blanton et al. (2003a)) are shown (blue, green, red, cyan, and magenta curves, respectively). The shapes of these relations are fairly, similar, with the exception of the u-band. This means that galaxy colors are more weakly dependent on halo mass. We also show the r-band $L_{\rm cen}(M)$ resulting from the LF measured by Blanton et al. (2001) (dashed red curve), which is similar to LFs observed in the 2dF (*e.g.*, Norberg et al. 2002); the luminosity-mass relation is higher because of the brighter M_* and it is not as steep at the faint end because of the steeper faint-end slope α. One should also note that Schechter function fits to the LF are less reliable at the faint-end, and hence the luminosity-mass relation is less reliable at faint luminosities and low masses.

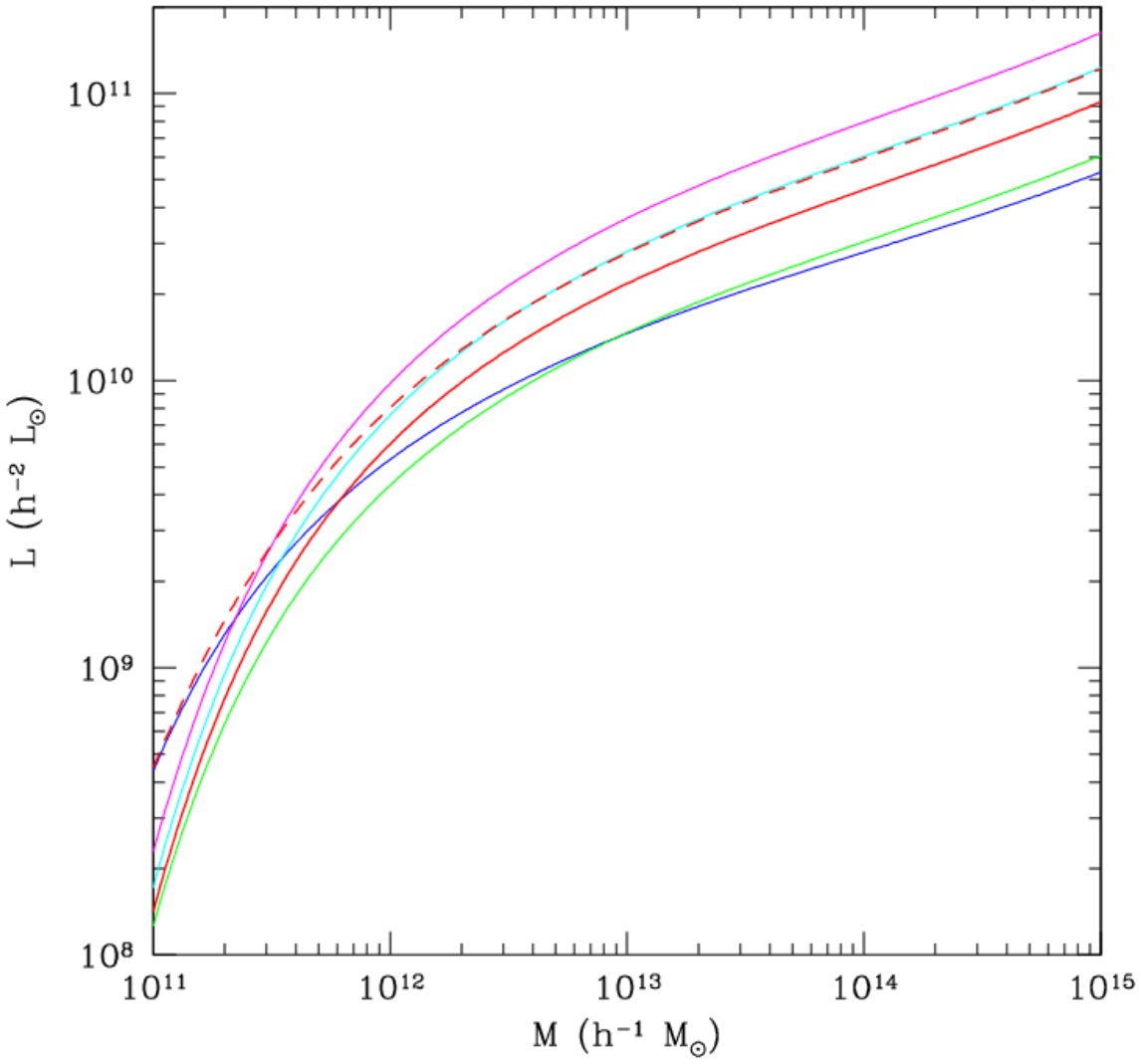

Figure 3.3: Central galaxy luminosity-halo mass relations for the *ugriz* band LFs (Blanton et al. 2003a), which are the blue, green, red, cyan and magenta curves, respectively. Also shown is the r-band $L_{\rm cen}(M)$ resulting from the Blanton et al. 2001 LF (dashed red curve).

What Figures 3.2 and 3.3 show is that the LF parameters must be known fairly accurate, especially if fainter luminosities are important, and that there is some room for error in the luminosity-dependent HOD. For correlation function predictions in the halo model, a higher (lower) central galaxy luminosity-halo mass relation entails that a particular luminosity threshold will correspond to a less (more) massive minimum halo mass, and thus a weaker (stronger) correlation function amplitude. For luminosity-marked statistics, the shape of the luminosity-mass relation matters more than its amplitude, such that a steeper $L_{\rm cen}(M)$ generally results in a stronger marked signal.

3.5 CONCLUSION

We have described how the relation between central galaxy luminosity and host halo mass is determined. We obtained this relation by matching the cumulative number densities of galaxies from the luminosity function on one hand, and the cumulative number densities of halos from the mass function and halo occupation distribution on the other. Our only assumptions were that sufficiently massive halos host galaxies and that there is a monotonic one-to-one relationship between luminosity and mass.

We have shown that the central galaxy luminosity-halo mass relation depends both on the parameters of the luminosity function and on the parameters of the halo occupation distribution, but moreso on the former. The amplitude of $L_{\rm cen}(M)$ can significantly affect the amplitude of correlation functions predicted by the halo model (*i.e.*, the strength of galaxy clustering), and the shape of $L_{\rm cen}(M)$ significantly affects the luminosity-marked correlation functions predicted by the halo model (*i.e.*, the strength of luminosity clustering, or its environmental dependence).

The relationship between luminosity and halo mass has additional uses for marked statistics. One can apply it to interpret color-marked statistics, as shown in Chapter 5, and given the stellar mass-to-light ratio, one can apply it to stellar mass- and metallicity-marked statistics, as shown in Chapter 6. Finally, given the mass-to-light ratio and the relation between stellar mass and star formation rate, one can determine the halo-model prediction

of SFR-marked statistics, although this is complicated by additional requirements such as information about the SFR-dependence of the HOD and satellite galaxy distribution.

4.0 DARK MATTER HALO MODEL OF GALAXY CLUSTERING AND MARKED CLUSTERING

4.1 INTRODUCTION

Most galaxy clustering analyses treat galaxies as mere points without attributes. However, galaxies have luminosities, masses, metallicities, concentrations, star formation rates, etc. The datasets are now sufficiently large and the quality of the data is sufficiently good for measurements of the spatial correlations of galaxy attributes themselves. Rather than measuring the clustering of galaxies for different luminosity cuts or color cuts, for example, one can weight galaxies by their attributes and measure the clustering of luminosity or color, thus more directly quantifying the correlations between these attributes and large-scale environment.

The theory of this approach of marked point processes and marked correlation functions was developed by Stoyan & Stoyan (1994). Marked statistics recently have been applied to astrophysical datasets by Beisbart & Kerscher (2000), Beisbart, Kerscher & Mecke (2002), Gottlöber et al. (2002), Faltenbacher et al. (2002), and Kasun & Evrard (2005), and to N-body simulations by Wechsler et al. (2005) and Harker et al. (2006). Sheth (2005) recently developed The theory of marked statistics in the halo model framework (see Cooray & Sheth 2002, for a review).

Marked statistics are novel and under-exploited tools useful for identifying and quantifying the environmental dependence of galaxy properties (Sheth, Connolly & Skibba 2005). They have been shown to be very sensitive to environmental effects (Sheth & Tormen 2004), obviating the need for making assumptions about environmental indicators. In the halo model, the formation and evolution of dark matter halos are correlated with the large-scale

environment only insofar halo abundances are environmentally dependent. That is, if particular halo properties or the properties of galaxies they host are environmentally dependent in away, this environmental dependence is *entirely* due to the correlation between halo mass and environment, with low-mass halos tending to reside in underdense regions and massive halos in overdense regions. In addition to halo mass, there also is a weak dependence on halo formation time (*e.g.*, Gao et al. 2005), but this appears to be *much* weaker in comparison (see Chapter 7 here, and Sheth & Tormen 2004).

By quantifying the environmental dependence of galaxy formation with measurements of marked statistics with various galaxy marks, and by comparing these to the correlations with environment predicted by the halo model, one can learn to what extent the observed environmental dependence is driven by that of halo mass. Such analyses have recently been performed for luminosity, color, stellar mass, and metallicity (Skibba et al. 2006; Skibba & Sheth, in preparation).

Since the clustering of real galaxies and their properties is strongly affected by redshift distortions, halo-model comparisons of marked clustering measurements must also be done in redshift space or projected space. In this paper we develop the halo-model description of marked statistics in redshift and projected space for these types of analyses. This description essentially consists of a combination of the results of Sheth (2005) with those of Seljak (2001).

This paper is organized as follows. In the next section we introduce marked statistics, and in particular, the marked two-point correlation function in real space, redshift space, and projected. We describe the components of the halo model required for modeling the galaxy correlation function in Section 4.3. In Section 4.4, we develop the halo-model description of the unmarked and marked power spectrum in redshift space. In Section 4.5, we describe another marked statistic, using the mean mark. We then conduct a simple test of the environmental sensitivity of marked statistics in Section 4.6. We discuss our results in the final section.

4.2 MARKED STATISTICS

In what follows, a mark is a weight or attribute associated with each point in a point process. A point process is a galaxy catalog, and a mark can be any observable property associated with a galaxy, such as luminosity, color, velocity dispersion, size, star formation rate, etc. Marked statistics measure the clustering of marks. Since the positions at which the marks are measured may themselves be clustered, marked statistics are defined in a way which accounts for this.

For example, let $\bar{\rho}$ denote the mean density of particles, and let $\bar{w}$ denote the mean mark, averaged over all particles. Now consider a particle with mark larger than this mean value. Are the particles neighboring it also likely to have larger marks? One way to quantify this is to compute the mark product, or 'marked' two-point correlation function, which consists of the ratio of the 'weighted' correlation function to that of the unweighted correlation function, ξ. The typical number of pairs at separation r is $\bar{\rho}^2[1+\xi(r)]$. Therefore, the marked correlation function is

$$\begin{aligned} M(r) &\equiv \frac{\sum w(\boldsymbol{x})w(\boldsymbol{y})\,\mathcal{I}(|\boldsymbol{x}-\boldsymbol{y}|-r)}{\bar{w}^2\,\sum \mathcal{I}(|\boldsymbol{x}-\boldsymbol{y}|-r)} \\ &= \frac{\sum w(\boldsymbol{x})w(\boldsymbol{y})\,\mathcal{I}(|\boldsymbol{x}-\boldsymbol{y}|-r)}{\bar{w}^2\,\bar{\rho}^2[1+\xi(r)]} \equiv \frac{1+W(r)}{1+\xi(r)}, \end{aligned} \tag{4.1}$$

where $\mathcal{I}(x) = 0$ unless $x = 0$, and the sum is over all galaxy pairs. The only difference between the sums in the numerator and the denominator is that, in the numerator, the ith particle contributes a weight $w_i/\bar{w}$, whereas in the denominator, the weight is unity for all particles. In effect, the denominator divides out the contribution to the weighted correlation function which comes from the spatial distribution of the points, leaving only the contribution from the fluctuations of the marks. Consequently, $M(r)$ is unity for all r if there are no correlations between the marks.

Similarly, the redshift-space marked correlation function is

$$M(s) \equiv \frac{1+W(s)}{1+\xi(s)}\,. \tag{4.2}$$

To obviate complicated redshift-space calculations and redshift distortions in the data, one can measure the *projected* two-point correlation function, which is free of redshift distortions (Davis & Peebles 1983):

$$w_p(r_p) \;=\; \int \mathrm{d}r\, \xi(r_p, \pi) \;=\; 2 \int_{r_p}^{\infty} \mathrm{d}r\, \frac{r\, \xi(r)}{\sqrt{r^2 - {r_p}^2}}, \tag{4.3}$$

where $r = \sqrt{{r_p}^2 + \pi^2}$, and r_p and π are the galaxy separations perpendicular and parallel to the line of sight. For the halo-model description of the redshift-distorted two-point correlation function, $\xi(r_p, \pi)$, see Tinker (2006), who describes it in terms of the pairwise galaxy velocity dispersion and the line-of-sight velocity PDF.

Similarly, the weighted projected correlation function is

$$W_p(r_p) \;=\; 2 \int_{r_p}^{\infty} \mathrm{d}r\, \frac{r\, W(r)}{\sqrt{r^2 - {r_p}^2}}, \tag{4.4}$$

which we use in the marked projected correlation function

$$M_p(r_p) \;=\; \frac{1 \,+\, W_p(r_p)/r_p}{1 \,+\, w_p(r_p)/r_p}\,, \tag{4.5}$$

which makes $M_p(r_p) \approx M(r)$ on scales larger than a few Mpc. It turns out that, in spite of the fact that our redshift-space model describes the unweighted correlation function $\xi(s)$ well and the marked correlation function $M(s)$ adequately, the marked projected correlation function is easier to interpret when comparing to measurements of real galaxy samples (see Figures 6 and 7 of Skibba et al. 2006).

In addition to the marked correlation functions $M(r)$, $M(s)$ and $M_p(r_p)$, which involve the product of marks, one can measure other marked statistics with the mean mark or the mark variance and covariance, and these can yield new information complementing the marked correlation function. One can also use more sophisticated models accounting for the dependence of the mark on the distance from the halo center (see Chapter 8 and Sheth (2005) for details).

The halo model prediction of $M(r)$, $M(s)$ and $M_p(r_p)$ represents the prediction of the 'standard' model: the shape of the weighted correlation function includes the effects of the statistical correlation between halo mass and environment, but no other physical effects.

4.3 HALO MODEL COMPONENTS

In the CDM halo model, the density field is made of dense objects called halos, which have a range of masses and densities approximately 200 times the background density, regardless of their masses. All mass is bound up in halos, and so all galaxies are also associated with halos.

Cooray & Sheth (2002), following Sherrer & Bertschinger (1991), described the model of two-point clustering for a continuous density field, in Fourier space. The two-point correlation function, $\xi(r)$ is determined by the sum of galaxy pairs in the same halo and pairs in separate halos. We refer to these as the 'one-halo term', which dominates at small scales $r < 1\,\mathrm{Mpc}/h$, and the 'two-halo term', which dominates at large scales. The one-halo term is determined by how halo density profiles depend on halo mass (described in Chapter 2), how halo abundances depend on mass, and on how the probability distribution of the number of galaxies occupying halos depends on mass (both described in Chapter 3).

The two-halo term depends on the linear theory power spectrum (described in Chapter 3), and is sensitive to the large-scale clustering of the halos themselves. Mo & White (1996) developed an analytical model for the spatial clustering of halos, and they showed that halo-halo correlations and mass-mass correlations differ by a mass- and redshift-dependent quantity. This is called the 'bias factor' because collapsed halos are biased tracers of the overall dark matter distribution. Although the bias relation depends on the merger histories of halos, it is well approximated given knowledge of the shape of the halo mass function (Sheth & Lemson 1999, Sheth & Tormen 1999). Sheth & Tormen (1999) updated the Press & Schechter mass function to be consistent with numerical dark matter simulations, and Sheth, Mo & Tormen (2001) and Sheth & Tormen (2002) extended it to allow for the ellipsoidal collapse of halos. The ellipsoidal collapse bias relation depends on the critical overdensity required for collapse, the linear growth factor, the variance of the power spectrum as a function of mass, and fitting parameters.

Using all of these halo model components described above, we develop the halo model description of the redshift-space correlation function and marked correlation function in the following section.

4.4 THE HALO MODEL DESCRIPTION OF GALAXY CLUSTERING

This section shows how to describe the unmarked and marked power spectra, which can be Fourier transformed to obtain the unmarked and marked correlation functions. In essence, our formalism combines the results of Sheth (2005) with those of Seljak (2001).

4.4.1 Power Spectrum in Redshift Space

In order to describe the clustering of objects, one first requires the number density of those objects. In the halo model, all mass, including baryonic and dark matter, is bound up in dark matter halos with a range of masses. Hence, the mean number density of galaxies simply is

$$\bar{\rho}_{\rm gal} = \int dm \, \frac{dn(m)}{dm} \, \langle N_{\rm gal}(m) \rangle, \tag{4.6}$$

where $dn(m)/dm$ denotes the number density of halos of mass m, and

$$\langle N_{\rm gal}(m) \rangle \; = \; \sum_N N \, p(N|m) \; = \; \langle N_{\rm cen}|m \rangle \; + \; \langle N_{\rm sat}|m \rangle \tag{4.7}$$

is the first moment of the probability distribution of the number of galaxies occupying halos. The number of central galaxies given m, $\langle N_{\rm cen}|m \rangle$, is usually assumed to be a step function (*i.e.*, unity for $M > M_{\rm min}$), and the number of satellite galaxies, $\langle N_{\rm sat}|m \rangle$, is usually fit with a power-law function of $m/m_{\rm min}$.

The redshift space correlation function is the Fourier transform of the redshift space power spectrum $P(k)$:

$$\xi(s) = \int \frac{dk}{k} \, \frac{k^3 P_z(k)}{2\pi^2} \, \frac{\sin ks}{ks}. \tag{4.8}$$

In the halo model, $P(k)$ is written as the sum of two terms: one that arises from particles within the same halo and dominates on small scales (the 1-halo term), and the other from particles in different halos which dominates on larger scales (the 2-halo term). That is,

$$P^z(k) = P^z_{1h}(k) + P^z_{2h}(k). \tag{4.9}$$

The effects of redshift distortions are different on small scales and large scales; consequently, the redshift-space 1-halo and 2-halo terms are affected differently. There are two

effects that modify the clustering properties of the power spectrum. At large scales, there is a significant boost of power due to the streaming of matter into overdense regions, which compresses those regions and stretches underdense ones along the line of sight (Kaiser 1987). Consequently, the redshift-space perturbation of the galaxy density field is affected by the peculiar velocity perturbations in the following way:

$$\delta_g^z(k) \;=\; b_g(m)\delta_g(k) \;+\; \delta_v\,\mu^2 \quad\;\Rightarrow\quad\; P^z(k) = P(k)\,(b_g + (\delta_v/\delta_g)\,\mu^2)^2, \tag{4.10}$$

where δ_g is the galaxy density contrast, δ_v is the velocity divergence, b_g is the galaxy bias factor, and we define $\mu = \hat{\boldsymbol{r}} \cdot \hat{\boldsymbol{k}}$. At scales in which linear theory is valid, $\delta_v \;=\; f(\Omega_m(z))\,\delta(\boldsymbol{k})$, where $f(\Omega_m) \;\equiv\; \mathrm{d}\ln D_1(a)/\mathrm{d}\ln a \;\approx\; \Omega_m^{0.6}$ is the derivative of the linear growth factor D_1 (Peebles 1980):

$$D_1(a) \;\equiv\; \frac{5}{2}\,\Omega_0\,H_0^2 H(a)\,\int_0^a \frac{\mathrm{d}a'}{(a'\,H(a'))^3}\;.$$

For equation 4.10, the redshift-distorted power spectrum should not be modified by $(1+f\mu^2)^2$ as done by Seljak (2001), or even $(b+f\mu^2)^2$ above, but rather by $(b(m_1)+f\mu^2)(b(m_2)+f\mu^2)$. However, for our purposes, our approximation is sufficiently good, potentially affecting the result by no more than a few percent (Ravi Sheth & Robert Smith, private communication).

On non-linear scales, the virial velocities of galaxies within collapsed halos act as a convolution in redshift space, which results in a small-scale suppression of power (Peacock & Dodds citeyearpd94, Seljak 2001). Since the one-dimensional distribution of virial motions, σ, can be described by a Gaussian (Sheth 1996), then the contrast in the redshift-space density field can be expressed as the following:

$$\delta_g^z(k) \;=\; \delta_g(k)\,e^{-(\mu k\sigma/H)^2/2}\,. \tag{4.11}$$

One can calculate a scale-dependent line-of-sight velocity dispersion within halos with an NFW density profile according to Sheth et al. (2001a), but the following isothermal sphere approximation is usually sufficient (Sheth & Diaferio 2001, Tinker et al. 2006):

$$\sigma^2_{\rm vir}(m) \;\approx\; \frac{G\,m}{2\,r_{\rm vir}(m)} \;=\; G\,(\frac{\pi}{6}m^2\Delta_{\rm vir}\bar{\rho})^{1/3}, \tag{4.12}$$

where $\Delta_{\rm vir} \approx 200$.

By averaging the square of equation 4.10 and radially averaging equation 4.11 over μ (White 2001, Seljak 2001), one can obtain the monopole moment of the redshift-space galaxy power spectrum, extending the real-space power spectrum described by Scoccimarro et al. (2001):

$$\begin{aligned} P_0^z(k) &= \frac{1}{2}\int_{-1}^{+1} \mathrm{d}\mu \left(\frac{\delta_g^z}{\delta_g}(k,\mu)\right)^2 L_0(\mu) \\ &= P^z_{1h,\,\mathrm{gal}}(k) \,+\, P^z_{2h,\,\mathrm{gal}}(k) \\ &\approx \int_{M_{\min}}^{\infty} \mathrm{d}m \,\frac{dn(m)}{dm}\,\frac{\langle N(N-1)|m\rangle}{\overline{\rho}^2_{\mathrm{gal}}}\,\mathcal{R}_p(k\sigma)\,|u_{\mathrm{gal}}(k|m)|^p \qquad (4.13) \\ &\quad + \left(F^2_{\mathrm{gal}}(k) + \frac{2}{3}\,F_{\mathrm{vel}}(k)\,F_{\mathrm{gal}}(k) \,+\, \frac{1}{5}\,F^2_{\mathrm{vel}}(k)\right) P_{\mathrm{lin}}(k) \qquad (4.14) \end{aligned}$$

where

$$F_{\mathrm{gal}}(k) = \int_{M_{\min}}^{\infty} \mathrm{d}m\,\frac{dn(m)}{dm}\,b^2(m)\,\frac{\langle N_{\mathrm{gal}}|m\rangle}{\overline{\rho}_{\mathrm{gal}}}\,\mathcal{R}_1(k\sigma)\,u_{\mathrm{gal}}(k|m) \qquad (4.15)$$

$$F_{\mathrm{vel}}(k) = f(\Omega_m(z))\int_{M_{\min}}^{\infty} \mathrm{d}m\,\frac{dn(m)}{dm}\,b(m)\,\frac{m}{\overline{\rho}}\,\mathcal{R}_1(k\sigma)\,u_{\mathrm{dm}}(k|m) \qquad (4.16)$$

$$\text{and } \mathcal{R}_p(k\sigma) \equiv \int_0^1 \mathrm{d}\mu \exp\left[-\frac{p}{2}\left(k\frac{\sigma_{\mathrm{vir}}(m)}{H(z)}\,\mu\right)^2\right] = \frac{\mathrm{erf}(k\,\sigma_{\mathrm{vir}}(m)\,\sqrt{p/2}/H)}{k\,\sigma_{\mathrm{vir}}(m)\,\sqrt{p/2}/H} \qquad (4.17)$$

is the integral of (4.11) over μ, and $u_{\mathrm{dm}}(k|m)$ is the Fourier transform of the halo density profile divided by the mass m. There is an extra factor of bias in $F_{\mathrm{gal}}(k)$ because of the linear distortions described in (4.10). Note that the expressions in equations 4.13 and 4.14 assume that the effects of non-linear virial motions only affect the 1-halo term and that the effects of the linear flow of matter into overdense regions only affect the 2-halo term. Note also that, although one can often reliably assume that the galaxy density profile $u_{gal}(k|m)$ and $u_{\mathrm{dm}}(k|m)$ are similar, depending on the galaxy properties being studied, it is sometimes necessary to account for the fact that the galaxy profiles are actually luminosity-dependent and even more strongly dependent on color, SFR, and morphology (*e.g.*, Diaferio et al. 2001, Collister & Lahav 2005, Kuehn & Ryden 2005). However, although the density profiles of galaxies may have different concentrations than those of dark matter, any dependence of $u_{gal}(k|m)$ on galaxy type is only important at very small scales of $r < 100\,\mathrm{kpc}/h$, smaller

than we can probe with clustering measurements. The dependence of the halo occupation distribution on galaxy type dominates that of the density profile.

Analyses of galaxy clustering in the halo model framework have clearly shown that central galaxies in halos are special and distinct from the population of satellite galaxies (Chapter 3, Berlind et al. 2005, Zehavi et al. 2005). Central galaxies have different properties, such as blah, than satellite galaxies, on average. Therefore, we must treat the one-halo term as the sum of center-satellite pairs and satellite-satellite pairs. Then for the integral for $P^z_{1h}(k)$ in equation 4.13 we have

$$\begin{aligned} &\langle N(N-1)|m\rangle\, \mathcal{R}_p(k\sigma)\, |u(k|m)|^p \;= \\ &2\,\langle N_{cen}|m\rangle\, \langle N_{sat}|m\rangle\, \mathcal{R}_1(k\sigma)\, u(k|m) \;+\; \langle N_{sat}|m\rangle^2\, \mathcal{R}_2(k\sigma)\, u(k|m)^2\,, \end{aligned} \tag{4.18}$$

where we have assumed that the satellite galaxies have a Poisson distribution and that there is exactly one central galaxy in each halo. In the center-satellite term there is $\mathcal{R}_1$ and in the satellite-satellite term there is $\mathcal{R}_2$ because we assume the central galaxy is at rest with respect to the satellites, which each have some velocity dispersion proportional to the halo's virial radius. Similarly, the center-satellite term has one power of $u(k|m)$ while the satellite-satellite term has two powers of $u(k|m)$ because we assume the central galaxy is at rest and the satellites have an average density profile with respect to the central. This is a realistic assumption, especially since the majority of halos only contain a single 'central' galaxy and no satellites. We find in Chapter 8 that if we focus on group galaxies alone, their clustering is affected by the velocity dispersion of central galaxies at scales of a few hundred kpc, and one must account for this.

In addition, note that one can recover the real-space power spectrum by setting $\sigma_{vir} \to 0$ (which results in $\mathcal{R}(k\sigma) \to 1$) and $F_{vel} \to 0$. When explicit calculations are made, we assume that the density profiles of halos have the form described by Navarro et al. (1996), so u has the form given by Scoccimarro et al. (2001), and that halo abundances and clustering are described by the parameterization of Sheth & Tormen (1999).

Some astrophysicists (*e.g.*, Hamilton 1992, Seljak 2001) have attempted to improve upon the large-scale "streaming model" described above by expanding it in terms of higher-order angular moments of the redshift-space power spectrum. However, as Scoccimarro (2004)

has convincingly argued, this approach is problematic. The large-scale "squashing effect" due to linear dynamics, in $P^z_{2h}(k)$, is not independent of the small-scale "finger-of-god" effect due to pairwise velocity dispersions, in $P^z_{1h}(k)$. Some of the dispersion effect may actually come from large-scale flows, as opposed to virial velocities, and this complicates calculations of the small-scale power suppression. Peacock & Dodds (1994) and White (2001) have attempted to account for this by using the "dispersion model", with $|\delta_g(k)|^2 \rightarrow |\delta_g(k)|^2\, b(k)^2\, (1 + f\mu^2/b)^2 \times exp[-k^2\sigma^2\mu^2/2]$, but this gives rise to an unphysical distribution of pairwise velocities. Moreover, since the pairwise velocity distribution has exponential wings on all scales, not just in the nonlinear regime (Sheth 1996), the pairwise velocity PDF prediction of linear perturbation theory is never a good approximation, even in the large-scale limit. A general and accurate calculation of the redshift-space power spectrum is far more complicated than one with the dispersion model and linear theory (see Scoccimarro 2004, Sections IV-VI). Nonetheless, since inaccuracies of a few percent are allowable in the marked statistics in this paper, our approximation of $P_0(k)$ in equation 4.13 will be sufficient.

4.4.2 Weighted Power Spectrum in Redshift Space

The Fourier transform of the weighted correlation function in redshift-space, $W^z_0(k)$, is similar to $P^z_0(k)$, but with each central and satellite galaxy weighted by a halo mass-dependent mark, $\langle w|m\rangle$, such as luminosity or color. We continue to separate the galaxies into centrals and satellites, and we express the weighted correlation function as the sum of two terms

$$W^z_0(k) \;=\; W^z_{1h,\,\rm gal}(k) \;+\; W^z_{2h,\,\rm gal}(k) \tag{4.19}$$

where, in the one-halo term, we have

$$W^z_{1h,\,\rm gal}(k) \;=\; \int_{M_{\rm min}}^{\infty} {\rm d}m\, \frac{dn(m)}{dm}\, \frac{\langle N(N-1)|m\rangle\, \langle w|m\rangle^2}{\overline{w}^2_{\rm gal}}\, \mathcal{R}_p(k\sigma)\, |u_{\rm gal}(k|m)|^p \tag{4.20}$$

in which the center-satellite and satellite-satellite terms are expressed as

$$\begin{aligned}
&\langle N(N-1)|m\rangle\, \langle w|m\rangle^2\, \mathcal{R}_p(k\sigma)\, |u(k|m)|^p \;= \\
&2\, \langle N_{\rm sat}|m\rangle\, \langle w_{\rm cen}|m\rangle\, \langle w_{\rm sat}|m\rangle\, \mathcal{R}_1(k\sigma)\, u(k|m) \;+\; \langle N_{\rm sat}|m\rangle^2\, \langle w_{\rm sat}|m\rangle^2\, \mathcal{R}_2(k\sigma)\, u(k|m)^2
\end{aligned} \tag{4.21}$$

and so equation 4.20 becomes

$$W^z_{1h,\,\mathrm{gal}}(k) \;=\; \int_{M_{\rm min}}^{\infty} \mathrm{d}m\, \frac{dn(m)}{dm} \left(\frac{2\,\langle N_{\rm sat}|m\rangle\,\langle w_{\rm cen}|m\rangle\,\langle w_{\rm sat}|m\rangle\,\mathcal{R}_1(k\sigma)\,u(k|m)}{\bar{w}^2} + \frac{\langle N_{\rm sat}|m\rangle^2\,\langle w_{\rm sat}|m\rangle^2\,\mathcal{R}_2(k\sigma)\,u(k|m)^2}{\bar{w}^2} \right). \tag{4.22}$$

In the two-halo term, we have

$$W^z_{2h,\,\mathrm{gal}}(k) \;=\; \left(F^2_{\rm w}(k) \;+\; \frac{2}{3}\,F_{\rm vel}(k)\,F_{\rm w}(k) \;+\; \frac{1}{5}\,F^2_{\rm vel}(k) \right) P_{\rm lin}(k) \tag{4.23}$$

where

$$F_{\rm w}(k) \;=\; \int_{M_{\rm min}}^{\infty} \mathrm{d}m\, \frac{dn(m)}{dm}\, b^2(m) \times \frac{\langle w_{\rm cen}|m\rangle \;+\; \langle N_{\rm sat}|m\rangle\,\langle w_{\rm sat}|m\rangle}{\overline{w}_{\rm gal}}\,\mathcal{R}_1(k\sigma)\,u_{\rm gal}(k|m)\,. \tag{4.24}$$

The unweighted $P^z_0(k)$ is normalized by the number density of galaxies $\overline{\rho}_{\rm gal}$, but $W^z_0(k)$ is normalized by the average weighted number density

$$\overline{w}_{\rm gal} \;=\; \int_{M_{\rm min}}^{\infty} \mathrm{d}m\, \frac{dn(m)}{dm} \left(\langle w_{\rm cen}|m\rangle \;+\; \langle N_{\rm sat}|m\rangle\,\langle w_{\rm sat}|m\rangle \right). \tag{4.25}$$

In equations 4.20-4.24, the weight for a central galaxy in an m-halo is $\langle w_{\rm cen}|m\rangle$ and the weight for satellite galaxies in an m-halo is $\langle w_{\rm sat}|m\rangle$. For example, if galaxy luminosity is the mark, then $w_{\rm cen}(m)$ requires a one-to-one relationship between luminosity and halo mass, and $\langle w_{\rm sat}|m\rangle$ is obtained by using this relation with the halo occupation distribution $\langle N_{\rm sat}|m\rangle$. One can estimate the dependence of luminosity or color on halo mass by matching number densities consistently with the galaxy luminosity function in the data and the halo mass function.

In practice, one compares to data from surveys or simulations with magnitude and redshift cuts, which means the average weight of satellites depends on the minimum weight, in the following way

$$\langle w_{\rm sat}|m, w_{\rm min}\rangle \;=\; w_{\rm min}(m_{\rm min}) \;+\; \int_{w_{\rm min}}^{\infty} \mathrm{d}w\, P(>w|m, w_{\rm min}) \tag{4.26}$$

where

$$P(> w, |m, w_{\rm min}) \equiv \frac{\langle N_{\rm sat}|m, w\rangle}{\langle N_{\rm sat}|m, w_{\rm min}\rangle} \tag{4.27}$$

In addition, since both $w_{\rm cen}(m)$ and $\langle w_{\rm sat}|m, w_{\rm min}\rangle$ can be obtained from the data, the mean weight of all galaxies in an m-halo (central + satellites),

$$\langle w_{\rm tot}|m, w_{\rm min}\rangle = \frac{\langle w_{\rm cen}|m\rangle + \langle N_{\rm sat}|m, w_{\rm min}\rangle \langle w_{\rm sat}|m, w_{\rm min}\rangle}{1 + \langle N_{\rm sat}|m, w_{\rm min}\rangle} \tag{4.28}$$

is completely determined by the halo occupation distribution. Our analysis shows that, in low mass halos, $w_{\rm tot} \approx w_{\rm cen}$ because $N_{\rm sat} \ll 1$, whereas in massive halos, $w_{\rm tot} < w_{\rm cen}$. The different quantities scale very differently with halo mass. If the central galaxies are *not* treated as special, then the contribution to the one-halo term would scale as $N_{\rm sat} w_{\rm tot}^2$ for the center-satellite term and $(N_{\rm sat} w_{\rm tot})^2$ for the satellite-satellite term. Consequently, marked statistics allow one to discriminate between models which treat the central object as special from models which do not.

As the above formalism makes clear, marked correlation functions significantly depend on the weighting scheme, the halo occupation distribution and its dependence on the mark, and on the $w_{\rm min}(m_{\rm min})$ relation. It is important to note that the uncertainties in cosmological parameters also affect the resultant marked statistics, and for the main parameters involved (Ω_m, Λ, σ_8) consensus has not been reached about their precise values (*e.g.*, Sánchez et al. 2006).

4.5 MARKED CORRELATIONS USING THE MEAN MARK

Another useful marked statistic is the mean mark, rather than the mark product that the marked correlation function measures. Analogous to equation 4.1, the mean mark as a function of pair separation is defined by

$$\begin{aligned} M_1(r) &\equiv \frac{\sum [w(\boldsymbol{x}) + w(\boldsymbol{y})]\, \mathcal{I}(|\boldsymbol{x}-\boldsymbol{y}| - r)}{2\bar{w} \sum \mathcal{I}(|\boldsymbol{x}-\boldsymbol{y}| - r)} \\ &= \frac{\sum [w(\boldsymbol{x}) + w(\boldsymbol{y})]\, \mathcal{I}(|\boldsymbol{x}-\boldsymbol{y}| - r)}{2\bar{w}\, \bar{\rho}^2 [1 + \xi(r)]}, \end{aligned} \tag{4.29}$$

where $\mathcal{I}(x) = 0$ unless $x = 0$, and the sum is over all galaxy pairs. We have normalized by unity, so $M_1(r)$ is unity for all r if there are no correlations between the marks.

The 1-halo and 2-halo terms of the mean-mark weighted correlation function in redshift space, analogous to equations 4.20-4.24, are consequently the following

$$W^z_{1,\,1h}(k) \;=\; \int_{M_{\rm min}}^{\infty} {\rm d}m\, \frac{dn(m)}{dm} \left(\frac{2\,\langle N_{\rm sat}|m\rangle\,(\langle w_{\rm cen}|m\rangle + \langle w_{\rm sat}|m\rangle)\,\mathcal{R}_1(k\sigma)\,u(k|m)}{2\,\bar{w}_{\rm gal}\,\bar{\rho}_{\rm gal}} + \frac{\langle N_{\rm sat}|m\rangle\,\langle w_{\rm sat}|m\rangle\,\mathcal{R}_2(k\sigma)\,u(k|m)^2}{\bar{w}_{\rm gal}\,\bar{\rho}_{\rm gal}} \right). \tag{4.30}$$

Just like equation 4.23, the 2-halo term in redshift space is

$$W^z_{1,\,2h}(k) \;=\; \left(F^2_{\rm w}(k) \;+\; \frac{2}{3}\,F_{\rm vel}(k)\,F_{\rm w}(k) \;+\; \frac{1}{5}\,F^2_{\rm vel}(k) \right) P_{\rm lin}(k)$$

but now the weighted component is

$$F^2_{\rm w}(k) \;=\; \left[\int_{M_{\rm min}}^{\infty} {\rm d}m\, \frac{dn(m)}{dm}\, b^2(m) \times \frac{\langle w_{\rm cen}|m\rangle \;+\; \langle N_{\rm sat}|m\rangle\,\langle w_{\rm sat}|m\rangle}{\overline{w}_{\rm gal}}\,\mathcal{R}_1(k\sigma)\,u(k|m) \right] F_{\rm gal}(k)\,. \tag{4.31}$$

We can also calculate the projected mean mark, $M_1(r_p) \equiv (1 + W_1/r_p)/(1 + w_p/r_p)$, with

$$W_1(r_p) \;=\; 2\int_{r_p}^{\infty} {\rm d}r\, \frac{r\,W_1(r)}{\sqrt{r^2 - {r_p}^2}}\,. \tag{4.32}$$

An example of the projected mean mark is shown in Figure 4.1, for the $M_r < -20.5$ volume-limited SDSS sample used in Skibba et al. (2006) with the r-band luminosity mark. The solid curve in the lower panel shows the result when we model the central and satellite galaxy marks separately and the dashed curve is the result when we do not distinguish between them. Except for the smallest scale point, there is good agreement between the center-satellite halo model and the data. This, complementing the results in Chapter 5, suggests that the correlation between galaxy luminosity and large-scale environment is actually driven by the correlation between halo mass and environment.

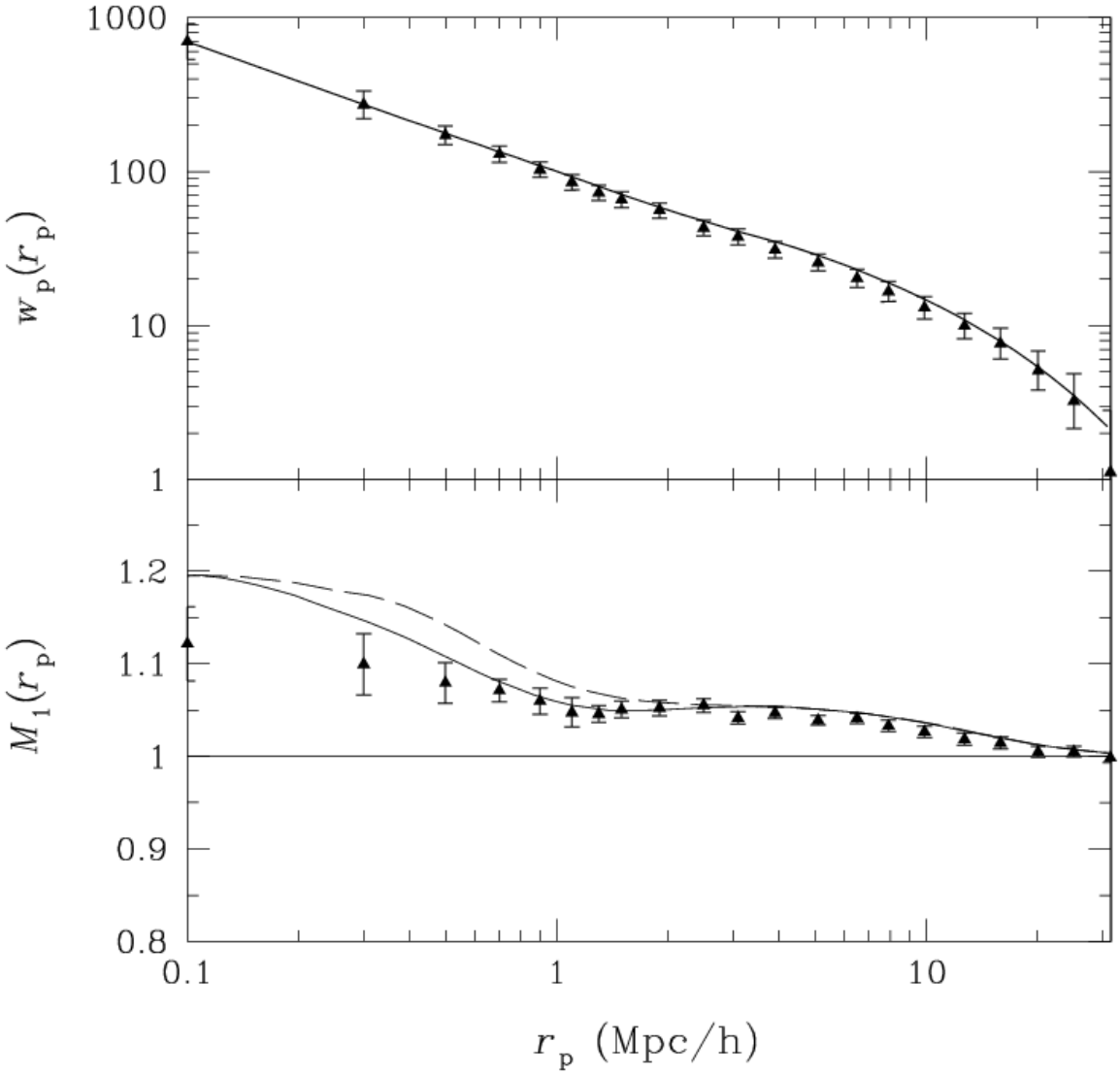

Figure 4.1: Projected mean mark of $M_r < -20.5$ volume-limited SDSS sample, compared to halo-model prediction assuming the Zehavi et al. 2005 HOD and Blanton et al. 2003a r-band LF.

4.6 ENVIRONMENTAL SENSITIVITY OF MARKED STATISTICS

We end by demonstrating how one can quantify the environmental sensitivity of marked statistics. This serves a very important purpose, to test whether a measured marked statistic is actually to some extent due to the correlation between halo mass and environment, or alternatively whether the measured marked signal is due to some other kind of environmental dependence and the consistency with the halo model is merely a coincidence.

In particular, we *impose* an environmental dependence on the luminosities in our $M_r < -20.5$ SDSS catalog and repeat the measurement of the marked two-point correlation function. Ummi Abbas kindly ran her neighbor-counting algorithm (described in Abbas & Sheth 2005; 2006) on our dataset, which counts the number of neighbors around each galaxy within a particularly-sized sphere. Spheres of 8 Mpc/h radii were chosen in order to attempt to account for "finger-of-god" redshift distortions. One could perhaps use a more sophisticated algorithm with ellipsoids or cylinders, with larger line-of-sight separations than projected separations, along the lines of the algorithm used by Berlind et al. (2006), but that is not necessary for our illustrative example here.

The number of galaxies withing 8 Mpc/h spheres is a proxy for environment because $N \propto 1 + \delta$, where dense regions have $\delta > 0$. To impose on environmental dependence, we modify our luminosity marks by multiplying by the number of galaxies within 8 Mpc/h to some power, which is equivalent to $L \rightarrow (1+\delta)^{\epsilon} L$. (Note that this means that galaxies with no neighbors now have weights of 'zero'.) If the marked statistic with these modified marks is significantly different than the original measurement, then this implies that the marked statistic is sensitively probing the environmental dependence of the mark.

Marked statistics are sensitive to the distribution of marks relative to the mean mark, and the correlation with environment we are imposing could alter the distribution. To ensure that our marked correlation function measurements will indicate the environmental sensitivity of the marked correlations and not merely different mark distributions, we compare the distributions in Figure 4.2. It is evident that the original distribution of r-band luminosities in our $M_r < -20.5$ catalog (black histogram) is virtually the same as that of the luminosities modified by $(1+\delta)^{\epsilon}$ with $\epsilon = 0.01$ and 0.05.

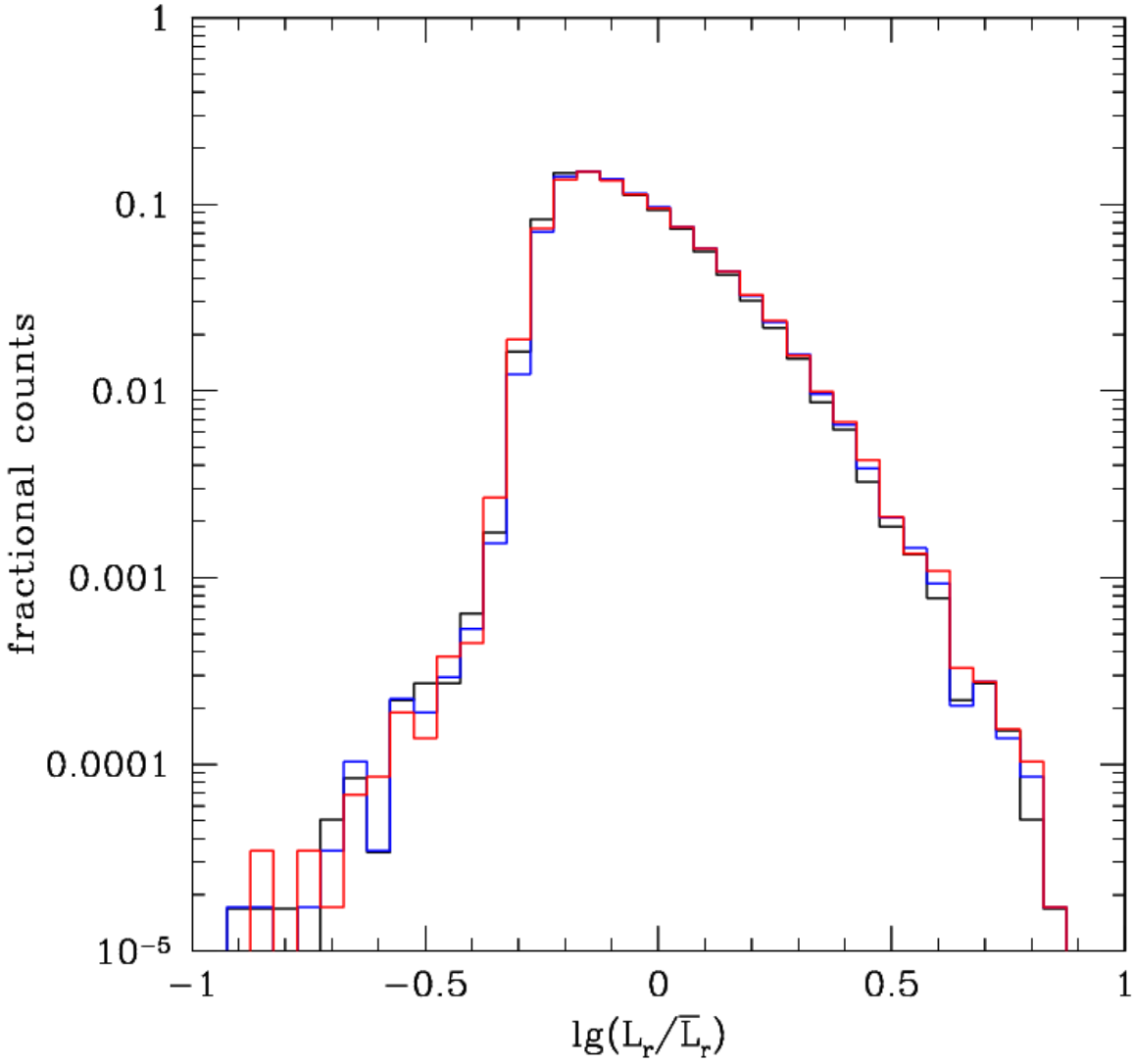

Figure 4.2: Distribution of r-band luminosities in the $M_r < -20.5$ SDSS volume-limited catalog. Original distribution (black histogram) is compared to the luminosities modified by $(1+\delta)^{0.01}$ (blue histogram) and by $(1+\delta)^{0.05}$ (red histogram).

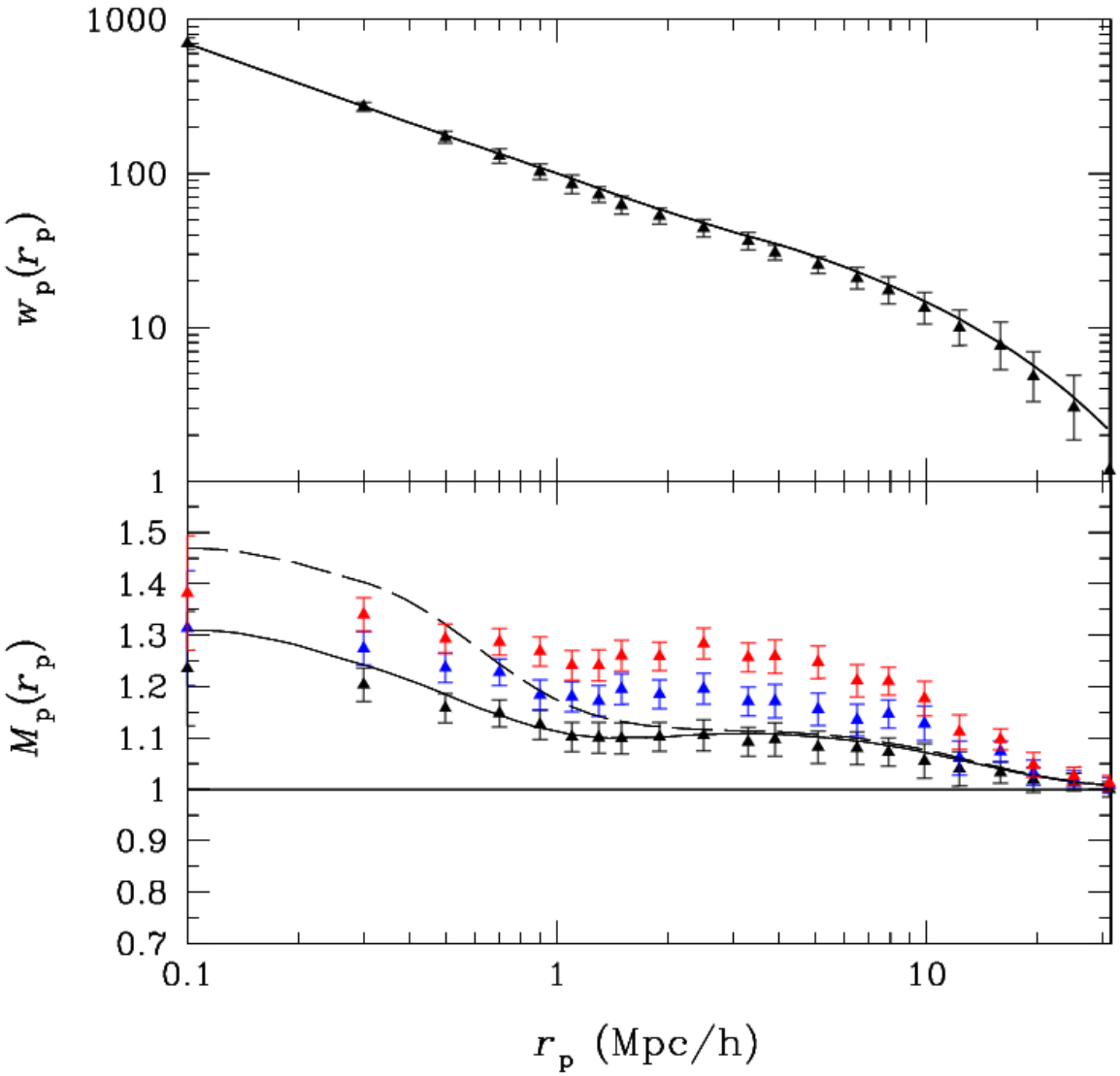

Figure 4.3: Projected luminosity-marked correlation function for $M_r < -20.5$ SDSS galaxies (black points). Measurement with luminosities modified by $(1+\delta)^{0.01}$ (blue points) and by $(1+\delta)^{0.05}$ (red points).

The measured luminosity-marked projected correlation function of the $M_r < -20.5$ catalog is compared to the measurements with the luminosities modified using $\epsilon = 0.01$ and 0.05 in Figure 4.3. As can be seen, even a tiny additional environmental dependence (blue points) results in a significantly different marked correlation function. The projected luminosity-marked correlation function is clearly *extremely* sensitive to environment. Future work is required to carefully assess the environmental sensitivity of different marked statistics and different marks.

4.7 CONCLUSION

In this paper we have developed the halo-model description of the marked correlation function and mean mark in redshift space and projected space. These are important because marked statistics are most instructive when comparing halo-model predictions to real datasets, in which the positions of galaxies are affected by small-scale and large-scale redshift distortions. The marked correlation function and mean mark, in addition to complementary marked statistics, are useful for identifying and quantifying the correlations between galaxy properties and large-scale environment, without having to make any contentious assumptions about environmental indicators. We have also provided a test showing that marked statistics are indeed sensitive to environment, and are hence useful tools for probing aspects of the environmental dependence of halo and galaxy formation.

5.0 LUMINOSITY-MARKED CORRELATION FUNCTIONS OF SDSS GALAXIES

5.1 INTRODUCTION

In hierarchical models of structure formation, there is a correlation between halo formation and abundances and the surrounding large scale structure—the mass function in dense regions is top-heavy (Mo & White 1996; Sheth & Tormen 2002). Galaxy formation models assume that the properties of a galaxy are determined entirely by the mass and formation history of the dark matter halo within which it formed. Thus, the correlation between halo properties and environment induces a correlation between galaxy properties and environment. The main goal of the present work is to test if this statistical correlation accounts for most of the observed trends between luminosity and environment (luminous galaxies are more strongly clustered), or if other physical effects also matter.

We do so by using the statistics of marked correlation functions (Stoyan & Stoyan 1994, Beisbart & Kerscher 2000) which have been shown to provide sensitive probes of environmental effects (Sheth & Tormen 2004; Sheth, Connolly & Skibba 2005). The halo model (see Cooray & Sheth 2002 for a review) is the language currently used to interpret measurements of galaxy clustering. Sheth (2005) develops the formalism for including marked correlations in the halo model of clustering, and we extend this in Chapter 4 to describe measurements made in redshift space. This halo model provides an analytic description of marked statistics when correlations with environment arise entirely because of the statistical effect.

Section 5.2 describes how to construct a mock galaxy catalog in which the luminosity function and the luminosity dependence of clustering are the same as those observed in the Sloan Digital Sky Survey (SDSS). In these mock catalogs, any correlation with environment

is *entirely* due to the statistical effect. Section 5.3 shows that the halo model description of marked statistics provides a good description of this effect, both in real and in redshift space. Section 5.4 compares measurements of marked statistics in the SDSS with the halo model prediction. The comparison provides a test of the assumption that correlations with environment arise entirely because of the statistical effect. A final section summarizes our results, and shows that marked statistics provide interesting information about the correlation between galaxies and their environments without having to separate the population into the two traditional extremes of 'cluster' and 'field'.

5.2 WEIGHTED OR MARKED CORRELATIONS IN THE 'STANDARD' MODEL

Zehavi et al. (2005) have measured the luminosity dependence of clustering in the SDSS (York et al. 2000; Adelman-McCarthy et al. 2006). They interpret their measurements using the language of the halo model (see Cooray & Sheth 2002 for a review). In particular, they describe how the distribution of galaxies depends on halo mass in a ΛCDM model with $(\Omega_0, h, \sigma_8) = (0.3, 0.7, 0.9)$ which is spatially flat. In this description, only sufficiently massive halos ($M_{\rm halo} > 10^{11} M_\odot$) host galaxies. Each sufficiently massive halo hosts a galaxy at its center, and may host satellite galaxies. The number of satellites follows a Poisson distribution with a mean value which increases with halo mass (following Kravtsov et al. 2004). In particular, Zehavi et al. report that the mean number of galaxies with luminosity greater than L in halos of mass M is

$$N_{\rm gal}(>L|M) = 1 + N_{\rm sat}(>L|M) = 1 + \left[\frac{M}{M_1(L)}\right]^{\alpha(L)} \tag{5.1}$$

if $M \geq M_{\rm min}(L)$, and $N_{\rm gal}(M) = 0$ otherwise. In practice, $M_{\rm min}(L)$ is a monotonic function of L; we have found that their results are quite well approximated by

$$\left(\frac{M_{\rm min}}{10^{12} h^{-1} M_\odot}\right) \approx \exp\left(\frac{L}{9.9 \times 10^9 h^{-2} L_\odot}\right) - 1, \tag{5.2}$$

$M_1(L) \approx 23\, M_{\rm min}(L)$, and $\alpha \sim 1$.

Later in this paper we will also study a parametrization in which the cutoff at $M_{\rm min}$ is less abrupt:

$$\begin{aligned} N_{\rm gal}(>L|M) &= {\rm erfc}\left[\frac{\log_{10} M_{\rm min}(L)/M}{\sqrt{2}\sigma}\right] + N_{\rm sat}(>L|M) \\ N_{\rm sat}(>L|M) &= \left[\frac{M}{M_1(L)}\right]^{\alpha(L)} . \end{aligned} \tag{5.3}$$

This is motivated by the fact that semi-analytic galaxy formation models show smoother cutoffs at low-masses (Sheth & Diaferio 2001; Zheng et al. 2005), and that parameterizations like this one can also provide good fits to the SDSS measurements (Zehavi et al. 2005).

We use the model in equation (5.1) to populate halos in the $z = 0.13$ outputs of the VLS ΛCDM simulation (Yoshida, Sheth & Diaferio 2001) as follows. We specify a minimum luminosity $L_{\rm min}$ which is smaller than the minimum luminosity we wish to study. We then select the subset of halos in the simulations which have $M > M_{\rm min}(L_{\rm min})$. We specify the number of satellites each such halo contains by choosing an integer from a Poisson distribution with mean $N_{\rm sat}(> L_{\rm min}|M)$. We then specify the luminosity of each satellite galaxy by generating a random number u distributed uniformly between 0 and 1, and finding that L for which $N_{\rm sat}(> L|M)/N_{\rm sat}(> L_{\rm min}|M) = u$. This ensures that the satellites have the correct luminosity distribution. Finally, we distribute the satellites around the halo center so that they follow an NFW profile (see Scoccimarro & Sheth 2002 for details). We also place a central galaxy at the center of each halo. The luminosity of this central galaxy is given by inverting the $M_{\rm min}(L)$ relation between minimum mass and luminosity. We assign redshift space coordinates to the mock galaxies by assuming that a galaxy's velocity is given by the sum of the velocity of its parent halo plus a virial motion contribution which is drawn from a Maxwell-Boltzmann distribution with dispersion which depends on halo mass (following equation 5.12 below). We insure that the center of mass motion of all the satellites in a halo is the same as that of the halo itself by subtracting the mean virial motion vector of satellites from the virial motion of each satellite (see Sheth & Diaferio 2001 for tests which indicate that this model is accurate).

The resulting mock galaxy catalog has been constructed to have the correct luminosity function (Figure 5.1) as well as the correct luminosity dependence of the galaxy two-point

correlation function. In addition, note that the number of galaxies in a halo, the spatial distribution of galaxies within a halo, and the assignment of luminosities all depend only on halo mass, and not on the surrounding large-scale structure. Therefore, the mock catalog includes only those environmental effects which arise from the environmental dependence of halo abundances.

For reasons described by Sheth, Connolly & Skibba (2005), the marked correlation function we measure in the mock catalogs is

$$M(s) \equiv \frac{1 + W(s)}{1 + \xi(s)}, \tag{5.4}$$

where $\xi(s)$ is the two-point correlation function in redshift space, and $W(s)$ is the same sum over galaxy pairs separated in redshift space by s, but now each member of the pair is weighted by the ratio of its luminosity to the mean luminosity of all the galaxies in the mock catalog. (Schematically, if the estimator for $1 + \xi$ is DD/RR, then the estimator for $1 + W$ is WW/RR, so the estimator we use for M is WW/DD.) This measurement of $M(s)$ represents the prediction of the 'standard' model: the shape of the luminosity weighted correlation function includes the effects of the statistical correlation between halo mass and environment, but no other physical effects.

5.3 THE HALO MODEL DESCRIPTION

This section shows how to describe the marked correlation in redshift space discussed above in the language of the halo model. Details are in Chapter 4;in essence, the calculation combines the results of Sheth (2005) with those of Seljak (2001).

In the halo model, all mass is bound up in dark matter halos which have a range of masses. Hence, the density of galaxies is

$$\bar{n}_{\rm gal} = \int dm \, \frac{dn(m)}{dm} \, N_{\rm gal}(m), \tag{5.5}$$

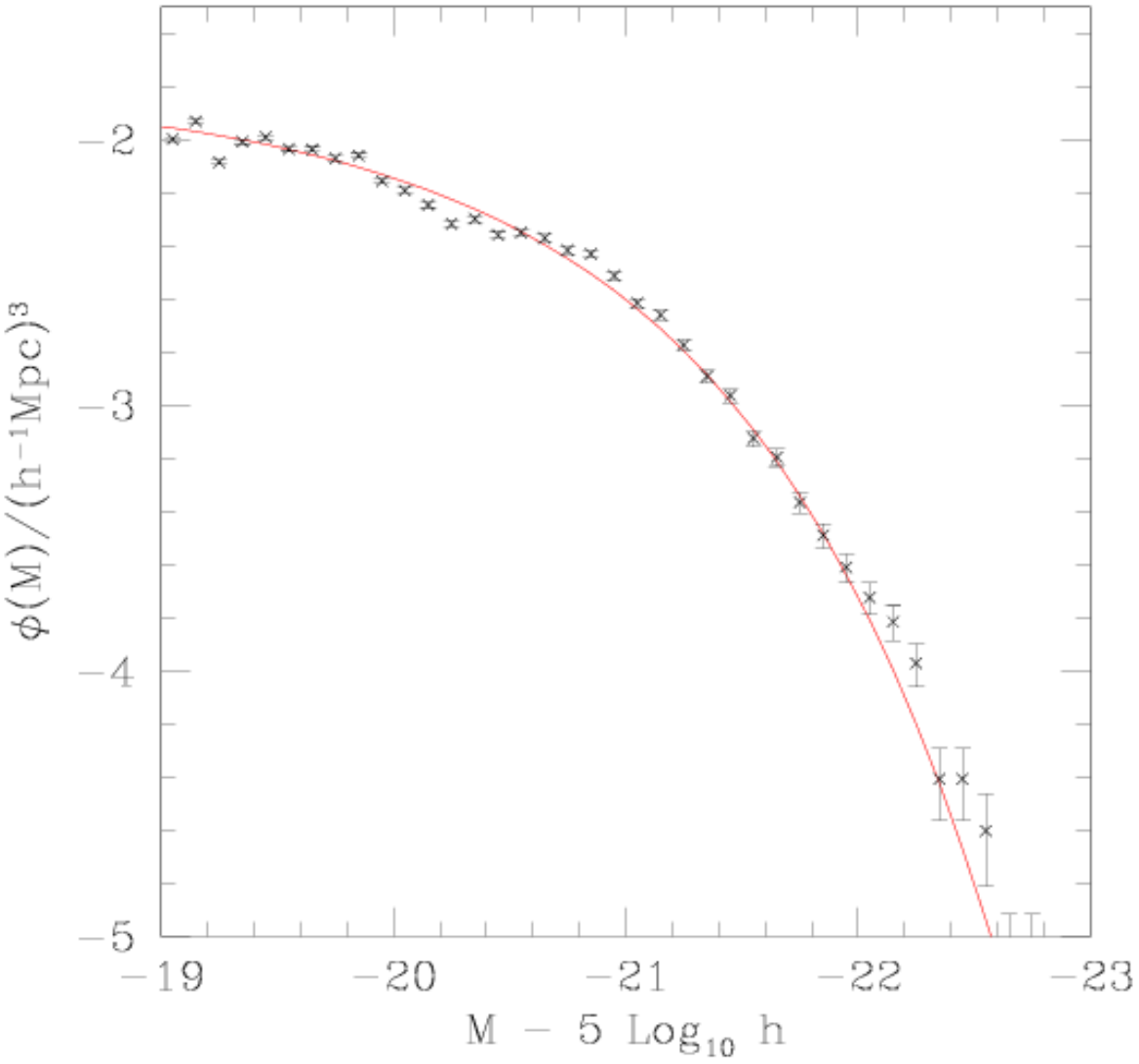

Figure 5.1: Luminosity function in the mock catalog (symbols with error bars); M refers to the absolute magnitude in the $r-$band. Smooth curve shows the SDSS luminosity function (Blanton et al. 2003a).

where $dn(m)/dm$ denotes the number density of halos of mass m. The redshift space correlation function is the Fourier transform of the redshift space power spectrum $P(k)$:

$$\xi(s) = \int \frac{dk}{k} \frac{k^3 P(k)}{2\pi^2} \frac{\sin ks}{ks}. \tag{5.6}$$

In the halo model, $P(k)$ is written as the sum of two terms: one that arises from particles within the same halo and dominates on small scales (the 1-halo term), and the other from particles in different halos which dominates on larger scales (the 2-halo term). Namely,

$$P(k) = P_{1h}(k) + P_{2h}(k), \tag{5.7}$$

where, in redshift space,

$$\begin{aligned}
P_{1h}(k) &= \int dm \frac{dn(m)}{dm} \left[\frac{2N_{\rm sat}(m)\, u_1(k|m)}{\bar{n}^2_{\rm gal}} \right. \\
&\qquad \left. + \frac{N^2_{\rm sat}(m)\, u^2_2(k|m)}{\bar{n}^2_{\rm gal}} \right], && (5.8) \\
P_{2h}(k) &= \left(F^2_{\rm g} + \frac{2F_{\rm g}F_{\rm v}}{3} + \frac{F^2_{\rm v}}{5} \right) P_{\rm Lin}(k), && (5.9) \\
u_1(k|m) &= \left[\frac{\sqrt{\pi}}{2} \frac{{\rm erf}(k\sigma_{\rm vir}(m)/\sqrt{2}H)}{k\sigma_{\rm vir}(m)/\sqrt{2}H} \right] u(k|m), && (5.10) \\
u^2_2(k|m) &= \left[\frac{\sqrt{\pi}}{2} \frac{{\rm erf}(k\sigma_{\rm vir}(m)/H)}{k\sigma_{\rm vir}(m)/H} \right] u^2(k|m), && (5.11)
\end{aligned}$$

$u(k|m)$ is the Fourier transform of the halo density profile divided by the mass m, H is the Hubble constant, and

$$\sigma^2_{\rm vir}(m) \approx \frac{Gm}{2r_{\rm vir}} = G \left(\frac{\pi}{6} m^2 \Delta_{\rm vir} \bar{\rho} \right)^{1/3} \tag{5.12}$$

is the line-of-sight velocity dispersion within a halo ($\Delta_{\rm vir} \approx 200$). In addition, the bias factor $b(m)$ describes the strength of halo clustering,

$$\begin{aligned}
F_{\rm v} &= f \int {\rm d}m \frac{{\rm d}n(m)}{{\rm d}m} \frac{m}{\bar{\rho}}\, u_1(k|m)\, b(m), && (5.13) \\
F_{\rm g} &= \int {\rm d}m \frac{{\rm d}n(m)}{{\rm d}m} \frac{1 + N_{\rm sat}(m) u_1(k|m)}{\bar{n}_{\rm gal}}\, b(m), && (5.14)
\end{aligned}$$

$f \equiv d\ln D(a)/d\ln a \approx \Omega^{0.6}$, and $P_{\rm Lin}(k)$ is the power spectrum of the mass in linear theory. The real space power spectrum is given by setting the terms in square brackets in equations (5.10) and (5.11) for u_1 and u_2 to unity, and $F_{\rm v} \to 0$. When explicit calculations are made, we assume that the density profiles of halos have the form described by Navarro et al. (1996), so u has the form given by Scoccimarro et al. (2001), and that halo abundances and clustering are described by the parameterization of Sheth & Tormen (1999).

To describe the effect of weighting each galaxy by its luminosity, let $W(r)$ denote the weighted correlation function, and $\mathcal{W}(k)$ its Fourier transform. Following Sheth, Abbas & Skibba (2004) and Sheth (2005), we write this as the sum of two terms:

$$\mathcal{W}(k) = \mathcal{W}_{1h}(k) + \mathcal{W}_{2h}(k), \tag{5.15}$$

where

$$\begin{aligned}
\mathcal{W}_{1h}(k) &= \int dm\, \frac{dn(m)}{dm} \\
&\quad \times \left[\frac{2L_{\rm cen}(m)\,\langle L|m, L_{\rm min}\rangle\, N_{\rm sat}(m)\, u_1(k|m)}{\bar{n}_{\rm gal}^2\ \bar{L}^2} \right. \\
&\quad\quad \left. + \frac{\langle L|m, L_{\rm min}\rangle^2 N_{\rm sat}^2(m)\, u_2^2(k|m)}{\bar{n}_{\rm gal}^2\ \bar{L}^2} \right], \\
\mathcal{W}_{2h}(k) &= \left(F_{\rm w}^2 + \frac{2F_{\rm w}F_{\rm v}}{3} + \frac{F_{\rm v}^2}{5} \right) P_{\rm Lin}(k),
\end{aligned}$$

with

$$\begin{aligned}
F_{\rm w} &= \int {\rm d}m\, \frac{{\rm d}n(m)}{{\rm d}m}\, b(m) \\
&\quad \times \frac{L_{\rm cen}(m) + N_{\rm sat}(m)\langle L|m, L_{\rm min}\rangle u_1(k|m)}{\bar{n}_{\rm gal}\ \bar{L}}
\end{aligned} \tag{5.16}$$

and

$$\bar{L} = \int dm\, \frac{dn(m)}{dm}\, \frac{L_{\rm cen}(m) + N_{\rm sat}(m)\,\langle L|m, L_{\rm min}\rangle}{\bar{n}_{\rm gal}}\,. \tag{5.17}$$

Here $\bar{L}$ is the average luminosity, $L_{\rm cen}(m)$ is the luminosity of the galaxy at the center of an m-halo, and $\langle L|m, L_{\rm min}\rangle$ is the average luminosity of satellite galaxies more luminous than $L_{\rm min}$ in m-halos. Thus, the calculation requires an estimate of how the central and the average satellite luminosity depend on m. As we show below, both are given by the luminosity

dependence of ξ (i.e., equation 5.1), so this halo model calculation of the weighted correlation function requires *no* additional information!

The luminosity of the central galaxy is obtained by inverting the relation between $M_{\rm min}$ and L (e.g., equation 5.2). Obtaining an expression for the average luminosity of a satellite galaxy is more complicated. Define

$$\begin{aligned} P(>L|m,L_{\rm min}) &\equiv \frac{N_{\rm sat}(>L|m)}{N_{\rm sat}(>L_{\rm min}|m)} \\ &= \int_L^\infty dL\, p(L|m,L_{\rm min}), \end{aligned} \tag{5.18}$$

where $N_{\rm sat}$ is given by equation (5.1). Then the mean luminosity of satellites in m halos,

$$\langle L|m,L_{\rm min}\rangle = \int_{L_{\rm min}}^\infty dL\, p(L|m,L_{\rm min})\, L, \tag{5.19}$$

can be obtained from the fact that

$$\int_{L_{\rm min}}^\infty dL'\, P(>L'|m,L_{\rm min}) \;=\; \langle L|m,L_{\rm min}\rangle - L_{\rm min}. \tag{5.20}$$

This shows that if we add $L_{\rm min}$ to the quantity on the left hand side (which is given by integrating equation 5.1 over L), we will obtain the quantity we are after.

Incidentally, since both $L_{\rm cen}(m)$ and $\langle L|m,L_{\rm min}\rangle$ can be estimated from the SDSS fits, the mean luminosity of the galaxies in an m-halo,

$$L_{\rm av}(m,L_{\rm min}) = \frac{L_{\rm cen}(m) + N_{\rm sat}(m,L_{\rm min})\langle L|m,L_{\rm min}\rangle}{1+N_{\rm sat}(m,L_{\rm min})}, \tag{5.21}$$

is completely determined by equation (5.1). The mass-to-light ratio of an m halo is $m/[N_{\rm gal}(m)\,L_{\rm av}(m)]$: this shows explicitly that the luminosity dependence of the galaxy correlation function constrains how the halo mass-to-light ratio must depend on halo mass. This halo-mass dependence has been used by Tinker et al. (2005); our analysis provides an analytic calculation of the effect. It shows that, in low mass halos, $L_{\rm av} \approx L_{\rm cen}$ because $N_{\rm sat} \ll 1$, whereas in massive halos, $L_{\rm av} < L_{\rm cen}$. Figure 5.2 compares the mass dependence of $L_{\rm av}$, $L_{\rm cen}$, and $L_{\rm sat}$ for galaxies restricted to $M_r < -20.5$ as predicted by Zehavi et al.'s (2005) halo model interpretation of the luminosity dependence of clustering in the SDSS.

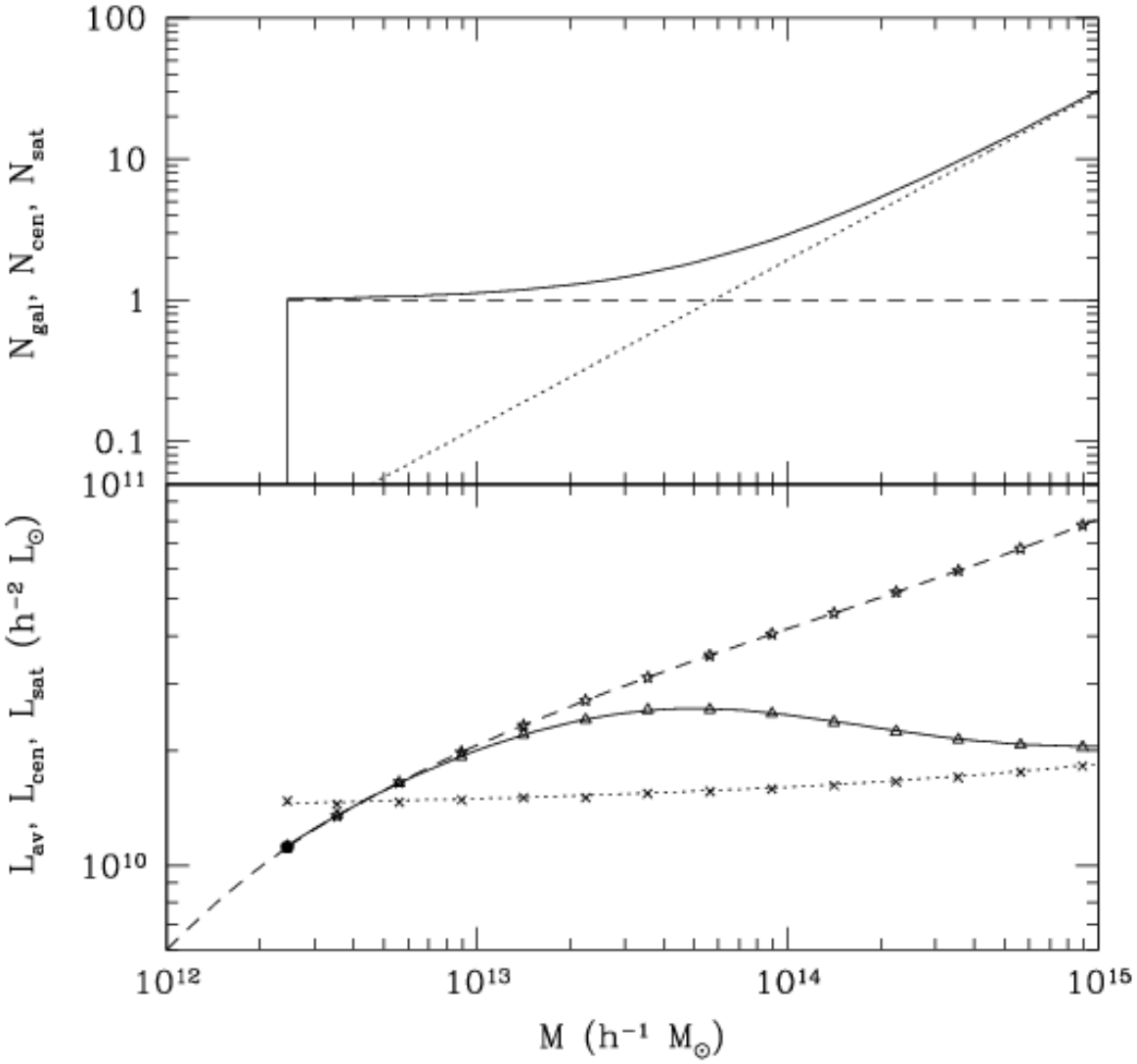

Figure 5.2: Mean number of galaxies in a halo (top) and mean luminosity in a halo (bottom) for SDSS galaxies with $M_r < -20.5$, as a function of the masses of their parent halos, predicted by the luminosity dependence of clustering. Different curves in bottom panel show the mean luminosity of the galaxies in a halo, the luminosity of the central galaxy, and the mean luminosity of the others, as a function of halo mass (solid, dashed, and dotted curves). Symbols show the result of computing these relations in our mock catalogs.

The symbols show measurements from our mock catalogs. The different quantities scale very differently with halo mass, with the following consequence.

Equation (5.15) treats the central galaxies differently from the others. If the luminosities of the central galaxies were not special, then the contribution to the one-halo term would scale as $N_{\rm sat}L_{\rm av}^2$ for the center-satellite term, and $(N_{\rm sat}L_{\rm av})^2$ for the satellite-satellite term. Note that, in this case, the luminosity weights are the same for the two types of terms—only the number weighting differs. However, because the mass dependence of $L_{\rm av}$ is so different from that of the other two terms (cf. Figure 5.2), marked statistics allow one to discriminate between models which treat the central object as special from models which do not (e.g. Sheth 2005).

To illustrate, the symbols in the top and bottom panels of Figure 5.3 show measurements of $\xi(r)$ and $M(r)$ measured in this mock catalog. Error bars were obtained with a "jackknife" procedure, as detailed in Scranton et al. (2002), in which the statistic is re-measured after omitting a random region, and repeating thirty times ($\sim$1.5 times the number of bins in separation for which we present results). Note that the errors in W are strongly correlated with those in ξ, so that the true error in M is grossly (more than a factor of ten) overestimated if one simply sums these individual errors in quadrature. A much better approximation of the uncertainties is obtained as follows. Randomly scramble the marks among the galaxies, remeasure M, and repeat many (~ 100) times. Compute the mean of M over these realizations. This mean, and the rms scatter around it are shown as dotted lines in the two panels. Note that this scatter is within a factor of two of the full jackknife error estimate; it is smaller than the jackknife estimate on scales $r > 1h^{-1}$Mpc and $s > 3h^{-1}$Mpc, and, on smaller scales, it is larger than the jackknife estimate.

The solid lines in the top panel show the halo model calculation of ξ. These show that the model is in excellent agreement with the measurements on all scales in real space. The solid and dashed curves in the bottom panel show the associated halo model calculations of the marked statistic M when central galaxies are special (solid), and when they are not (dashed). Note that both these curves give the *same* prediction for the unweighted statistic ξ.

Comparison of these curves with the measurements yields two important pieces of infor-

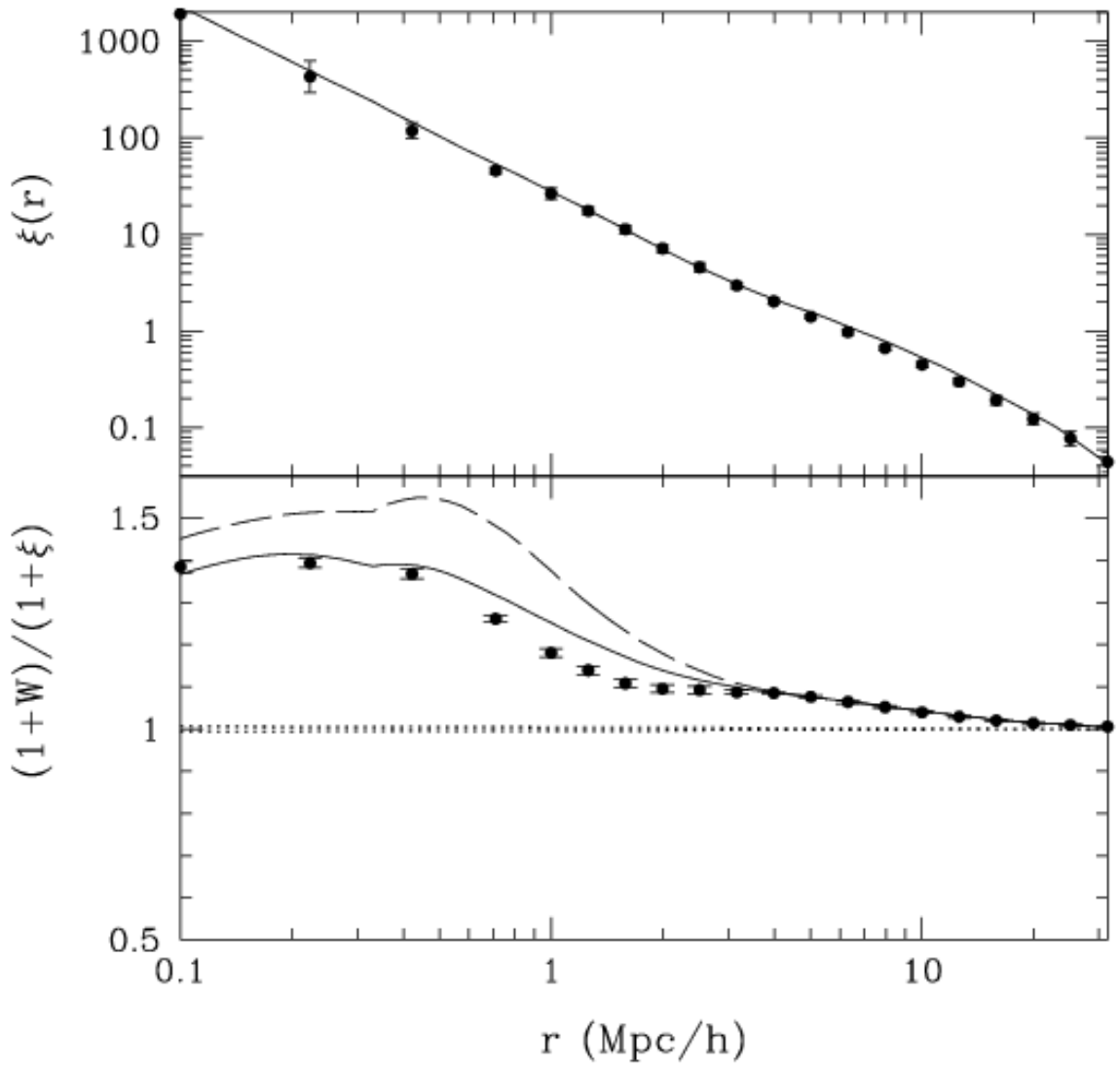

Figure 5.3: Luminosity-weighted real-space correlation function measured in a mock catalogs which resembles an SDSS volume limited sample with $M_r < -20.5$. Symbols show the measurements; smooth curves show the associated halo-model predictions when the luminosity of the central galaxy in a halo is assumed to be different from the others; dashed curves show the prediction when the central object is not special. Dotted curves show the mean and rms values of the statistic M, obtained by randomizing the marks and remeasuring M one hundred times.

mation. First, on large scales ($r > 4h^{-1}$Mpc), the solid and dashed curves are identical, and they are in excellent agreement with the measurements. This indicates the large-scale signal is well described by a model in which there are no additional correlations with environment other than those which arise from the correlation between halo mass and environment. This is not reassuring, since the mock catalogs were constructed to have no correlations other than those which are due to halo bias. Second, on smaller scales, the solid curves are in substantially better agreement with the measurements than are the dashed curves. (A χ^2 estimate of the goodness of fit of the two marked correlation models yields values which smaller by a factor of ten when the central galaxy is treated specially compared to when it is not.) Since the mock catalogs do treat the central galaxies differently from the others, it is reassuring that the halo model calculation which incorporates this difference is indeed in better agreement with the measurements.

In the next section, we will present measurements of marked statistics in redshift space. To see if we can use our halo model calculation to interpret the measurements, Figure 5.4 compares measurements of $\xi(s)$ and $M(s)$ in the mock catalog with our halo model calculation. The format is similar to Figure 5.3: solid curves in the bottom panels show the predicted marked statistic M when the central galaxy in a halo is treated differently from the others, and dashed curves show what happens if it is not. Both curves give the same prediction for the unweighted statistic ξ.

The top panel shows that the halo model calculation of $\xi(s)$ is in excellent agreement with the measurements on scales larger than a few Mpc, as it was for $\xi(r)$. However, it is not as accurate when the redshift separations are of order a few Mpc. Nevertheless, the model is able to reproduce the factor of ten difference between $\xi(r)$ and $\xi(s)$ on small scales. We will discuss the reason for the discrepancy on intermediate scales shortly.

Similarly, the bottom panel shows excellent agreement between measurements and model for the marked statistic $M(s)$ on large scales ($s > 8h^{-1}$Mpc), both when the central object is treated specially and when it is not. In addition, the model in which the central object is special is in better agreement with the measurements on small scales. (A χ^2 estimate of the goodness of fit of the two marked correlation models yields values which smaller by more than a factor of two when the central galaxy is treated specially compared to when it is not.)

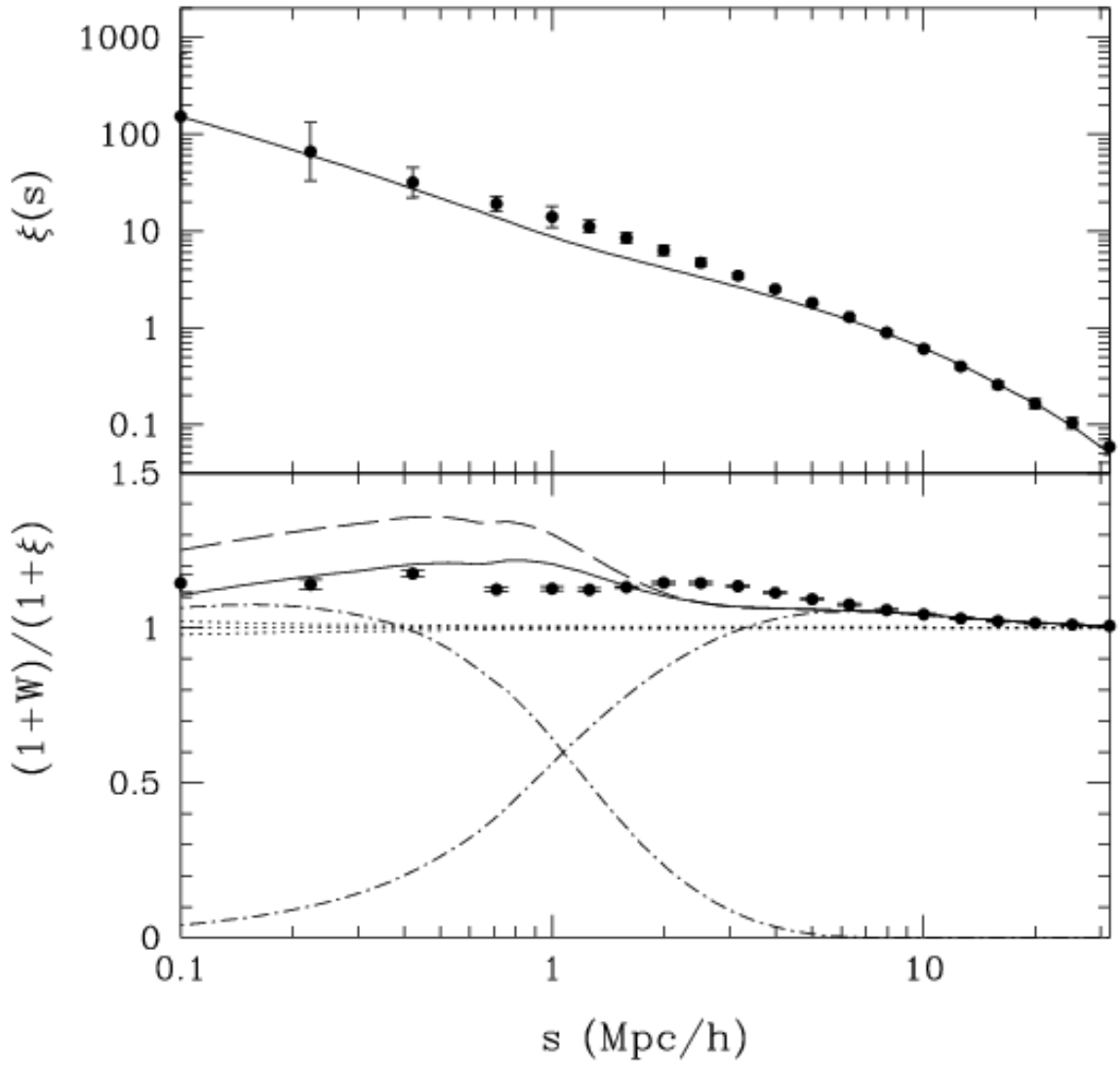

Figure 5.4: Luminosity-weighted redshift-space correlation functions measured in mock catalogs which resemble an SDSS volume limited sample with $M_r < -20.5$. Symbols show the measurements; smooth curves show the associated halo-model predictions when the luminosity of the central galaxy in a halo is assumed to be different from the others; dashed curves show the prediction when the central object is not special. Dotted curves show the mean and rms values of the statistic M, obtained by randomizing the marks and remeasuring M one hundred times. Dot-dashed curves show the one- and two-halo contributions to the statistic in our model when the central object in a halo is special.

On intermediate scales, however, there is substantial discrepancy discrepancy between the model and the mocks; the discrepancy is more pronounced for $M(s)$ than for $\xi(s)$.

To study the cause of this discrepancy, dot-dashed lines show the two contributions to the statistic, $\mathcal{W}_{1h}/(1+\xi)$ and $(1+\mathcal{W}_{2h})/(1+\xi)$, separately. This shows that it is on scales where both terms contribute that the model is inaccurate. There are two reasons why it is likely that this inaccuracy can be traced to our simple treatment of the two-halo term. The suppression of power due to virial motions means that we must model the two-halo term more accurately in redshift-space than in real-space. Our halo-model calculation incorrectly assumes that linear theory is a good approximation even on small scales (*e.g.*, Scoccimarro 2004 shows that this is a dangerous assumption even on scales of order 10 Mpc) and that volume exclusion effects (Mo & White 1996) are negligible (Sheth & Lemson 1999 discuss how one might incorporate such effects). Because our mocks make use of both the positions and velocities of the halos in the simulations, they incorporate both these effects. Thus, our simple halo-model likely underestimates $M(s)$ on intermediate scales, but overestimates it on smaller scales. Since this is in the sense of the discrepancy with the measurements in the mock catalogs, it is likely that this inaccuracy can be traced to our simple treatment of the two-halo term. We will have cause to return to this discrepancy in the next section, where we use our halo model calculation to interpret measurements of marked statistics in the SDSS dataset.

5.4 MEASUREMENTS IN THE SDSS

Figure 5.5 shows $\xi(r_p,\pi)$ and $W(r_p,\pi)$, the unweighted (solid) and weighted (dashed) correlation functions of pairs with separations r_p and π, perpendicular and parallel to the line of sight. The measurements were made in a volume limited catalog (59,293 galaxies with $M_r < -20.5$) extracted from the SDSS DR4 database (Adelman-McCarthy et al. 2006). Contours show the scales at which the correlation functions have values of 0.1, 0.2, 0.5, 1, 2, and 5, when averaged over bins of $2h^{-1}$Mpc in r_p and π. This format, due to Davis & Peebles (1983), allows one to isolate redshift space effects on the correlation functions, since these

act only in the π direction. The dotted quarter circles at separations of 5, 10 and $20h^{-1}$Mpc are drawn to guide the eye—they serve to highlight the fact that both $\xi(r_p, \pi)$ and $W(r_p, \pi)$ are very anisotropic. In contrast, the corresponding real-space quantities would be isotropic. The figure shows clearly that W has a slightly higher amplitude than ξ on the scales shown.

The quantity studied in the previous section, for which we have analytic (halo-model) estimates, can be derived from this plot as follows. Counting pairs in spherical shells of radius $s = \sqrt{r_p^2 + \pi^2}$ yields the redshift space correlation function $\xi(s)$. This measure of clustering is sensitive to the fact that the correlation function in redshift-space is anisotropic; in particular, it contains information about the typical motions of galaxies within halos (which are responsible for the elongation of the contours along the π direction at $r_p \leq 5h^{-1}$ Mpc), as well as the motions of the halos themselves (which are responsible for the squashing along the π direction at $r_p \geq 5h^{-1}$ Mpc). The result of counting pairs of constant r_p, whatever their value of π, yields the projected correlation function $w_p(r_p)$; since r_p is not affected by redshift space distortions, this quantity contains no information about galaxy or halo motions, so is more closely related to the real-space correlation function. Figures 5.6 and 5.7 compare both these quantities with the corresponding halo-model calculations.

Figure 5.6 shows $\xi(s)$ and $M(s)$ measured in two volume limited catalogs extracted from the SDSS database. One of these catalogs is the same as that which resulted in Figure 5.5, and the other is for a slightly fainter sample ($M_r < -19.5$, with 61,821 galaxies). Error bars are estimated by jack-knife re-sampling, as discussed previously. The solid curves show the redshift-space halo-model calculation in which central galaxies are special, and dashed curves show the expected signal if they are not. (Recall that both have the same $\xi(s)$.)

On large scales, both the solid and dashed curves provide an excellent description of the measurements on large scales. This agreement suggests that correlations with environment on scales larger than a few Mpc are *entirely a consequence of the correlation between halo abundances and environment*, just as they were in the mock catalogs. Since the model calculation incorporates the assumption that the halo mass function is top-heavy in dense regions, the agreement with the measured $M(s)$ provides strong evidence that this is indeed the case.

The discrepancy between the halo-model calculation and the measurements on interme-

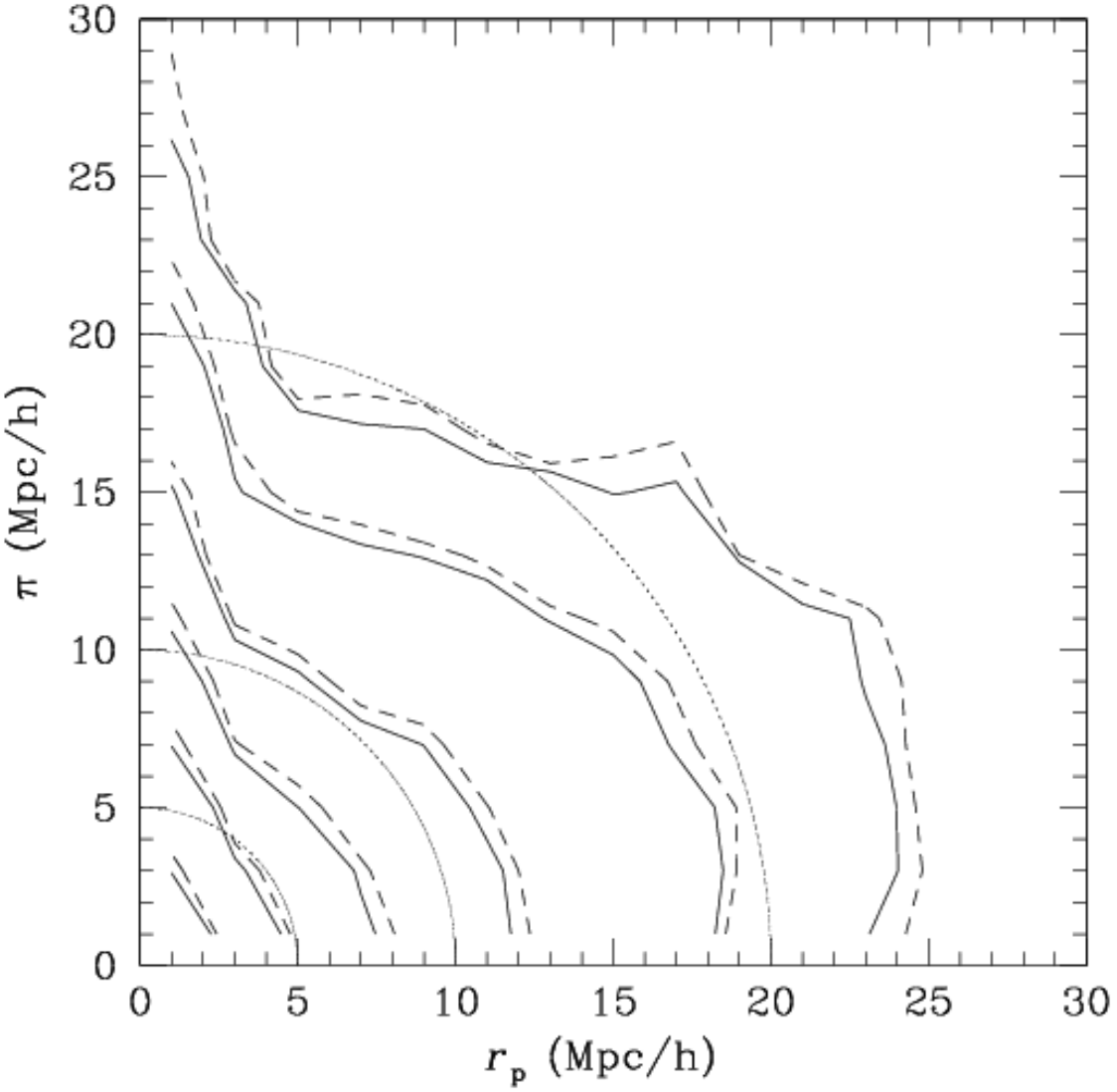

Figure 5.5: Unweighted (solid) and weighted (dashed) correlation functions measured in volume limited catalog with $M_r < -20.5$ in the SDSS. Dotted curves show that the measured correlation functions are significantly anisotropic.

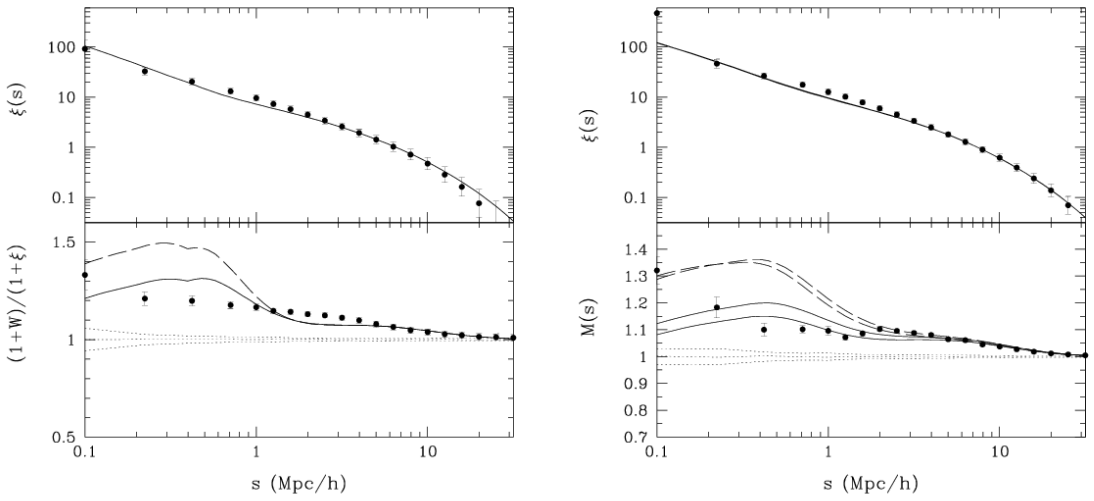

Figure 5.6: Redshift-space correlation functions measured in volume limited catalogs with $M_r < -19.5$ (left) and $M_r < -20.5$ (right) in the SDSS. Top panels show the unweighted correlation function $\xi(s)$, and bottom panels show the marked statistic $M(s)$. Smooth curves show the associated redshift-space halo-model predictions; solid curves are when the central galaxy in a halo is treated differently from the others, whereas this is not done for the dashed curves. Dotted curves show the mean and rms values of the statistic M, estimated by randomizing the marks and remeasuring M one hundred times. Two sets of curves are shown in the right hand panels; the top set of solid and dashed curves shows the halo model calculation in which the relation between the number of galaxies and halo mass is given by equation (5.1), and the bottom set follow from equation (5.3).

diate scales is similar to the discrepancy between the halo-model and the mock catalogs studied in the previous section. There we argued that this is almost certainly due to our simple treatment of the two-halo contribution to the statistic. Indeed, the marked statistics in the mock catalogs behave qualitatively like those in the SDSS data (compare Figures 5.4 and 5.6), suggesting that the discrepancy between the halo-model calculation and the measurements are due to this, rather than to any environmental effects operating on intermediate scales.

On small scales, the solid curves are in substantially better agreement with the data than are the dashed curves (χ^2 smaller by a factor of four in both plots). Evidently, central galaxies are indeed a special population in the data. This provides substantial support for the assumption commonly made in halo-model interpretations of the galaxy correlation function that the central galaxy in a halo is different from all the others.

However, even the solid curves are not in particularly good agreement with the measurements. Before attributing the discrepancy to environmental effects not included in the halo model description, we have explored the effect of modifying our parametrization of the relation between the number of galaxies and halo mass which we use (equation 5.1). Figure 5.6 shows that the parametrization in equation (5.3), with $\sigma = 0.5$ and $M_1/M_{\rm min} = 30$, provides equally good fits to $\xi(s)$, but a slightly better description of $M(s)$. In this parameterization of the scaling of $N_{\rm gal}$ with halo mass, the minimum halo mass required to host a galaxy is not a sharp step function.

Further evidence in support of the parametrization in which the minimum mass cutoff is not sharp, and in which the central galaxy is different from the others is shown in Figure 5.7. The top and bottom panels compare measurements of the projected correlation functions $w_p(r_p)$ and $M_p(r_p)$, where

$$\begin{aligned} w_p(r_p) &= \int {\rm d}y\, \xi(r_p, y) = 2\int_{r_p}^{\infty} {\rm d}r\, \frac{r\,\xi(r)}{\sqrt{r^2 - r_p^2}}, \quad \text{and} \\ M_p(r_p) &= \frac{1 + W_p(r_p)/r_p}{1 + w_p(r_p)/r_p}, \qquad \text{where} \\ W_p(r_p) &= 2\int_{r_p}^{\infty} {\rm d}r\, \frac{r\,W(r)}{\sqrt{r^2 - r_p^2}} \text{ and } r = \sqrt{r_p^2 + y^2}, \end{aligned} \tag{5.22}$$

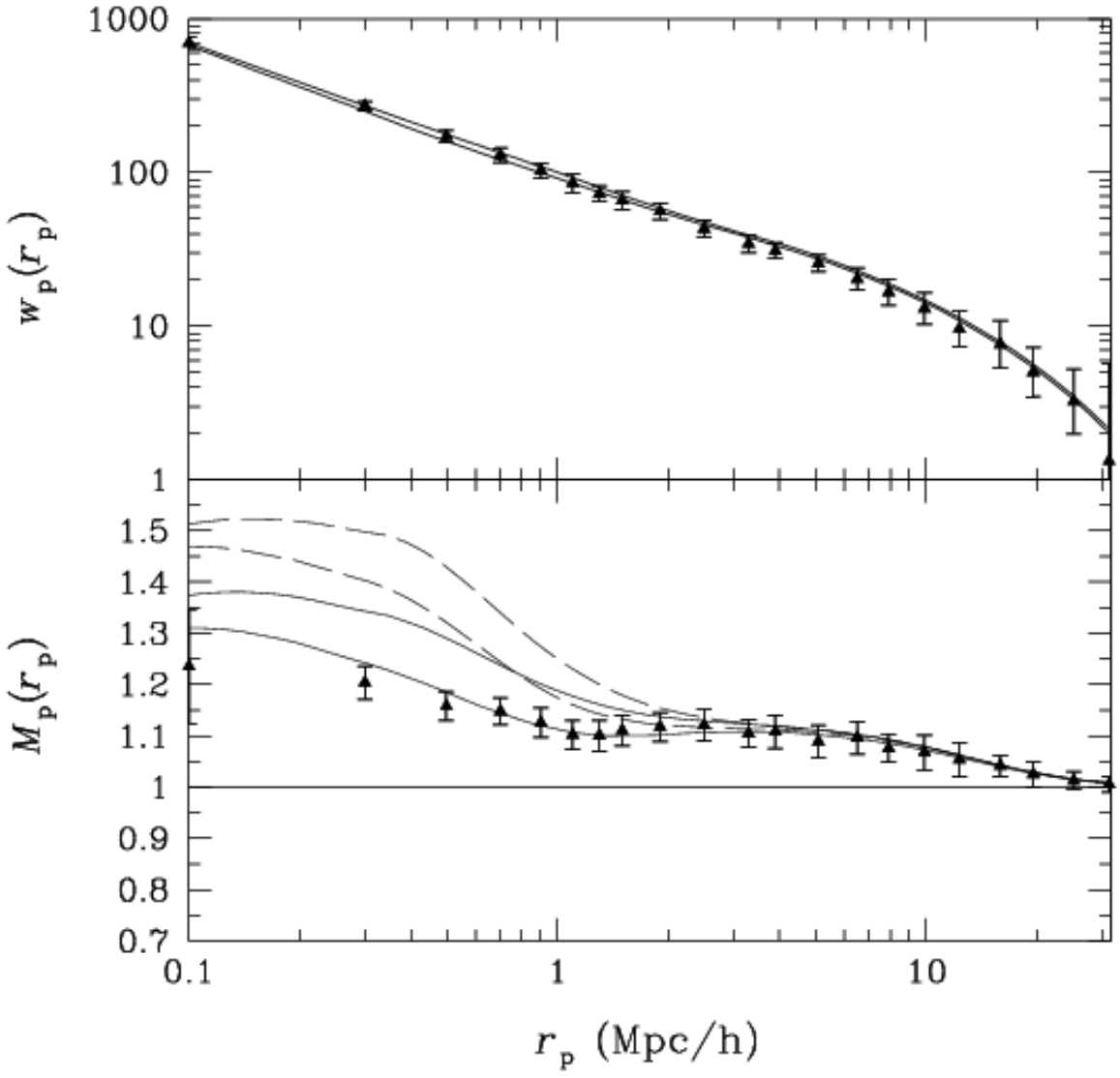

Figure 5.7: Projected correlation function measured in volume limited catalog with $M_r < -20.5$ in the SDSS. Top panel shows the unweighted projected correlation function $w_p(r_p)$, and bottom panels show the marked statistic $M_p(r_p)$. Smooth curves show the associated projected halo-model predictions; solid curves are when the central galaxy in a halo is treated differently than the others, whereas this is not done for the dashed curves. The upper set of dashed and solid curves show halo-model calculations which follow from equation (5.1); the lower set of curves assume equation (5.3).

with the associated halo-model calculations. (In the halo model, the real-space quantities $\xi(r)$ and $W(r)$ which appear in the expressions above, are related to $\xi(s)$ and $W(s)$ by setting setting $F_{\rm v} = 0$ and taking the limit $\sigma_{\rm vir} \to 0$ in u_1 and u_2. See Chapter 4for our particular definition of M_p.) Note that these projected quantities are free of redshift space distortions, making them somewhat easier to interpret.

As was the case for the redshift space measurements, both parameterizations of $N_{\rm gal}(M)$ provide good descriptions of the unweighted statistic $w_p(r_p)$, and in both cases, the weighted statistic is in better agreement when the central object is treated specially. However, the Figure shows clearly that when the central object is special, then equation (5.3) provides a substantially better description of M_p—the agreement with the measurements is excellent over all scales.

Since this paper was published, we have repeated the measurements above with the recent SDSS Data Release 5 volume-limited catalogs, which have 23% more galaxies in the case of the $M_r < -20.5$ sample. We also accounted for incompleteness due to uncertainties in the selection function and measured redshifts, and we applied evolution corrections as well as K-corrections to the magnitudes. The resulting $w_p(r_p)$ had a few points that were no more than 1% lower, and the luminosity-marked $M_p(r_p)$ was virtually the same as that of Figure 5.7. The only significant difference was that the $w_p(31.6\,{\rm Mpc}/h)$ point was much higher, almost right on top of the theory curve.

5.5 DISCUSSION

We showed how to generate a mock galaxy catalog which has the same luminosity function (Figure 5.1) and luminosity dependent two-point correlation function as the SDSS data. We used the mock catalog to calculate the luminosity-weighted correlation function in a model where all environmental effects are a consequence of the correlation between halo mass and environment (Figures 5.3 and 5.4 show results in real and redshift space). We then showed how to describe this luminosity-weighted correlation function in the language of the redshift-space halo model (equation 5.15). The analysis showed that estimates of the

luminosity dependence of clustering constrain how the mass-to-light ratio of halos depends on halo mass (equation 5.21 and Figure 5.2). The central galaxy in a halo is predicted to be substantially brighter than the other objects in the halo, and although the luminosity of the central object increases rapidly with halo mass, the mean luminosity of the other objects in the halo is approximately independent of the mass of the host halo.

Our analysis also showed that measurements of clustering as a function of luminosity completely determine the simplest halo model description of marked statistics. In addition, measurements of the marked correlation function allow one to discriminate between models which treat the central object in a halo as special, from those which do not (Figures 5.3 and 5.4). Also, in hierarchical galaxy formation models, the marked correlation function is expected to show a signal on large scales if the average mark of the galaxies in a halo correlates with halo mass. The signal arises because massive halos populate the densest regions; it is present even if there are no physical effects which operate to correlate the marks over large scales.

We compared this halo model of marked statistics with measurements in the SDSS (Figures 5.6 and 5.7). The agreement between the model and the measurements on scales smaller than a few Mpc provides strong evidence that central galaxies in halos are a special population—in general, the central galaxy in a halo is substantially brighter than the others. (Berlind et al. 2005 come to qualitatively similar conclusions, but from a very different approach.) Substantially better agreement is found for a model in which the minimum halo mass required to host a luminous central galaxy does not change abruptly with luminosity. This is in qualitative agreement with some semi-analytic galaxy formation models, which generally predict some scatter in central luminosity at fixed halo mass (e.g. Sheth & Diaferio 2001; Zheng et al. 2005).

The agreement between the halo model calculation and the data on scales larger than a few Mpc indicates that the standard assumption in galaxy formation models, that halo mass is the primary driver of correlations between galaxy luminosity and environment, is accurate. In particular, these measurements are consistent with a model in which the halo mass function in large dense regions is top-heavy, and, on these large scales, there are no additional physical or statistical effects which affect the luminosities of galaxies. In this

respect, our conclusions are similar to those of Mo et al. (2004), Kauffmann et al. (2004), Blanton et al. (2005), Weinmann et al. (2006a), and Abbas & Sheth (2006), although our methods are very different.

We note in passing that there is a weak statistical effect for which the halo-model above does not account: at fixed mass, halos in dense regions form earlier (Sheth & Tormen 2004). Gao, Springel & White (2005) show that this effect is more pronounced for low mass halos (related marked correlation function analyses by Harker et al. 2006 and Wechsler et al. 2005 come to similar conclusions). The agreement between our halo-model calculation and the measurements in the SDSS suggests that this correlation between halo formation and environment is not important for the relatively bright galaxy population we have studied here. This is presumably because these SDSS galaxies populate more massive halos. Comparisons with larger upcoming SDSS datasets, with a fainter luminosity threshold (such as $M_r < -18$), may bear out the correlation between low-mass halo formation and environment.

As a final indication of the information contained in measurements of marked statistics, Figure 5.8 compares $M(s)$ when the $u-$, $g-$ and $r-$band luminosities are used as the mark. The underlying population is the same as that for Figures 5.5–5.7: the sample is volume limited to $M_r < -20.5$. Thus, $\xi(s)$ is fixed, and only $W(s)$ changes with wave-band. Notice that there is a clear trend with wavelength: $M(s)$ is slightly smaller than unity on all but the very smallest scales when the mark is $u-$band luminosity, it is greater than unity when $g-$band luminosity is the mark, especially on small scales, and this scale dependence is even stronger when L_r is the mark. This trend with wavelength is qualitatively consistent with the predictions of semi-analytic galaxy formation models (Sheth, Connolly & Skibba 2005) and indicates that the mean u-band luminosity of the galaxies in a halo depends less strongly on halo mass than does the mean $r-$band luminosity in a halo (Sheth 2005). In the models, the $u-$band luminosity is an indicator of star formation, although Fig. 5 in Sheth (2005) shows that this is only true at small halo masses. Therefore, it will be interesting to compare our u-band measurement with future measurements which use estimators of the star-formation rate as the mark.

Figure 5.9 shows the result of weighting these same galaxies by their colors. The top panel shows results where the weight is the difference in the absolute magnitudes, $M_u - M_r$ and

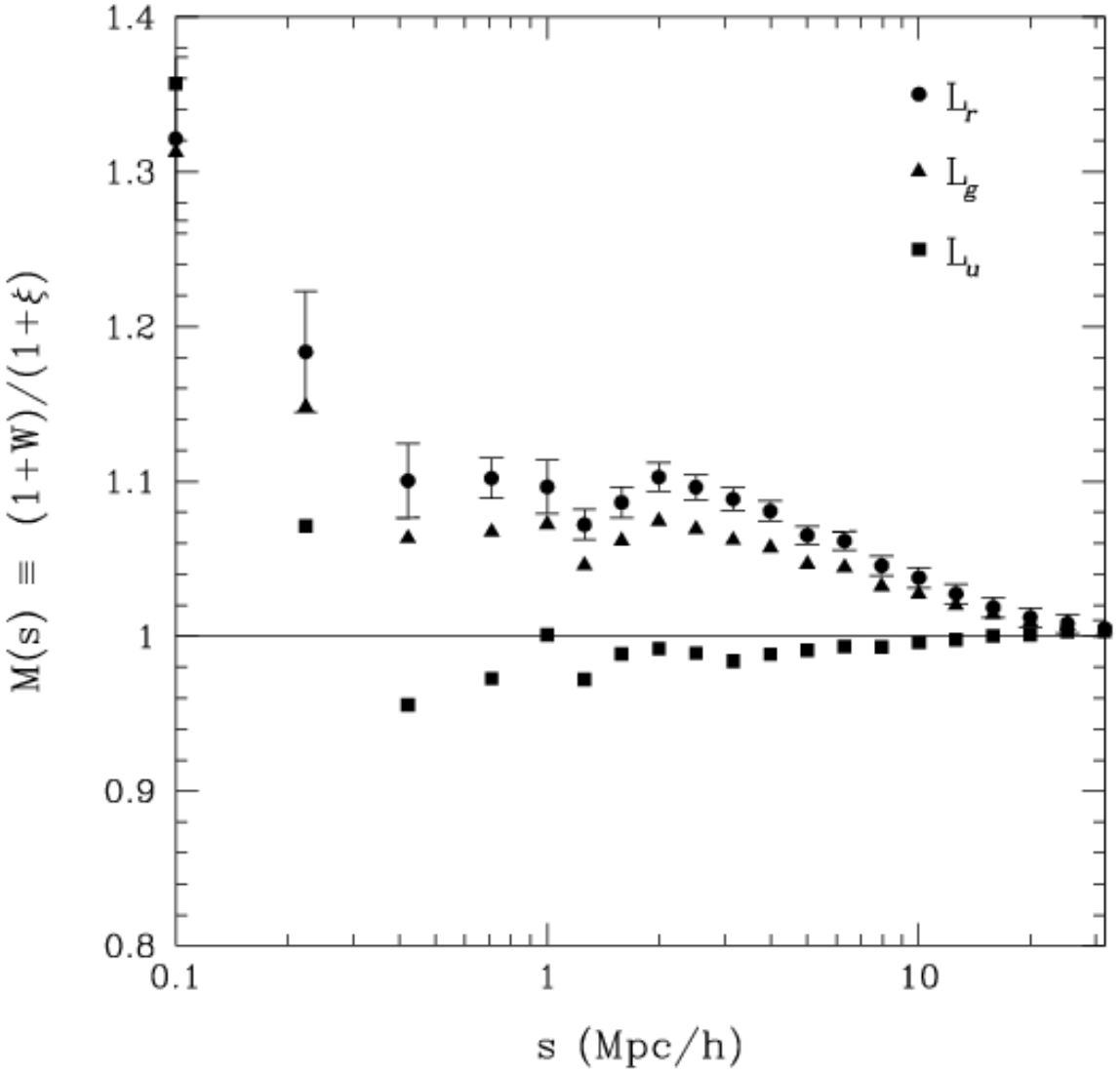

Figure 5.8: Redshift-space luminosity-weighted correlation functions measured in volume limited catalogs with $M_r < -20.5$ in the SDSS. Circles, triangles and squares show $M(s)$ when the weight is $r-$, $g-$ and $u-$band luminosity respectively. For clarity, jack-knife error bars are only shown for the $r-$band measurement, since the uncertainties are similar in the other bands.

$M_g - M_r$, whereas the weights in the bottom bottom panel were the ratios of the luminosities in two bands. Comparison of the two panels shows the effect on $M(s)$ of rescaling the weights while preserving their rank-ordering—while there are quantitative differences, the results in both panels are qualitatively similar. The $M(s)$ measurements shown in the bottom panel are more widely separated because the luminosity ratio involves $10^{\rm color}$, which has the effect of weighting the redder galaxies more heavily. Figure 5.10 compares the distributions of the $M_u - M_r$ and $M_g - M_r$ color marks and $10^{\rm color}$ marks, and the distributions of the luminosity ratios are indeed significantly different. The L_r/L_u distribution is much more spread out than that of the L_r/L_g distribution, contributing to the stronger marked correlation function. The L_r/L_u mark distribution is also offset by 7 galaxies with extremely large values of the mark, due to abnormally faint u-band magnitudes, and these would also affect the marked correlations.

This analysis of color-marked correlation functions indicates clearly that close pairs of galaxies tend to be redder than average. Sheth, Connolly & Skibba (2005) show that this is also the case in semi-analytic galaxy formation models.

The measurements shown in Figures 5.8 and 5.9 are consistent with models in which galaxies in clusters are more massive and have smaller star formation rates than galaxies in the field (to the extent that the $u-$band luminosity is an indicator of star-formation). In effect, these figures demonstrate the environmental dependence of galaxy luminosities and colors, without having to divide the galaxy sample up into discrete bins of 'field', 'group', and 'cluster'. Thus, marked statistics allow one to study correlations with environment over a continuous range in density, rather than in somewhat arbitrary discrete bins in environment. In this respect, our use of marked statistics to quantify and interpret environmental trends is very different from recent approaches which address the same problem. Since marked statistics are simple to measure and interpret, we hope that they will become standard tools for quantifying the correlation between the properties of galaxies and their environments.

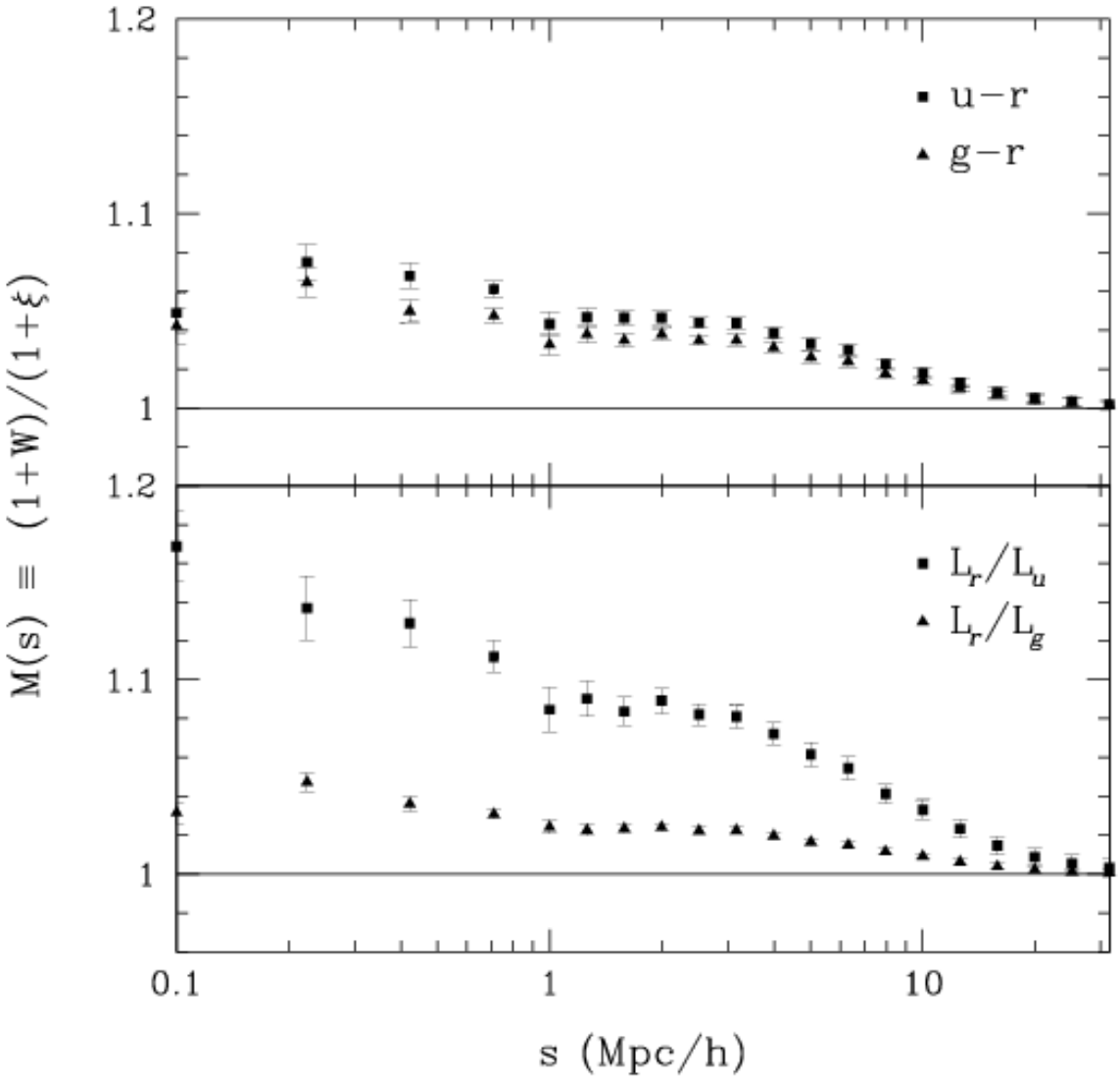

Figure 5.9: Redshift-space color-weighted correlation functions measured in volume limited catalogs with $M_r < -20.5$ in the SDSS. The top panel shows results when the color weight is the difference in absolute magnitudes. In the bottom panel, galaxies were weighted by the ratio of the luminosities in the two bands. Both panels show that close pairs of galaxies tend to have redder colors.

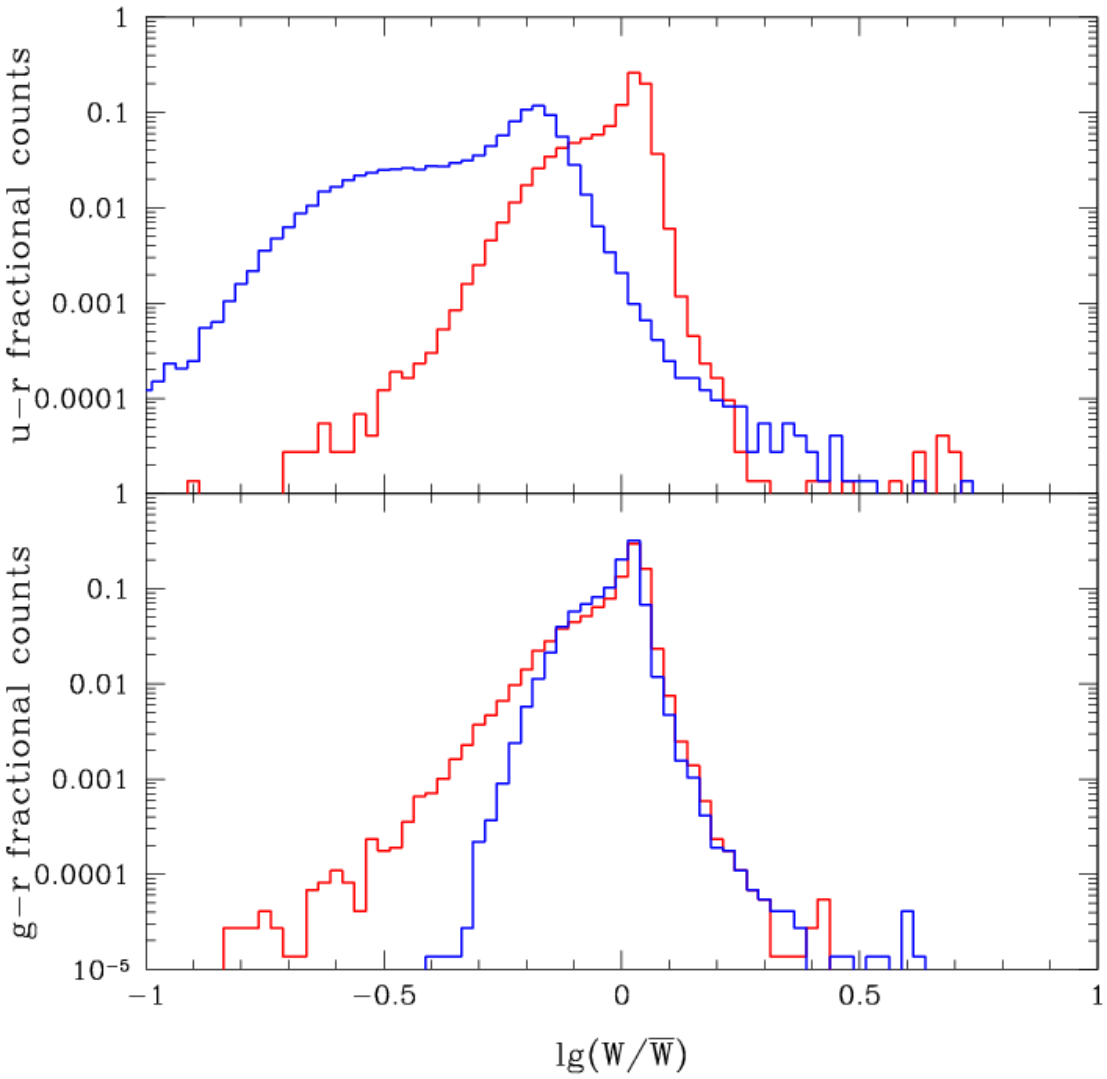

Figure 5.10: Comparison of distributions of color marks and 10^{color}, relative to mean mark. Top panel: $u - r$ (red histogram) and L_r/L_u (blue); bottom panel: $g - r$ (red) and L_r/L_g (blue). The L_r/L_u distribution is slightly offset because of a small number of galaxies with extremely faint u-band magnitudes.

6.0 LUMINOSITY- AND STELLAR MASS-MARKED CORRELATION FUNCTIONS OF SDSS AND MILLENNIUM SIMULATION GALAXIES

6.1 INTRODUCTION

Hierarchical models of structure formation have shown that there is a correlation between the formation and abundances of halos and large-scale structure (Mo & White 1996; Sheth & Tormen 2002). Dark matter halo properties, especially their masses, are correlated with their environments: less massive halos tend to form in underdense regions and more massive halos tend to form in overdense regions.

Observational studies of galaxy surveys have shown that galaxy properties are similarly correlated with their environments: luminous galaxies tend to be more clustered than faint ones, and massive, red, passively star-forming, elliptical galaxies tend to be more clustered than their less massive, blue, actively star-forming, spiral counterparts (*e.g.*, Zehavi et al. 2005, Berlind et al. 2005, Blanton et al. 2005, Bernardi et al. 2005, Best et al. 2005, Kuehn & Ryden 2005, Li et al. 2006). Based on such evidence, galaxy formation models assume that the properties of a galaxy are determined entirely by the mass and formation history of the halo within which it formed. This would entail that the correlation between halo properties and environment induces correlations between galaxy properties and environment. We have found that this statistical correlation accounts for most of the observed trends between luminosity and environment (Skibba et al. 2006), and the main goal of the present work is to determine to what extent this is also the case with galaxies in the Millennium Simulation.

We are doing so by using a novel and versatile set of tools known as marked statistics (Stoyan & Stoyan 1994, Beisbart & Kerscher 2000), which have been shown to provide sensitive probes of environmental effects (Sheth & Tormen 2004; Sheth, Connolly & Skibba

2005). Rather than measuring clustering by treating galaxies as points without attributes and making semi-arbitrary cuts in the data, marked clustering statistics allow us to describe the clustering of those continuous galaxy marks themselves, thus allowing us to probe the environmental dependence of those galaxy properties.

The halo model (see Cooray & Sheth 2002 for a review) is the framework currently used to interpret measurements of galaxy clustering. Sheth (2005) develops the formalism for including marked correlations in the halo model of clustering, and we extend this to describe measurements made in redshift space in Chapter 4. The halo model provides an analytical description of marked clustering statistics when correlations with environment arise entirely because of the statistical correlation between halo mass and environment. This paper incorporates the halo model description of luminosity-marked statistics of Skibba et al. (2006) and applies them to Millennium Simulation galaxies. We also develop a description of stellar mass-marked statistics. We compare our halo model predictions of luminosity- and stellar mass-marked correlations to our measurements in the Millennium Simulation, and we discuss how these results contrast with similar results from the SDSS.

Because galaxy properties are assumed to be determined by the mass and formation histories of the host halos, galaxy formation modelers have taken the approach of carefully simulating the distribution of dark matter, and then semi-analytically simulating the formation and evolution of galaxies within those dark matter objects (Kauffmann et al. 1999). The Millennium Run is the largest calculation of the evolution of dark matter structure with the concordance ΛCDM cosmology (Springel et al. 2005). Croton et al. (2005) have used semi-analytic models to simulate the growth of galaxies and their central black holes on the output of the Millennium simulation. They applied models of gas cooling, star formation, supernova feedback, plus a new model of the 'radio mode' of AGN feedback, which results in suppressing cooling flows and hence, star formation.

We have used samples of galaxies that resulted from this simulation to investigate their clustering and the clustering of their luminosities and stellar masses. This kind of marked statistics analysis yields information directly about the environmental dependence of these galaxy properties and provides a strong test of some of the assumptions included in the semi-analytic models. This interaction between theories, simulations, and observations is very

useful in that it allows us to inform and refine the halo model of marked galaxy clustering and the semi-analytic models of galaxy formation.

This paper is organized as follows. The next section provides the halo model description of luminosity-marked correlations. In Section 6.3, we show and discuss our clustering measurements of Millennium Run simulation galaxies in samples with three different luminosity thresholds. We compare these results to the halo model predictions and to our measurements with SDSS data, and we discuss the disagreement between them. We do a similar analysis for stellar mass-marked correlation functions in Section 6.4, and we also compare our halo model metallicity-marked correlation functions to our SDSS measurements. We compare measurements of star formation rate- and specific SFR-marked correlation functions of the SDSS and Millennium simulation in Section 6.5. In the last section we discuss our results and their implications.

6.2 LUMINOSITY-MARKED CORRELATION FUNCTIONS IN THE HALO MODEL

6.2.1 Measuring Marked Correlation Functions

In what follows, a mark is a weight or attribute associated with each point in a point process. A point process is a galaxy catalog, and a mark can be any observable property associated with a galaxy, such as luminosity, color, velocity dispersion, stellar mass, star formation rate, etc. Marked statistics measure the clustering of marks. Since the positions at which the marks are measured may themselves be clustered, marked statistics are defined in a way which accounts for this.

The marked correlation function we measure is

$$M(r) \equiv \frac{1+W(r)}{1+\xi(r)}, \tag{6.1}$$

where $\xi(r)$ is the two-point correlation function and $W(r)$ is the same sum over galaxy pairs separated by r, but now each member of the pair is weighted by the ratio of its

mark to the mean mark of all the galaxies in the catalog, as explained by Sheth, Connolly & Skibba (2005). In effect, the denominator divides-out the contribution to the weighted correlation function which comes from the spatial contribution of the points, leaving only the contribution from the fluctuations of the marks.

A crude estimator of the unweighted and weighted correlation functions are DD/RR and WW/RR, where the RR counts are the sums of pairs in a random catalog of the same geometry. In principle, one could then use WW/DD as an estimator for the marked correlation function in (1), obviating the need for a random catalog. Because edge effects appear in both the numerator and denominator, this marked correlation ratio is less sensitive to edge effects. However, in practice, we use the estimator of Landy & Szalay (1993) for both $\xi(r)$ and $W(r)$, which minimizes the variance on large scales. Moreover, the marked correlation function is more useful and easier to interpret when examined alongside the unweighted correlation function, as a bias or other problem with the latter will significantly affect the former.

6.2.2 Halo Model Description of the Marked Correlation Function

This section shows how to describe the luminosity-marked correlation function in the language of the halo model.

In the halo model, all mass is bound up in dark matter halos which have a range of masses. Hence, the density of galaxies is

$$\bar{n}_{\rm gal} = \int_{m_{\rm min}} dm \, \frac{dn(m)}{dm} \, \langle N_{\rm gal}|m\rangle, \tag{6.2}$$

where $dn(m)/dm$ denotes the number density of halos of mass m, and $\langle N_{\rm gal}|m\rangle$ characterizes the distribution of galaxies within halos of mass m.

Zehavi et al. (2005) have measured the luminosity dependence of clustering in the SDSS. They interpret their measurements using the language of the halo model (see Cooray & Sheth 2002 for a review). In the ΛCDM model, only sufficiently massive halos ($M_{\rm halo} > 10^{11} M_{\odot}$) host galaxies. Each sufficiently massive halo hosts a galaxy at its center, and may host 'satellite' galaxies in 'subhalos'. We model the distribution of galaxies occupying halos by

accounting for the fact that central and satellite galaxies are distinctly different populations:

$$\langle N_{\rm gal}|M\rangle \;=\; \langle N_{\rm cen}|M\rangle \;+\; \langle N_{\rm sat}|M\rangle \;\approx\; 1 \;+\; \left(\frac{M}{\mu\, M_{\rm min}(L_{\rm min})}\right)^{\alpha(L_{\rm min})} \tag{6.3}$$

where the number of satellites follows a Poisson distribution with a mean value which increases with halo mass (following Kravtsov et al. 2004). The minimum mass of halos that host the population of galaxies in a volume-limited catalog is $M_{\rm min}$; the critical mass above which halos typically host one or more satellite galaxies is $M_1 \equiv \mu M_{\rm min}$; and the power-law slope α characterizes the mass dependence of the efficiency of galaxy formation. In describing the clustering of Millennium simulation galaxies, we have found that for distinct low halo masses, the satellite distribution $\langle N_{\rm sat}|M\rangle$ rolls off faster than a power-law, and we have accounted for this effect with an exponential $\exp(-\beta/(M/M_{\rm min}))$ (Tinker et al. 2005, Conroy et al. 2005).

Not surprisingly, the best-fit parameters of the halo occupation distribution are highly dependent on galaxy properties such as luminosity and color (Zehavi et al. 2005, Collister & Lahav 2005, Yang et al. 2005b). As a result, when studying a galaxy catalog with data cuts, and especially when studying a particular galaxy mark, it is crucial to carefully model the dependence of the halo occupation distribution on that galaxy property in an observationally-motivated way.

To describe the clustering of galaxies, the halo model requires additional components: the density profile and bias factor, which are determined from dark matter simulations, and the linear power spectrum. We use the halo mass function and bias factor of Sheth & Tormen (2002). We use the linear power spectrum of Efstathiou, Bond & White (1992); we have found that no nonlinear power spectrum was required to adequately model the transition from small to large scales. We assume the halos have an NFW profile (Navarro, Frenk, White 1997), although the profile of Moore et al. (1999) yields virtually the same correlation functions for the scales we can reliably probe. At smaller scales, r <100 kpc/h, the shape of the density profile matters. The density profile is expected to depend on galaxy properties, especially color, star formation rate, and morphology, and one must account for this when studying such marks at small scales.

The two-point correlation function is the Fourier transform of the power spectrum:

$$\xi(r) = \int \frac{dk}{k} \frac{k^3 P(k)}{2\pi^2} \frac{\sin kr}{kr}. \tag{6.4}$$

In the halo model, $P(k)$ is written as the sum of two terms: one that arises from galaxies within the same halo and dominates on small scales (the 1-halo term), and the other from galaxies in different halos which dominates on larger scales (the 2-halo term). That is,

$$P(k) = P_{1h}(k) + P_{2h}(k), \tag{6.5}$$

where,

$$\begin{aligned} P_{1h}(k) &= \int_{M_{\rm min}} dM \, \frac{dn(M)}{dM} \left[\frac{2 \, \langle N_{\rm sat}|M\rangle \, u_{\rm gal}(k|M)}{\bar{n}^2_{\rm gal}} \right. \\ &\quad \left. + \frac{\langle N_{\rm sat}(N_{\rm sat}-1)|M\rangle \, u_{\rm gal}(k|M)^2}{\bar{n}^2_{\rm gal}} \right], \qquad (6.6) \\ P_{2h}(k) &= \left[\int_{M_{\rm min}} dM \, \frac{dn(M)}{dM} \right. \qquad (6.7) \\ &\quad \left. \times \frac{1 + \langle N_{\rm sat}|M\rangle \, u_{\rm gal}(k|M)}{\bar{n}_{\rm gal}} \, b(M) \right]^2 P_{\rm lin}(k), \end{aligned}$$

where $u_{\rm gal}(k|M)$ is the Fourier transform of the galaxy density profile, which we assume to be an NFW profile, so u has the form given by Scoccimarro et al. (2001). The second moment of the halo occupation distribution,
$\langle N_{\rm sat}(N_{\rm sat}-1)|M\rangle = \langle N_{\rm sat}|M\rangle^2$, because we assume the satellites to have a Poisson distribution. The two parts of the 1-halo term in equation (6.6) can be thought of as the 'center-satellite term' and the 'satellite-satellite term', which are the two possibilities for pairs of galaxies within the same host halo.

6.2.3 Luminosity Mark

To describe the effect of weighting each galaxy, let $W(k)$ denote the Fourier transform of the weighted correlation function. Following Sheth, Abbas & Skibba (2004) and Sheth (2005), we write this as the sum of two terms:

$$W(k) \;=\; W_{1h}(k) \;+\; W_{2h}(k). \tag{6.8}$$

As described in Sheth (2005), the simplest way to model marked correlations is to weight all galaxies the same way, so that, given a halo of a particular mass and a mass-dependent mark, each galaxy is weighted by the average weight of all the galaxies within the halo. However, since central galaxies and satellite galaxies have distinctly different properties, it is more realistic to weight central galaxies by their mass-dependent mark and weight satellite galaxies by their average mass-dependent mark. When luminosity is the mark, the center-satellite model of the 1-halo and 2-halo terms is the following:

$$\begin{aligned} W_{1h}(k) \;&=\; \int_{M_{\rm min}} dM\, \frac{dn(M)}{dM} \\ &\quad\times \left[\frac{2\, L_{\rm cen}(M)\, \langle L_{\rm sat}|M, L_{\rm min}\rangle\, \langle N_{\rm sat}|M\rangle\, u_{\rm gal}(k|M)}{\bar{n}_{\rm gal}^2\, \bar{L}^2} \right. \\ &\quad\left. + \frac{\langle N_{\rm sat}|M\rangle^2\, \langle L_{\rm sat}|M, L_{\rm min}\rangle^2\, u_{\rm gal}^2(k|M)}{\bar{n}_{\rm gal}^2\, \bar{L}^2} \right], \end{aligned} \tag{6.9}$$

$$\begin{aligned} W_{2h}(k) \;&=\; \left[\int_{M_{\rm min}} dM\, \frac{dn(M)}{dM} \right. \\ &\quad\left. \frac{L_{\rm cen}(M) \;+\; \langle N_{\rm sat}|M\rangle\, \langle L_{\rm sat}|M, L_{\rm min}\rangle\, u_{\rm gal}(k|M)}{\bar{n}_{\rm gal}\, \bar{L}}\, b(M) \right]^2 P_{\rm lin}(k), \end{aligned} \tag{6.10}$$

where

$$\bar{L} = \int_{M_{\rm min}} dM\, \frac{dn(M)}{dM}\, \frac{L_{\rm cen}(M) \;+\; \langle N_{\rm sat}|M\rangle\, \langle L_{\rm sat}|M, L_{\rm min}\rangle}{\bar{n}_{\rm gal}}\,. \tag{6.11}$$

Here $\bar{L}$ is the average luminosity, $L_{\rm cen}(M)$ is the luminosity of the galaxy at the center of an M-halo, and $\langle L_{\rm sat}|M, L_{\rm min}\rangle$ is the average luminosity of satellite galaxies more luminous than $L_{\rm min}$ in M-halos.

We determine the relationship between halo mass and central galaxy luminosity as follows. The number density of halos as a function of mass can be easily obtained from the

halo mass function. To obtain the *cumulative* number density of only the halos of central galaxies (*i.e.*, excluding subhalos hosting satellite galaxies), we integrate the mass function with the observed luminosity-dependent halo occupation distribution (Zehavi et al. 2005). This yields the cumulative number density of galaxies in halos with mass greater than $M_{\rm min}$. For each mass, we match these number densities with the cumulative number densities of galaxies of luminosity greater than $L_{\rm min}$, obtained from the SDSS luminosity function (Blanton et al. 2003a). This constitutes a monotonic relationship between central luminosity and halo mass, which has the shape of a double power-law for each of the five SDSS bands. This procedure is similar to, but different than, that of Vale & Ostriker (2004) and Shankar et al. (2006).

Our calculation of the average satellite luminosity, $\langle L_{\rm sat}|M, L_{\rm min}\rangle$, and a comparison of the mass-dependence of it and of $L_{\rm cen}(M)$ are shown in Section 3 of Skibba et al. (2006). The central and satellite luminosity marks, as well as the stellar mass and metallicity marks in Section 4, are coupled to each other in the sense that at fixed halo mass luminous (massive) central galaxies have fairly luminous (massive) satellites on average. This naturally produces the 'galactic conformity' phenomenon observed by Weinmann et al. (2006a).

Now that we have the mass-dependent marks, we can calculate the expected luminosity-marked correlation function for different luminosity thresholds by Fourier transforming the unweighted and weighted power spectra:

$$M(r) \; = \; \frac{1 + W_{1h}(r) + W_{2h}(r)}{1 + \xi_{1h}(r) + \xi_{2h}(r)}.$$

6.3 MEASUREMENTS IN THE MILLENNIUM RUN SIMULATION

In order to compare our measurements to analyses we have done with volume-limited SDSS catalogs, we used cut the full Millennium catalog of Croton et al. (2005) into subcatalogs with luminosity thresholds of $M_r < -19.5$, $M_r < -20.5$, and $M_r < -21.5$. These catalogs contained 1,599,488, 519,892, and 63,556 simulated galaxies, respectively, which were large enough to yield good statistical measurements but small enough to allow for measurements

that are not prohibitively computationally intensive. All of our halo model calculations assumed $\Omega_m = 0.25$ and $\Omega_\Lambda = 0.75$ for consistency. We assumed $\sigma_8 = 0.9$, and all quantities are in terms of $h \equiv H_0/100$ km s^{-1} Mpc$^{-1} = 1$. We also used the CMBFAST linear power spectrum, $P_{\rm lin}(k)$.

In Section 3.1, we discuss issues involving the number densities of our catalogs, the minimum halo mass of the measurements, and the halo occupation distribution of our calculations. In Section 3.2, we compare the distributions of the luminosity marks of central and satellite galaxies, and explain how we rescale the luminosities of the Millennium simulation catalogs in order to do a more direct comparison to the halo model. We show our measurements and compare them to our halo model predictions in Section 3.3.

6.3.1 Halo Occupation Distribution

Our previous halo model predictions of the redshift-space $\xi(s)$ and projected $w_p(r_p)$ correlation functions for various luminosity thresholds were consistent with our measurements with SDSS catalogs (Skibba et al. 2006). However, we found a systematic discrepancy when comparing to our correlation function measurements of our Millennium simulation catalogs. The measured correlation functions were systematically low, by an apparently *constant* value, for $100\,{\rm kpc}/h \leq r \leq 30\,{\rm Mpc}/h$. This discrepancy was larger for our more luminous catalogs.

In general, theoretical and observational two-point correlations increase when the luminosity threshold is increased. In the halo model, this occurs because increasing $L_{\rm min}$ means increasing $M_{\rm min}(L_{\rm min})$, which results in decreasing the galaxy number density, $\bar{n}_{\rm gal}$ in (2), more than the un-normalized correlation functions themselves are decreased. In the measurements, it occurs simply because brighter galaxies tend to be more strongly clustered relative to the number density of galaxy catalogs.

Clustering measurements depend on number density and on the range of luminosities or halo masses considered, and comparisons with analytic calculations must account for these. Our halo model calculations assumed number densities based on halo abundances and the SDSS luminosity function, and we assumed halo occupation distributions based on SDSS observations. It is therefore critical to compare the number densities of our Millennium

catalogs with their SDSS counterparts. As expected, the number densities of the Millennium catalogs were *significantly* larger, especially for the more luminous catalogs. (Our SDSS number densities are similar to those in Table 2 of Zehavi et al. (2005), even though our catalogs came from a more recent data release.) To account for the discrepant number densities, we rank-ordered the Millennium galaxies by their luminosities, and then excluded the required number of faintest galaxies so that each catalog had the same number density as its SDSS counterpart. For our most luminous sample, $M_r < -21.5$, this resulted in reducing the number of galaxies by more than a third and shifted the effective luminosity threshold brighter by about 0.2 magnitudes; the correction had a smaller effect for fainter thresholds. We repeated our measurements, and this correction resulted in ameliorating some, but not all, of the discrepancy with the $\xi(r)$ predicted by the halo model.

Secondly, all of our catalogs are defined by luminosity thresholds, and our $M_{\rm min}(L_{\rm min})$ relation significantly depends on the luminosity function. Although we assumed an SDSS luminosity function in all of our other measurements, to determine the minimum mass from the luminosity threshold we used a 2dF-like luminosity function, with a brighter 'knee' and steeper faint-end slope. This resulted in an $M_{\rm min}$ that was lower by a factor of approximately 1.3, 1.5 and 2 in our $M_r < -19.5$, $M_r < -20.5$ and $M_r < -21.5$ catalogs, respectively. This ameliorated much of the remaining discrepancy with the measurements.

The effects of these two corrections can be seen in Figure 6.1, which compares the halo occupation distributions of satellite galaxies for $M_r < -19.5$.

The cyan points are the original mean satellite distribution and the blue points are the distribution subsequent to the number density correction. The green line shows the SDSS satellite distribution of Zehavi et al. (2005) for $M_r < -19.5$, and the red curve is the satellite distribution we used in our halo model calculations, with parameters $\mu \approx 28$, $\alpha \approx 1.1$, and $\beta \approx 1$. For our minimum halo mass, the measured halo occupation distribution demands a lower value of $\mu = M_1/M_{\rm min}$ (and of the slope α); however, a lower values of μ means more satellites per halo, on average, resulting in an unweighted correlation function that is not consistent with the measured $\xi(r)$ (upper panels of Figures 6.5 and 6.6). Our adopted HOD is therefore a compromise between the measured $\xi(r)$ and the measured $\langle N_{\rm sat}|M\rangle$.

It is important to stress that, even after correcting for the number densities and for

the LF-dependence of $M_{\rm min}(L_{\rm min})$, the minimum halo masses of the Millennium simulation are still slightly too low. For $M_r < -19.5$, the minimum halo mass should be $M_{\rm min} = 3.4 - 6.4 \times 10^{11} M_\odot$, due to uncertainties in the luminosity function. As can be seen in Figure 6.1, there are a considerable number of halos less massive than this, and some of them even host multiple subhalos. The corresponding minimum mass of the simulated $M_r < -19.5$ catalog is $M_{\rm min} \approx 2.1 \times 10^{11} M_\odot$, while accounting for the scatter in the mass-luminosity relation. This remaining discrepancy is slightly larger, and even more important, at brighter luminosity thresholds. This is crucial because the mass threshold directly affects the amplitude of the unweighted correlation function and the shape of marked correlation functions.

Now that the halo model fairly successfully describes the unweighted two-point correlation functions of the Millennium Run simulation, we can now examine the marked correlation functions.

6.3.2 Distributions of Central and Satellite Galaxy Luminosity Marks

For the luminosity-marked correlation functions of our Millennium Simulation catalogs, we used the r-band luminosities because the luminosity mark and luminosity-dependent HOD of our halo model calculations are based on r-band SDSS observations. However, the Millennium simulation data outputted by Croton et al. (2005) included absolute magnitudes that were *converted* to SDSS *ugriz* bands. Although such conversions are inherently imperfect, more important is the fact that 2dF luminosity functions were assumed (Norberg et al. 2002). The luminosity function naturally has a strong impact on luminosity-marked correlations.

To account for this, we rank-ordered the Millennium galaxies by luminosity and generated the same number of luminosities with an SDSS luminosity function (Blanton et al. 2003a), and assigned the galaxies with those luminosities, while preserving the same rank order. The luminosity functions are significantly different, as can be seen in Figure 6.2, which shows the original (2dF) and modified (SDSS) luminosity function of our $M_r < -20.5$ catalog. Note that $10^{11} L_\odot$ galaxies are extremely rare in the SDSS.

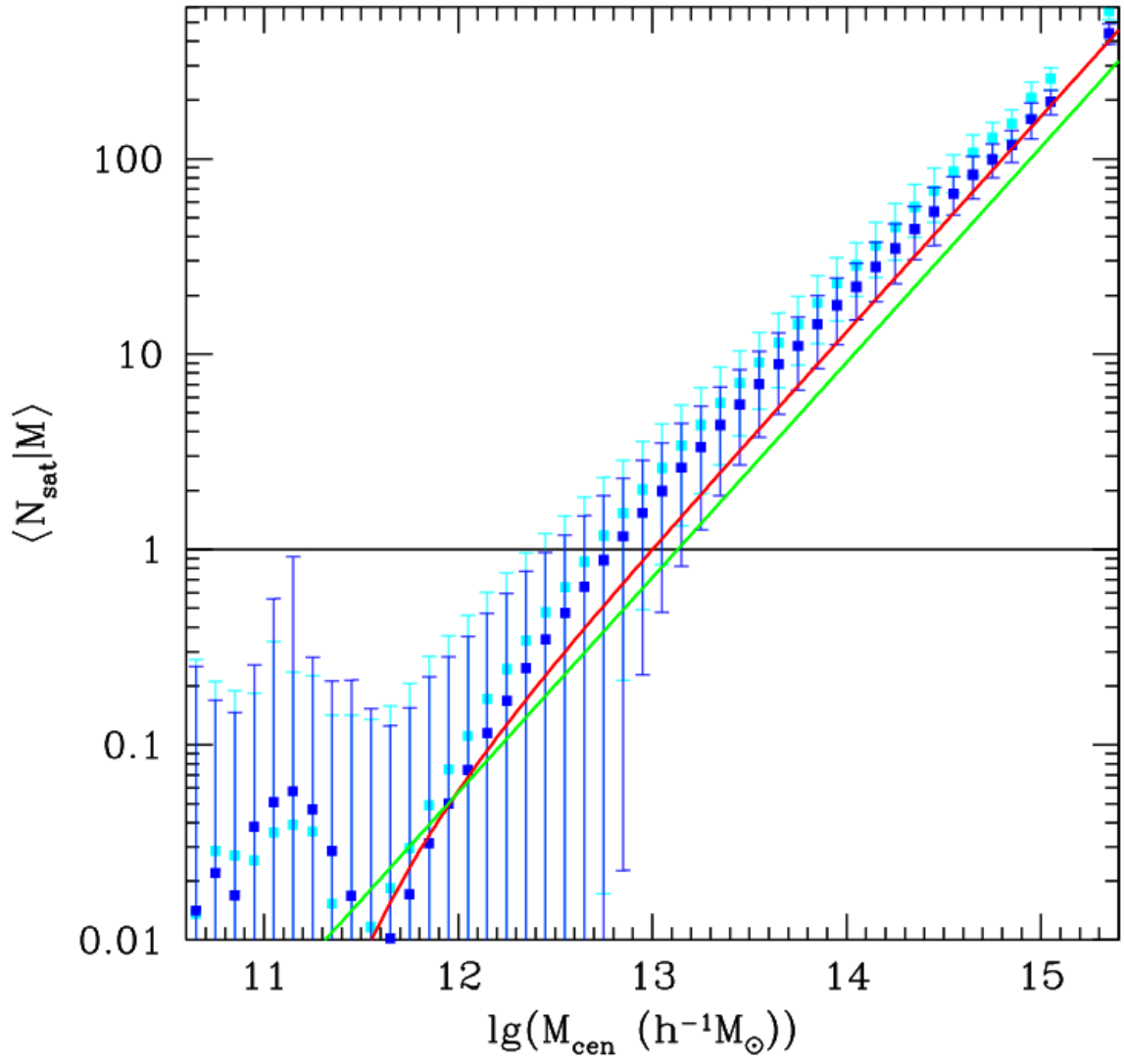

Figure 6.1: Satellite galaxy halo occupation distribution for $M_r < -19.5$. The red line is our assumed mean HOD, which is consistent with the measured $\xi(r)$. The green line is the corresponding result of Zehavi et al. (2005). The cyan and blue points are our measurements of the Millennium galaxy catalog prior and subsequent to doing the number density correction.

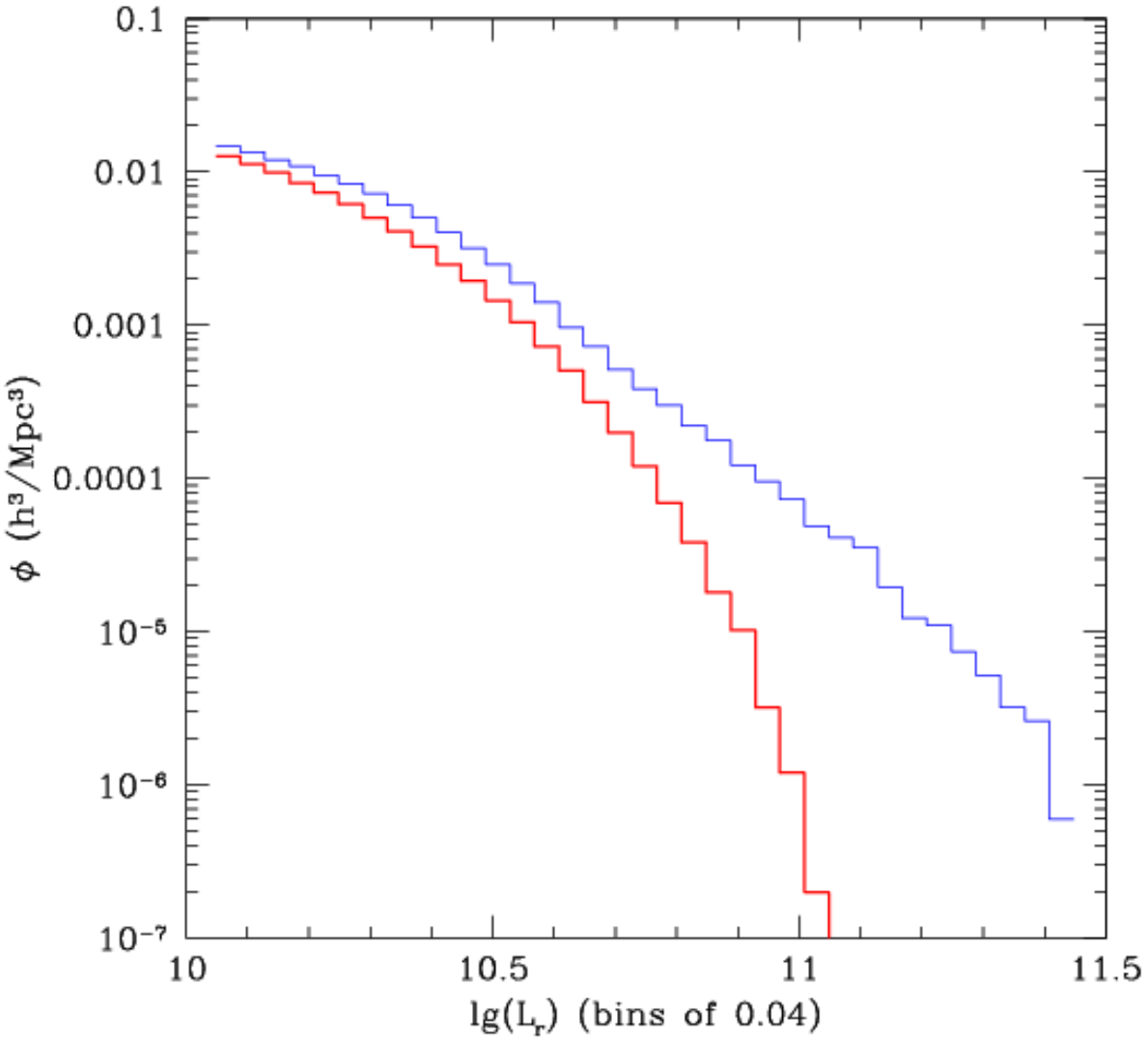

Figure 6.2: Luminosity function of $M_r < -20.5$ Millennium Run catalog. Blue histogram: unmodified LF (2dF); red histogram: rescaled LF (SDSS, from Blanton et al. 2003a).

The problem with such a rescaling of the luminosity marks is that it naïvely ignores the fact that central and satellite galaxies have significantly different distributions of their marks, whether the mark is luminosity, stellar mass, color, or star formation rate. Central galaxies tend to be more luminous, massive, redder, and more passively star forming than satellite galaxies (*e.g.* Weinmann et al. 2006a). This is also indirectly implied by the fact that subhalos are less massive than host halos and luminosity increases with halo mass, and color (SFR) increases (decreases) with stellar mass, and hence with halo mass (*e.g.,* Kauffmann et al. 2004, Brinchmann et al. 2004, Baldry et al. 2004).

To do a more direct comparison of luminosity-marked correlation functions to our halo model prediction, it is necessary to rescale the central and satellite galaxies separately since they have different luminosity functions. Our central galaxy luminosity function for the halo model is dependent on the halo mass function, dn/dM, and the mass-luminosity relation, $M(L)$:

$$\phi_{\rm cen}(L_{\rm cen}) = \frac{dn(M)}{dM}\frac{dM_{\rm halo}}{dL_{\rm cen}}.$$

The satellite luminosity function is slightly more complicated. Using the central/satellite information and halo masses, made available to us by Darren Croton, we created the list of host halo masses and the number of subhalos (hosting satellites) associated with each host. With this list we used the acceptance-rejection method to generate the required number of luminosities as a function of host halo mass, given the luminosity-dependent halo occupation distribution, $\langle N_{\rm sat}|M_h, L_{\rm min}\rangle$.

The luminosity functions of the central galaxies, satellite galaxies, and total population of galaxies of the $M_r < -19.5$ sample, prior to any rescaling, are the faint magenta, cyan, and gray histograms shown in Figure 6.3. At fainter thresholds, one expects a significant satellite population with luminosities that are mostly faint but also that overlap with those of the central galaxy population. This is evidently the case with the Millennium catalogs: at the faint end the satellite fraction clearly increases with decreasing luminosity.

The rescaled central, satellite, and total luminosity functions are the red, blue, and black histograms in Figure 6.3. As expected from the SDSS luminosity function, the measured luminosity functions drop more rapidly at the bright end. This can also be seen in the relation between central luminosity and halo mass prior and subsequent to rescaling, shown

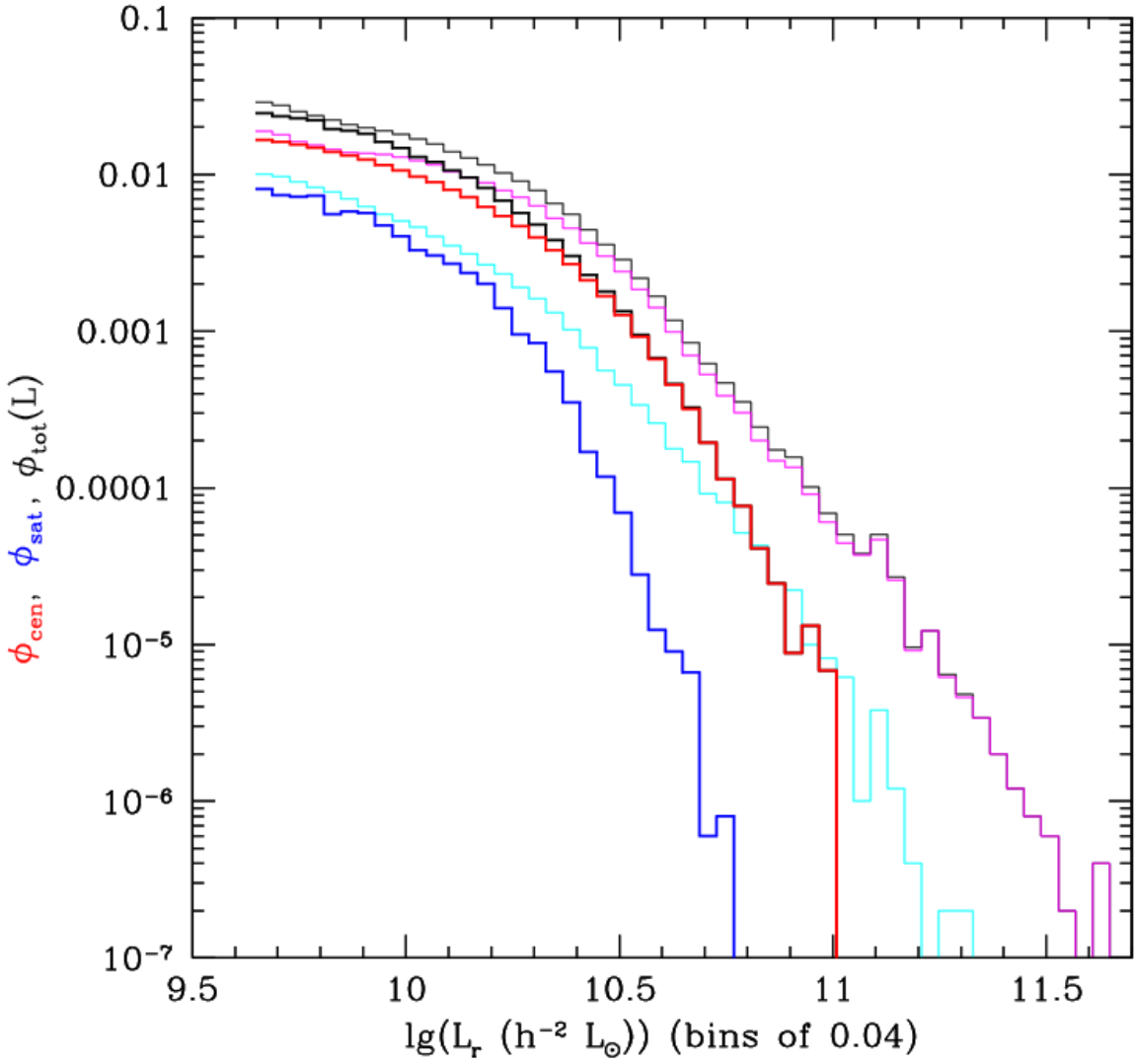

Figure 6.3: Central galaxy, satellite galaxy, and total luminosity functions for $M_r < -19.5$. Fainter histograms: unmodified central (magenta), satellite (cyan), and total (black) LFs. Dark histograms: rescaled central (red), satellite (blue), and total (black) LFs.

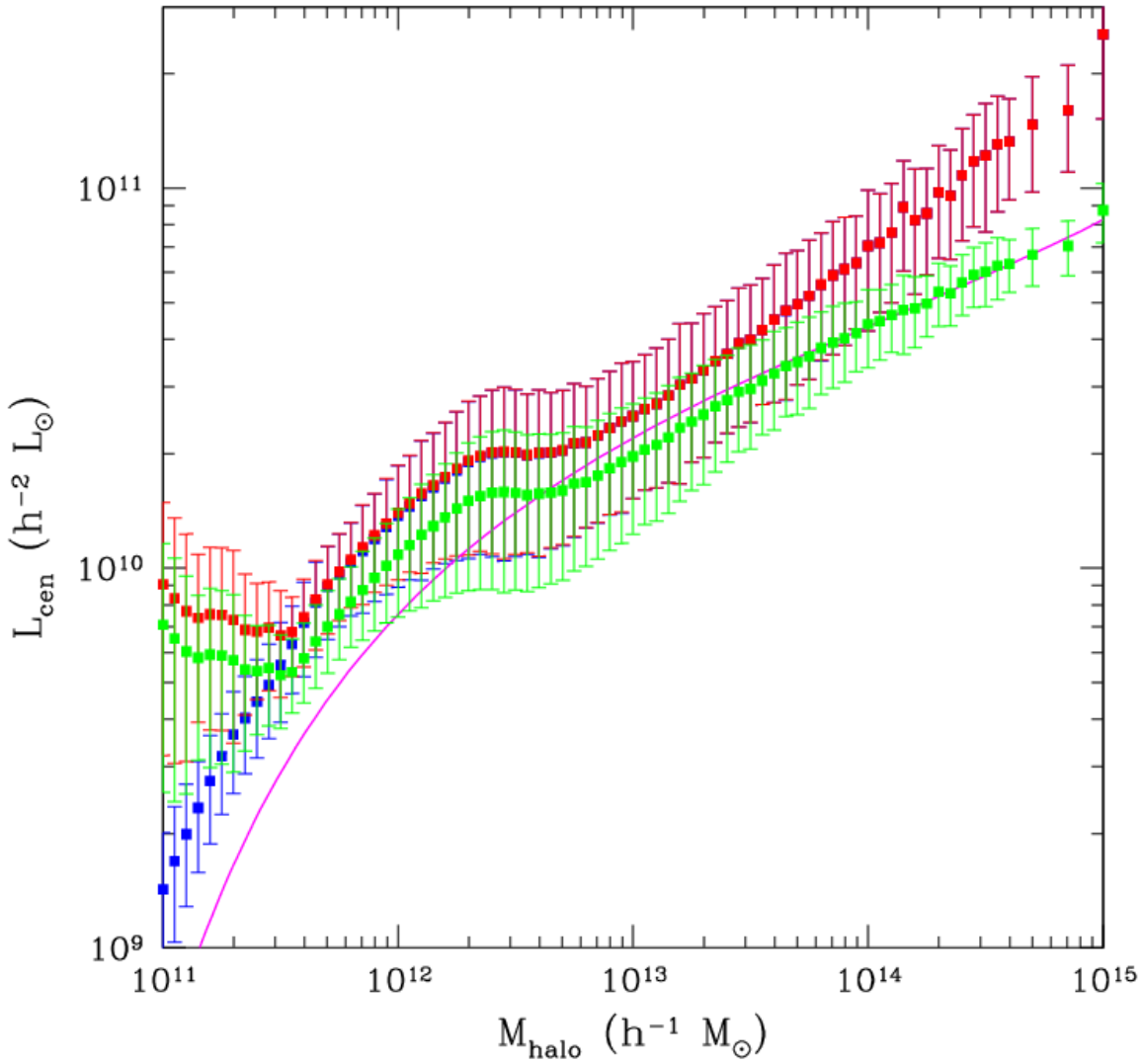

Figure 6.4: Relations between central galaxy Luminosity and host halo mass. $L(M)$ of whole Millennium Run (blue), of the $M_r < -19.5$ sample (red), of the rescaled $M_r < -19.5$ sample (green), and of our halo model, used for the rescaling (magenta curve).

in Figure 6.4. The rescaled luminosities (green points) are fainter at high masses than the unmodified luminosities (red points). Strangely, $L_{\rm cen}(M_h)$ appears to be independent of mass for $2 - 4 \times 10^{12} M_\odot$, which also affects the rescaled luminosities fainter than this. This curious feature in the simulation likely affects luminosity-marked statistics for fainter and brighter luminosity thresholds in different ways.

For the rescaled satellite luminosities, they are significant relative to the centrals now only up to luminosities of about $10^{10.4} L_\odot$, which corresponds to halo masses of 1 or $2 \times 10^{13} M_\odot$, according to Figure 6.7. The mean satellite luminosity is not *much* fainter than the mean central luminosity in Figure 6.3. When we use these rescaled luminosities to do luminosity-marked correlation functions, the numerous central galaxies will dominate the mean weight but luminosity-weighted galaxy pairs that consist of one or two satellites will consequently be weighted much less than pairs of central galaxies. Luminosity-marked correlation functions are clearly strongly affected by the central and satellite luminosity functions and the central and satellite fractions.

The satellite fraction for the luminosity thresholds under consideration is not negligible. For our Millennium catalogs, about 27%, 21%, and 15% of the galaxies are satellites in the $M_r < -19.5$, -20.5, and -21.5 samples, respectively. It is important to compare the relative number of satellite galaxies to the corresponding fractions in our SDSS-based halo model calculations. Since the number density of galaxies, $\bar{n}_{\rm gal}$, is given by equation (6.2), the fraction of satellites given a mass/luminosity threshold is given by

$$f_{\rm sat} = 1 - \frac{\int_{M_{\rm min}} dM \, \frac{dn(M)}{dM} \, N_{\rm cen}(M)}{\int_{M_{\rm min}} dM \, \frac{dn(M)}{dM} \, \langle N_{\rm gal} | M \rangle} \, ,$$

where $N_{\rm cen}(M)$ is assumed to be unity. The luminosity thresholds give a minimum mass $M_{\rm min}(L_{\rm min})$, which results in $f_{\rm sat} \approx 22\%$, 17% and 6% for $M_r < -19.5$, -20.5 and -21.5, respectively. This means that there are indeed more satellites than expected in the Millennium Run simulation: for the $M_r < -19.5$ catalog of 1.3m galaxies, there are more than 60,000 satellites–about one-fifth of them–that should be centrals.

It is not clear why there are so many satellite galaxies in the Millennium catalogs, especially because we have used the same cosmology and virialization density $\Delta_{\rm vir} = 200$. It is possible that some low-mass halos are flowing too rapidly through filaments and accreting

onto massive host halos earlier than expected, or that fewer subhalos than expected are being tidally disrupted and merging with the host halo. In any case, the excessive number of satellite galaxies will result in many satellites assigned with very faint luminosities when we rescale them. This will be important when we discuss our luminosity-marked results in the next section.

6.3.3 Luminosity-Marked Correlation Functions: Results

Marked correlations are always relative to the *mean* mark. Although for each catalog, using the SDSS luminosities decreases the mean luminosity somewhat, relative to the mean mark there are fewer very bright galaxies than if the original luminosities were used. In other words, there are fewer galaxies with large $L_i/\bar{L}$. As a result, it is not surprising that rescaling the luminosities to the SDSS luminosity function results in decreasing the luminosity-marked correlation function.

Our unweighted and luminosity marked correlation functions for $M_r < -19.5$ and $M_r < -20.5$ are shown in Figures 6.5 and 6.6. The $M_r < -21.5$ result (not shown) is fairly similar to the $M_r < -20.5$ result, but with larger error bars. The errors of $\xi(r)$ and $M(r)$ are estimates based on the variance of the Landy-Szalay (1993) estimator and on the variance of WW/DD (see the appendices of Sheth, Connolly & Skibba (2005) for details). These error estimates are comparable to jack-knife errors on scales less than a few Mpc, and they are underestimates at large scales.

Our halo model predictions for $\xi(r)$ were approximately consistent with the measurements when we used a halo occupation distribution with $M_1/M_{\rm min} \approx 30$, with an exponential low-mass roll-off in $\langle N_{\rm sat}|M\rangle$ with $\beta \approx 1$. In comparison, Zehavi et al. (2005) observed $M_1/M_{\rm min} = 23$ and found that an exponential roll-off was generally unnecessary for SDSS galaxies. We did adopt their same luminosity-dependent slope of $\langle N_{\rm sat}|M\rangle$, which was $\alpha \approx 1.1$, 1.2 and 1.9 for $M_r < -19.5$, -20.5 and -21.5, respectively. With this parameterization, and with the number density corrections to the data, $\xi(r)$ of the theory are in fairly good agreement with those of the Millennium catalogs.

Two halo model predictions of the marked correlations are shown in Figures 6.5 and 6.6.

The dashed curves are the results when all galaxies in a halo are naïvely weighted the same way, by the mean luminosity of all the galaxies hosted by the halo. The solid curves are the results when central galaxies are weighted by their luminosities $L_{\rm cen}(M)$ and satellite galaxies are weighted by their mean luminosities $\langle L_{\rm sat}|M\rangle$. The better agreement between the center-satellite model and the measurements provides evidence that central galaxies in halos are a decidedly special population.

The luminosity-marked correlation functions are clearly much weaker in the Millennium than they are in the halo model predictions. Even when we rescale the central and satellite luminosities, the discrepancy remains. In the brighter samples, after rescaling the luminosities the marked correlation functions weaken even further. The satellite fractions may be part of the explanation, but in any case, the small-scale environmental dependence of luminosity appears to be too weak in the simulation.

In general, in the halo model and SDSS volume-limited catalogs and our SDSS-based mock catalogs, the luminosity-marked correlations increase with decreasing $M_{\rm min}$ because of the increasing range of luminosities. This is not the case here: the faintest sample has weaker, not stronger marked correlations.

Furthermore, the small-scale marked correlation function is somewhat weaker for $M_r < -19.5$ than for $M_r < -20.5$. It is true that the one-halo satellite-satellite term has a larger contribution, relative to the center-satellite term, for the fainter threshold, but the larger range of luminosities should have a stronger effect. In general, we expect the luminosity-marked correlation functions to increase when the luminosity threshold is decreased, because that is what occurs in the SDSS measurements and in the halo model. When there is a larger range of weights, one is more likely to obtain galaxies with weights far from the mean, thus strengthening the marked signal. Our observation of the opposite trend in the Millennium simulation is somewhat troubling. However, the issue of the environmental dependence of luminosity in the Millennium Run is likely a tractable problem, as it is highly dependent on the assumed luminosity function in both simulations and the halo model. Interestingly, if one fits the central galaxy distribution as a function of mass and fits the first and second moments of the satellite halo occupation distribution, a discrepancy still occurs on small scales ($r < 1\,{\rm Mpc}/h$) even after accounting for the scatter in the HOD fits (see Figure 7.4

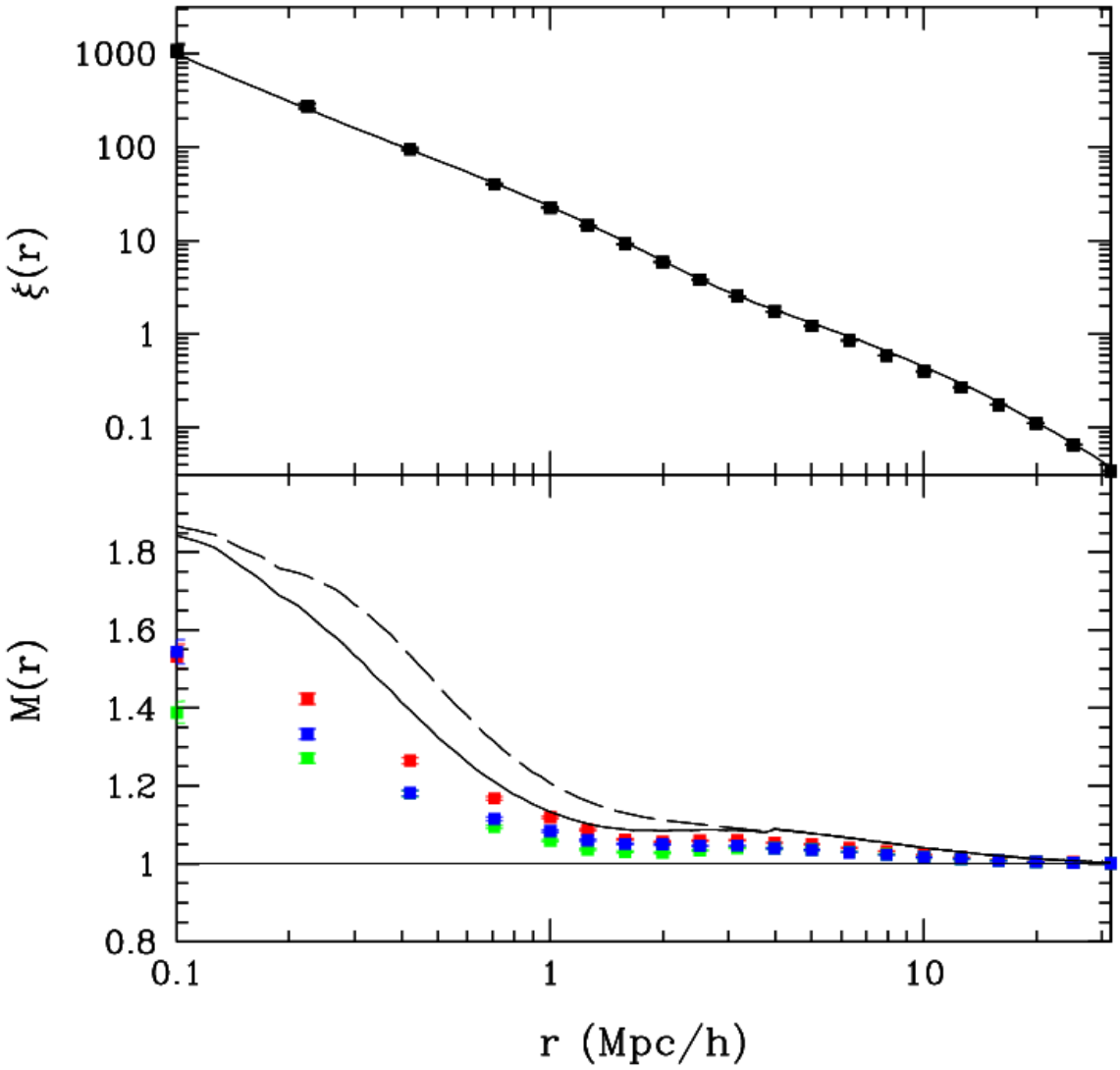

Figure 6.5: Unweighted and L_r-marked correlation function for $M_r < -19.5$. Top panel: red points show $\xi(r)$ for the unmodified Millennium catalog and blue points are the result with the correct SDSS number density. Top panel: black points show $\xi(r)$ for the Millennium catalog after correcting the number density. Bottom panel: red points show $M(r)$ for the unmodified luminosities and green points are the result when the luminosities have an SDSS luminosity function. Blue points are the result when we rescale the luminosities of the central and satellite galaxies separately. The dashed curve is the halo model prediction when all galaxies are weighted by the average luminosity in a halo, and the solid curve is the prediction for the center-satellite model.

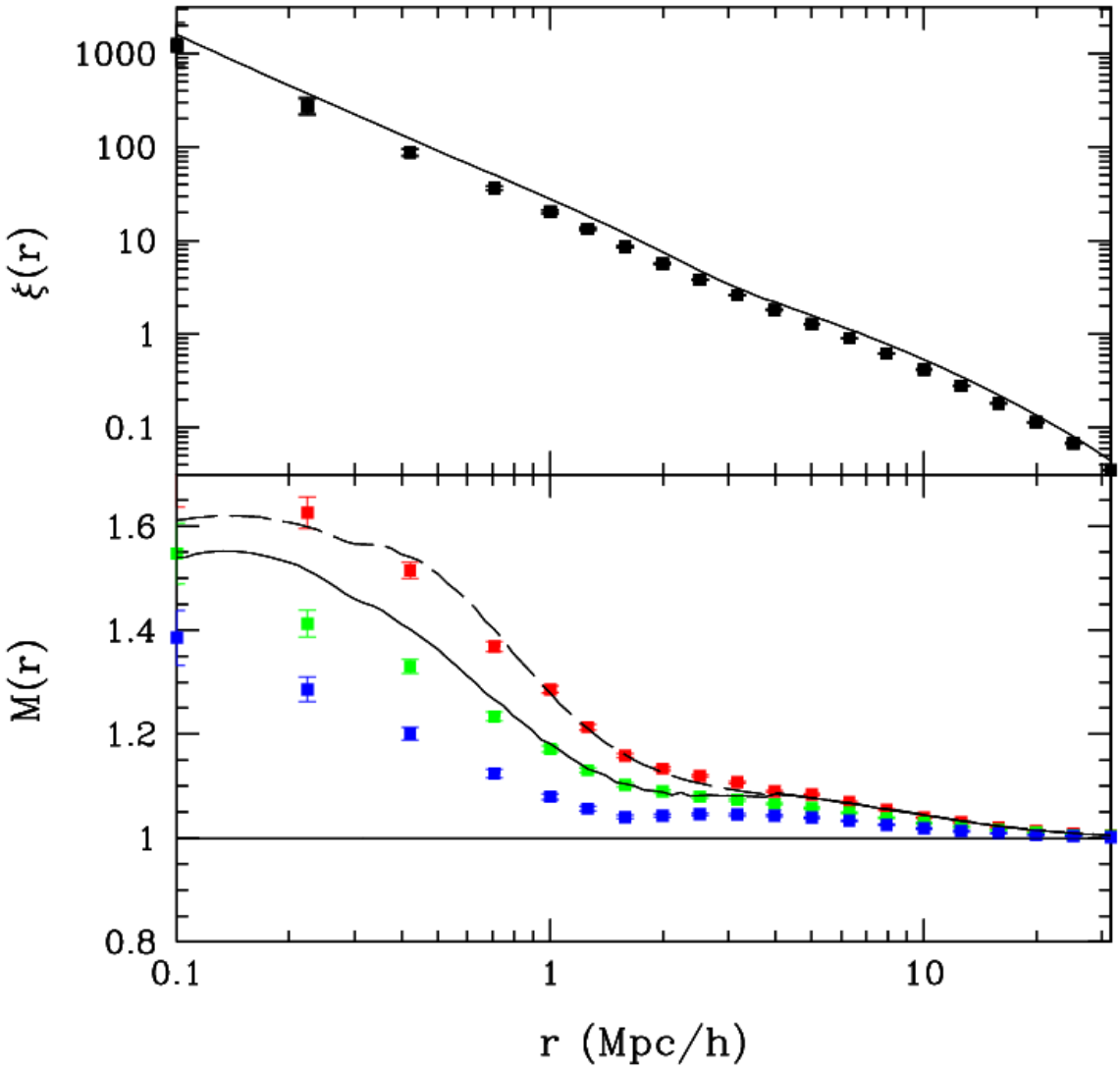

Figure 6.6: Unweighted and L_r-marked correlation function for $M_r < -20.5$. Top panel: red points show $\xi(r)$ for the unmodified Millennium catalog and blue points are the result with the correct SDSS number density. Bottom panel: red points show $M(r)$ for the unmodified luminosities and blue points are the result when the luminosities have an SDSS luminosity function. The dashed curve is the halo model prediction when all galaxies are weighted by the average luminosity in a halo, and the solid curve is the prediction for the center-satellite model.

and accompanying text). If one uses the Blanton et al. (2001) luminosity function, which is more '2dF-like', the small-scale discrepancy mostly disappears but then the simulation's marked correlation function is too strong on scales of $r \sim 1\,\mathrm{Mpc}/h$. It will be interesting to explore this issue further in the future.

As we will see in the next two sections, the environmental dependence of stellar mass and star formation rate in the Millennium Run may pose larger problems.

6.4 STELLAR MASS-MARKED CORRELATION FUNCTIONS

In this section we will compare stellar mass-marked correlations of the halo model to measurements with SDSS and MS data.

6.4.1 Stellar Mass Mark

For the stellar mass mark, it is necessary to determine an observationally-motivated relation between stellar mass and halo mass. We adopted two independent approaches for this: we used the r-band mass-to-light ratio observed by Bell et al. (2003) with the halo mass-luminosity relation we obtained earlier; we also used the relation between stellar mass and halo mass obtained from the weak gravitational lensing measurements of Mandelbaum et al. (2006).

These two $M_\star(M_h)$ relations differed slightly in amplitude but had a strikingly similar shape. It is consequently not surprising that they produced similar stellar mass-marked correlation functions, which are normalized by the mean stellar mass mark. Bell et al. (2003) found the following r-band mass-to-light ratio, as a function of $g-r$ color: . We used this to derive the mass-to-light ratio as a function of luminosity:

$$\left\langle \frac{M_\star}{L} \right\rangle = 4.2 \left(\frac{L}{L_*} \right)^{0.14} \frac{M_\odot}{L_\odot} \tag{6.12}$$

Bernardi et al. (2003) found a similar mass-to-light ratio for their sample of early-type galaxies, with a slope of 0.15 and an amplitude of approximately 3.9 for the L_* of our luminosity function.

Combined with our $L_r(M_h)$ from section 2.3, the mass-to-light ratio (12) produces an $M_\star(M_h)$ with similar slopes and a slightly lower amplitude compared to the one measured by Mandelbaum et al. (2006):

$$M_\star = \frac{2\, M_{\star,\,\rm norm}}{(M_h/M_0)^{-a} + (M_h/M_0)^{-b}}\,, \tag{6.13}$$

with the normalization $M_{\star,\,\rm norm} \approx 2 \times 10^{10} M_\odot/h$, the low-mass and high-mass slopes $a \approx 2$ and $b \approx 0.4$, and $M_0 \approx 3 \times 10^{11} M_\odot/h$ is the halo mass at which the slope of the relation changes. There is much uncertainty in these measurements, but mass-to-light ratio constraints will improve with additional weak lensing measurements in the future.

The $M_\star(M_h)$ result obtained by Shankar et al. (2006) is somewhat similar, but their high-mass slope is *much* steeper than that of both of these relations, possibly due to the different luminosity function they assumed. On the other hand, Wang et al. (2006) determine a weaker high-mass but their uncertainties are quite large.

Both of our stellar mass-halo mass relations, from the Bell et al. (2003) mass-to-light ratio and the Mandelbaum et al. (2006) weak lensing measurement, are shown in Figure 6.7. Their amplitudes differ by as much as 1.5, but it is the shape of the relation with halo mass on which the marked correlation function crucially depends. As will be shown below, their shapes are sufficiently similar to yield fairly similar predictions of stellar mass-marked clustering.

Our above two $M_\star(M_h)$ relations are estimates of the relationship between *central* galaxy stellar mass and host halo mass. We obtain the mean satellite stellar mass as a function of halo mass, $\langle M_{\star,\,\rm sat}|M, L_{\rm sat}\rangle$, the same way we obtained the mean satellite luminosity, by using $M_{\star,\,\rm min}(L_{\rm min})$ to integrate the stellar mass-dependent halo occupation distribution. In order to calculate stellar mass-marked correlations with the halo model, one needs the stellar mass dependence of the halo occupation distribution, which we assume to be similar to the luminosity dependence because of the strong correlation between them. We also assume that the density profile of the satellite galaxy distribution is mostly independent of stellar mass.

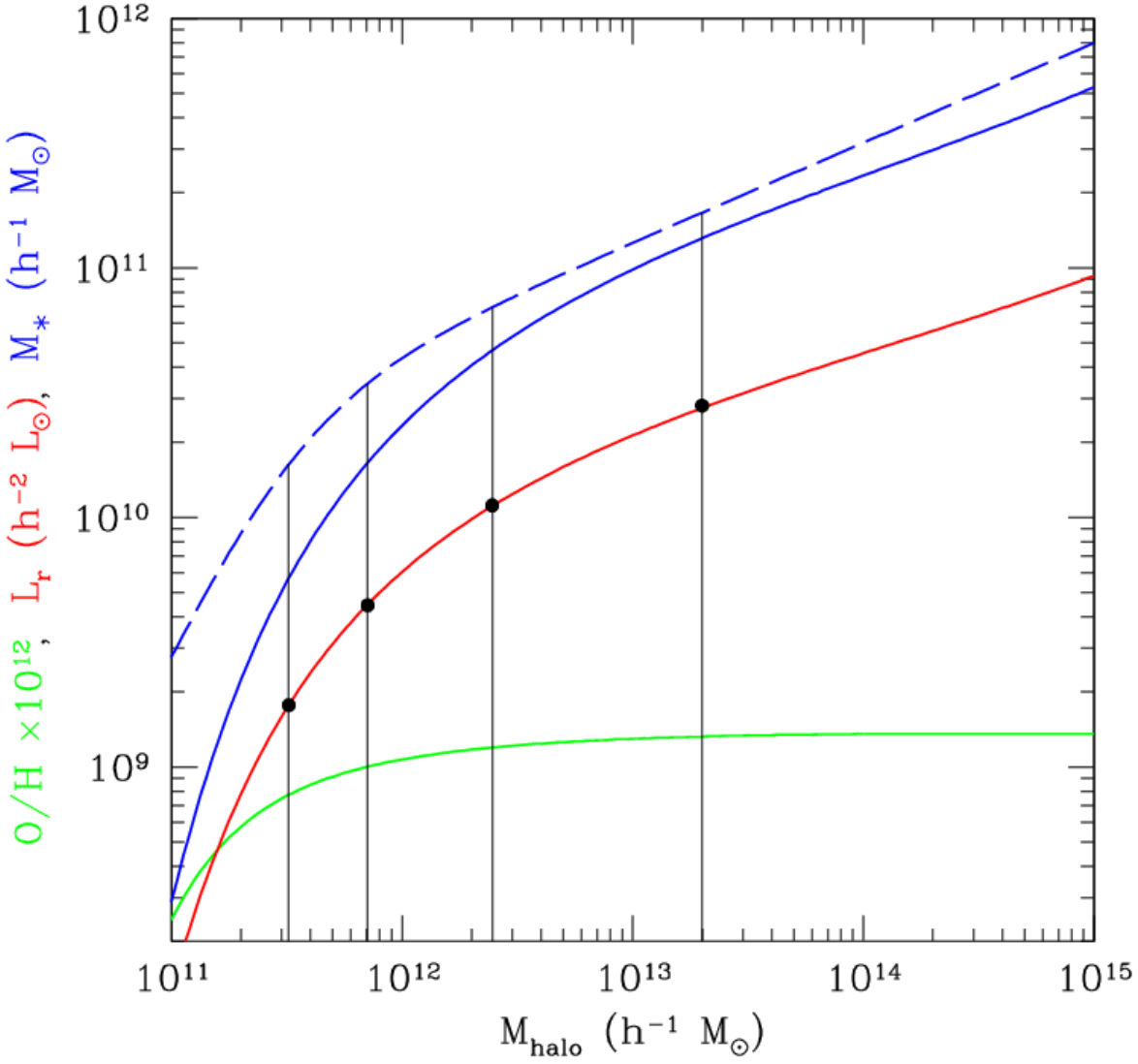

Figure 6.7: Relations between galaxy marks and halo mass. Red curve: r-band luminosity; solid blue curve: stellar mass from mass-to-light ratio (Bell et al. 2003); dashed blue curve: stellar mass from weak lensing measurement (Mandelbaum et al. 2005); green curve: O/H metallicity (Tremonti et al. 2004). All relations shown, except for dashed blue curve, assume SDSS cosmology, LF of Blanton et al. (2003a) at $z = 0.1$, and luminosity-dependent HOD of Zehavi et al. (2005).

6.4.2 Halo Model *vs.* SDSS

We used the stellar mass catalog of Kauffmann et al. (2006), who estimated the stellar masses using the procedures described in Kauffmann et al. (2003). The galaxies in the catalog are in the SDSS Data Release 4 (Adelman-McCarthy et al. 2006). We K-corrected the apparent magnitudes using the kcorrect code of Blanton et al. (2003a). This allowed us to create multiple volume-limited catalogs, which are necessary for accurate measurements of marked statistics. The two we will compare to Millennium catalogs have the following cuts: $-23.5 < M_r < -19.5$, $0.017 < z < 0.082$; $-23.5 < M_r < -20.5$, $0.019 < z < 0.125$. These catalogs contained 102,302 and 150,046 galaxies, respectively.

For our halo model calculations, we used the halo occupation distribution of Zehavi et al. (2004), with a step function for the central galaxies and a power-law with $M_1/M_{\rm min} = 23$ and a stellar mass-dependent slope. We used the mean redshift of each of the samples, $\bar{z} = 0.065$ and $\bar{z} = 0.095$, and we assumed $\Omega_m = 0.3$ and the linear power spectrum of Efstathiou, Bond & White (1992), because it is the power spectrum Zehavi et al. used, and it is slightly lower than that of CMBFAST at large scales. For the stellar mass marks, we have used the same halo model code as for luminosity marks, but we are now weighting by $L\,\langle M_\star/L\rangle(M_h)$ using the mass-to-light ratio of Bell et al. (2003) and by $M_\star(M_h)$ using the lensing measurement of Mandelbaum et al. (2006).

Our resulting projected correlation function and stellar mass-marked correlation function are compared to measurements of our SDSS samples in Figures 6.8 and 6.9. In both plots, each of the stellar mass-halo mass relations produce similar marked correlation functions, and this is important because they were determined from independent observations. In each case, the center-satellite model, in which the central and satellite galaxies are weighted differently by their respective mean stellar masses, is in good agreement with the data. This is strong evidence that central and satellite galaxies are distinct populations that develop differently.

This is also strong evidence that the environmental dependence of the formation and evolution of their stellar masses is primarily driven by the environmental dependence of halo mass. This is also the conclusion we reached about galaxy luminosity in Skibba et al.

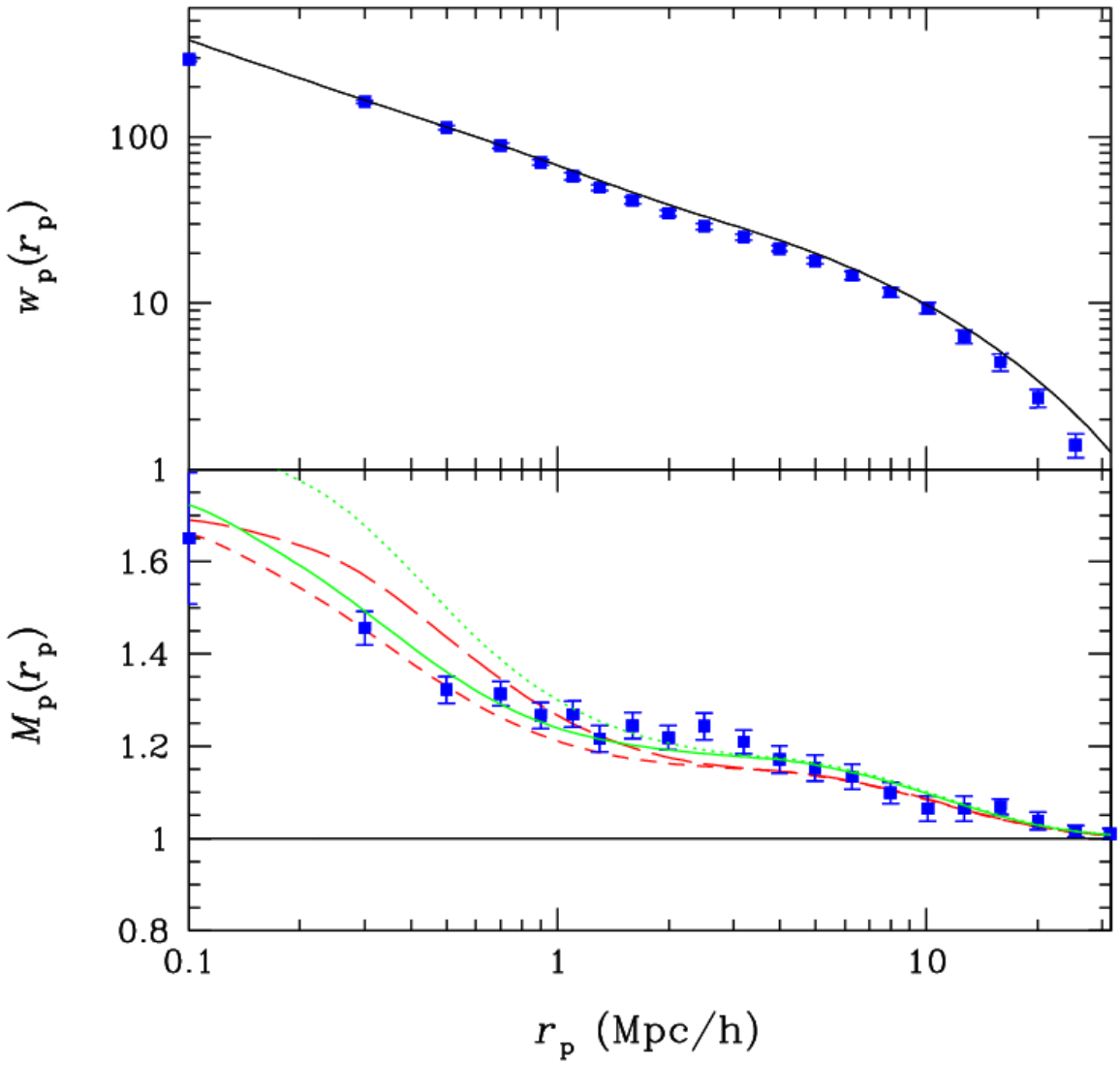

Figure 6.8: Stellar mass-marked correlations of $M_r < -19.5$ SDSS catalog. Solid and dotted (green) curves: $M_{M_\star}(r_p)$ from $L(M) + \langle M_\star/L \rangle$. Short-dashed and long-dashed (red) curves: $M_{M_\star}(r_p)$ from the lensing relation. The pair of lower, solid and short-dashed curves are the result of the center-satellite model.

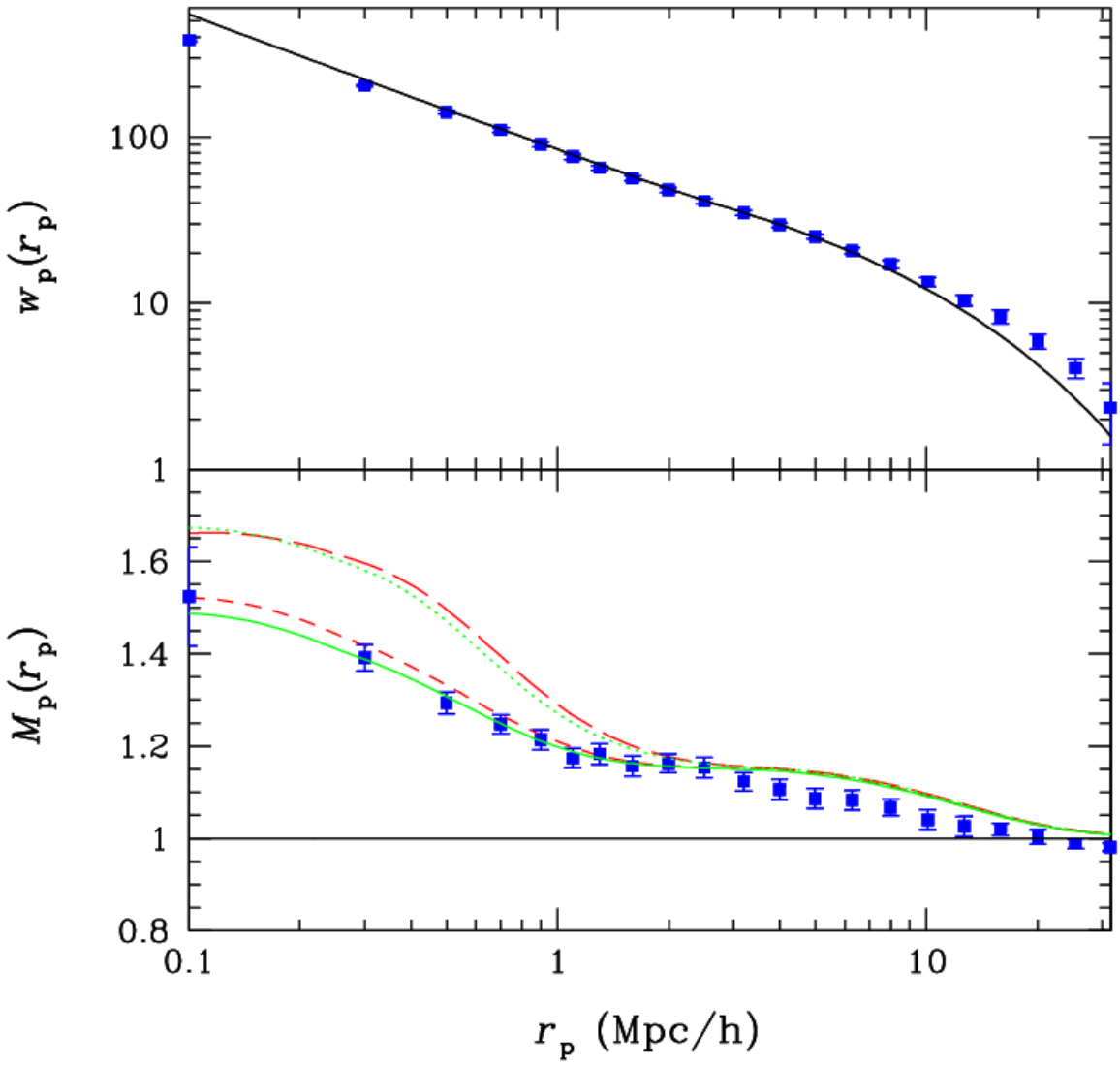

Figure 6.9: Stellar mass-marked correlations of $M_r < -20.5$ SDSS catalog. Solid and dotted (green) curves: $M_{M_\star}(r_p)$ from $L(M) + \langle M_\star/L \rangle$. Short-dashed and long-dashed (red) curves: $M_{M_\star}(r_p)$ from the lensing relation. The pair of lower, solid and short-dashed curves are the result of the center-satellite model.

(2006), although the environmental dependence of stellar mass is stronger. In addition, we have not included any scatter in the $M_\star(M_h)$ relation for the stellar mass marks. That is, we have assumed that the scatter in the mass-to-light ratio (0.1-0.2 dex scatter in the Bell et al. (2003) color versus M/L ratio) *is not correlated with the large-scale environment.* Our results suggest, although many galaxies experience numerous mergers, much of the observed scatter in mass-to-light ratios may be due to local small-scale processes of galaxy formation and evolution, such as supernova and AGN feedback, ram-pressure stripping, and galaxy 'strangulation'.

6.4.3 Halo Model *vs.* Millennium Run

The stellar mass-marked correlation functions of the Millennium simulation catalogs are much different than those of the SDSS. Using the Millennium HOD from Section 2 (and $\Omega_m = 0.25$, $z = 0$, and CMBFAST $P_{\rm lin}(k)$), we obtained stellar mass-marked correlation functions from the halo model to compare to Millennium Run simulation measurements. Our stellar mass-marked correlation functions are in excellent agreement with the SDSS, as shown in the preceding section, and we have used the same modeling for the stellar mass mark here. The SDSS and MS stellar mass distributions are compared in Figure 6.10, and they are different both below the "minimum" stellar mass corresponding to the minimum halo mass, and above it, with the simulation containing many more galaxies with $M_* > 10^{12} M_\odot$, which will amount to more strong weights in the weighted correlation function.

As can be seen in Figure 6.12, the environmental dependence of stellar mass is *much* stronger in the Millennium simulation than in the halo model and SDSS up to scales of 10 Mpc/h. The vast discrepancy increases at small scales, such that the environmental dependence in the Millennium at $r < 150\,{\rm kpc}/h$ is stronger by a factor of more than 1.5 in the brighter sample and stronger by a factor of 1.7 in the fainter one. This discrepancy is far larger than what we have observed with luminosity, and it is not so easily resolved.

To do a more direct comparison, we used the same $M_\star(M_h)$ relations obtained from the Bell et al. (2003) mass-to-light ratio and the Mandelbaum et al. (2006) weak lensing measurements to rescale the central and satellite galaxy stellar masses. The central and

satellite stellar mass functions, prior and subsequent to the rescaling, are shown in Figure 6.11 for $M_r < -19.5$. The rescaling changes both distributions substantially, resulting in very few central galaxies with $M_\star > 10^{12} M_\odot$, and resulting in lower stellar masses for the satellite galaxies, although they still constitute a significant fraction of the galaxies with masses of $10^{11} M_\odot$. The relatively lower satellite stellar masses are now more different than the unmodified central galaxy mass function, and consequently with the stellar mass mark one expects a weaker marked correlation function.

We compare our stellar mass-marked correlation function from the halo model to measurements from the $M_r < -19.5$ and $M_r < -20.5$ Millennium catalogs in Figure 6.12. Only the marked correlation functions are shown, as the unmarked correlation functions $\xi(r)$ are the same as the upper panels of Figures 6.5 and 6.6. Although the two rescaled measurements are not exactly the same in either panel, they are fairly close to the theoretical predictions, in spite of the small error bars. This suggests that our modeling of the stellar mass mark and the halo occupation distribution's dependence on it are reasonable.

It is strikingly clear that the environmental dependence of stellar mass in the Millennium simulation is *much* stronger than the environmental dependence we observe in SDSS stellar masses and in the halo model predictions using mass-to-light ratios and lensing measurements. When we rescale the stellar masses the marked correlation functions agree well, but this begs the question of the origin of the present-day stellar mass distribution. The distributions of SDSS and MS stellar masses are clearly different, and this results in very different correlations between stellar mass and large-scale environment. This vast discrepancy is a critical issue that semi-analytic modelers must grapple with. Just as the discrepancy in the environmental dependence of luminosity cannot be resolved simply be adopting a different luminosity function, this is not something that can be simply resolved by adopting a different initial mass function. The chosen IMF is obviously crucially important, but stellar mass-marked statistics also depend on the uncertainties in the stellar masses themselves, the relative fractions of central and satellite galaxies, and their respective stellar mass distributions.

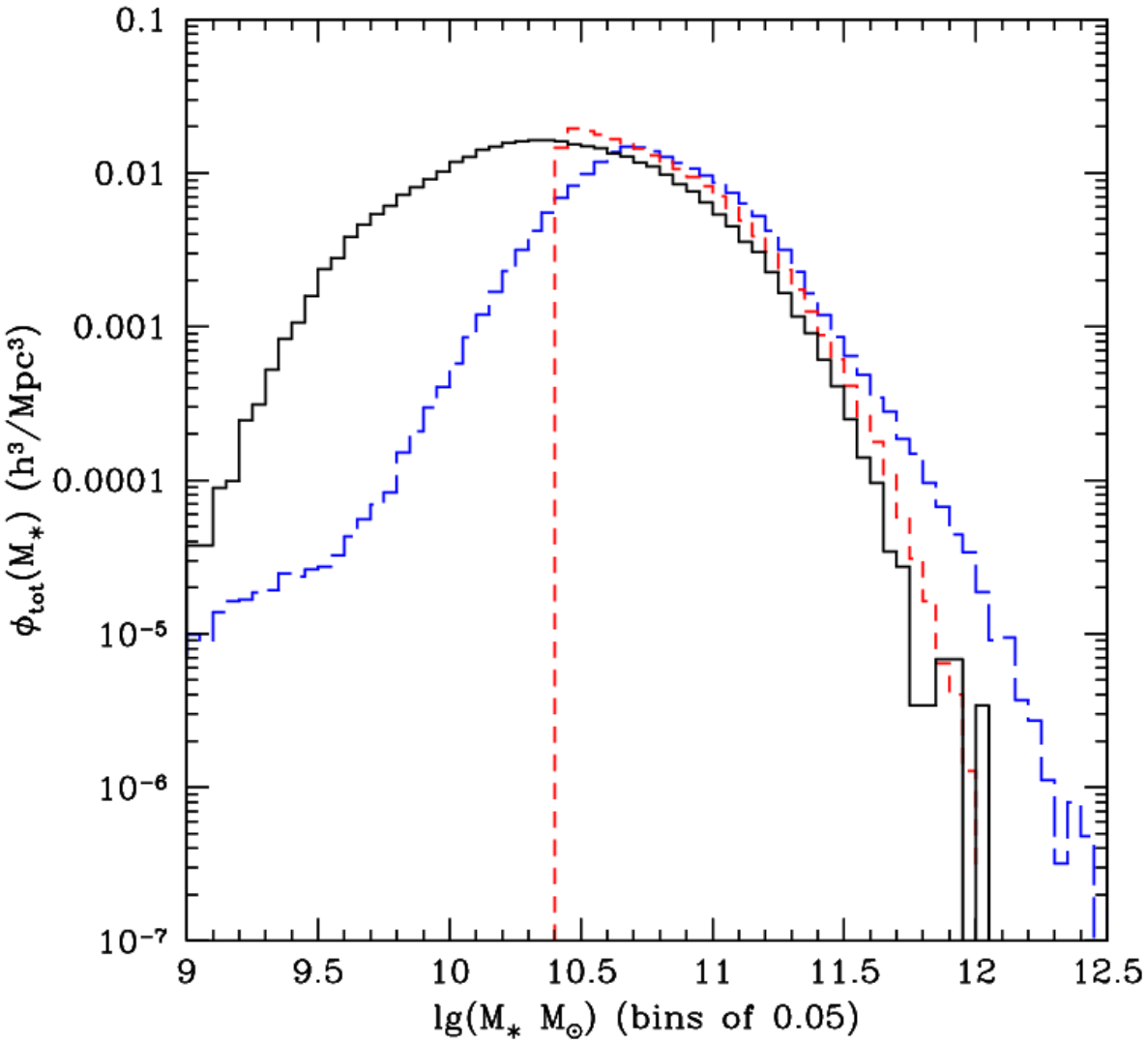

Figure 6.10: Stellar mass distributions for $M_r < -19.5$. Black histogram: SDSS; blue histogram: unmodified Millennium Run catalog; red histogram: rescaled Millennium catalog.

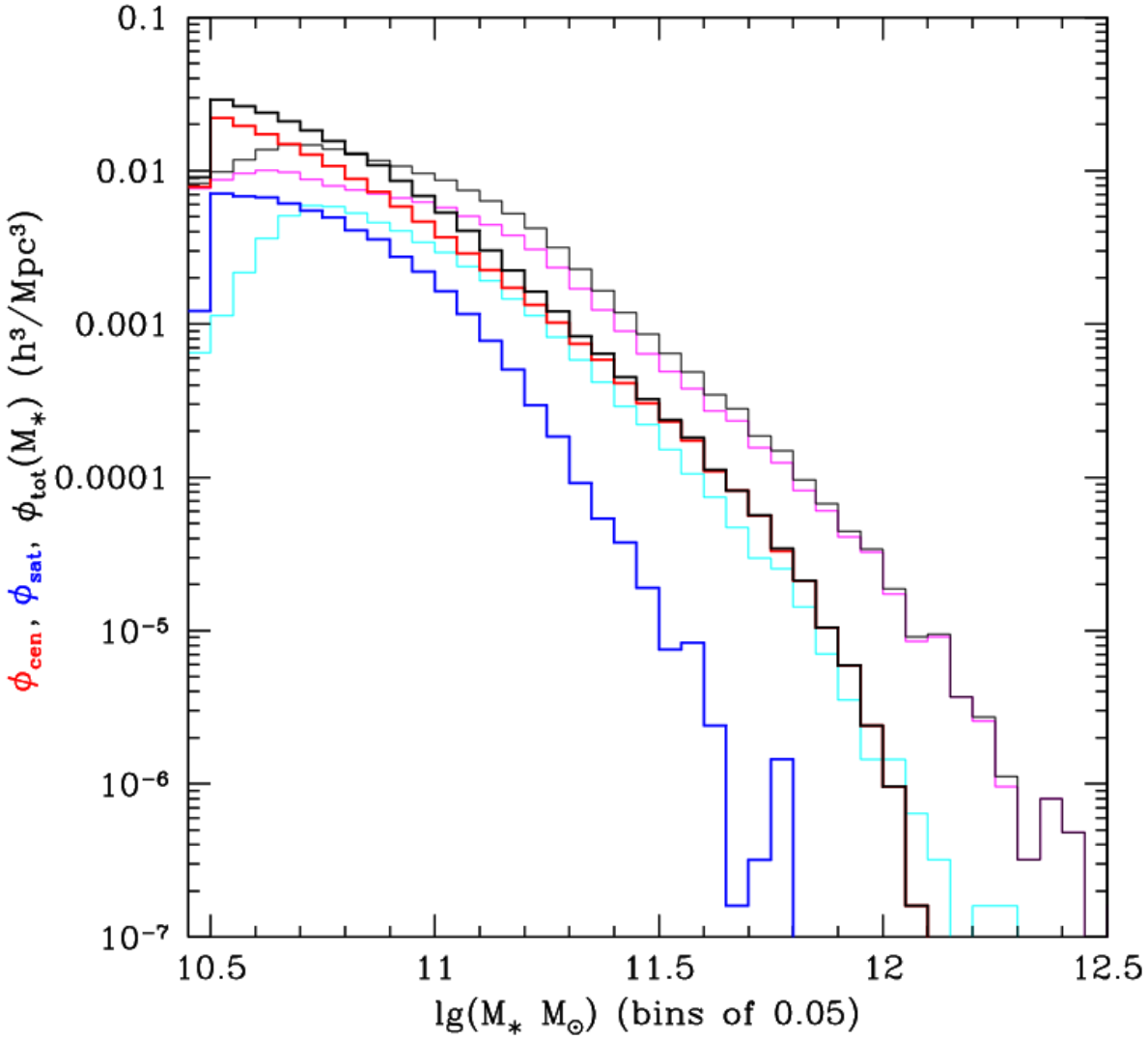

Figure 6.11: Stellar mass distributions for $M_r < -19.5$ Millennium Run catalog. Original distributions of central galaxies (magenta), satellite galaxies (cyan), and all galaxies (gray). Rescaled distributions of central galaxies (red), satellite galaxies (blue), and all galaxies (black).

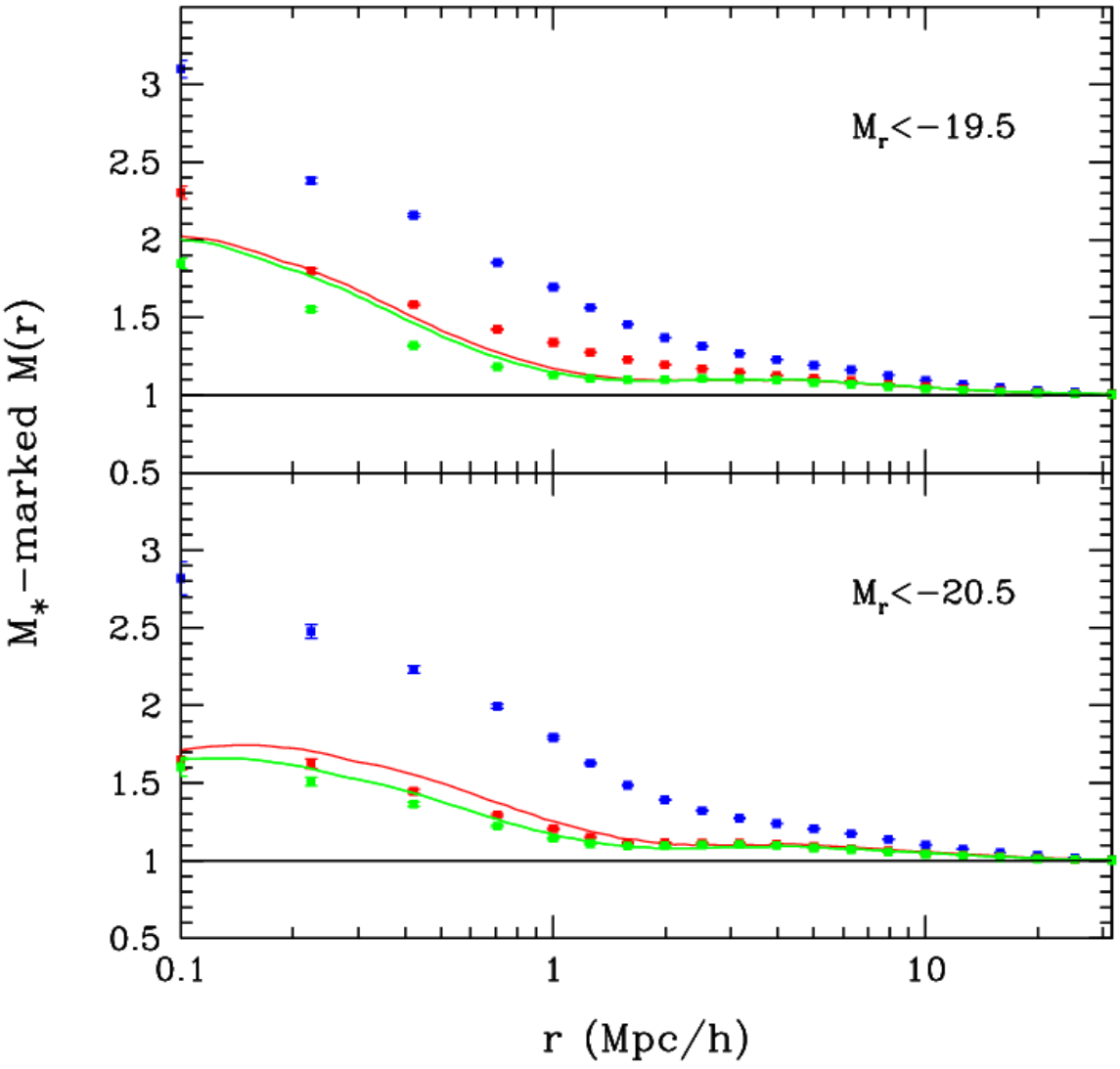

Figure 6.12: Stellar mass-marked correlations of $Mr < -19.5$ and $M_r < -20.5$ Millennium catalogs. As in Figures 6.8 and 6.9, the green curves are the results using the mass-to-light ratio measurement and the red curves are the results using the weak lensing measurement. Millennium Simulation measurements are shown for the original catalogs (blue points), rescaled masses using $M_*(M_h)$ from mass-to-light ratio (green points), rescaled masses using $M_*(M_h)$ from weak lensing (red points).

6.4.4 Metallicity Mark

In Section 4.2, our marked correlation functions show that the environmental dependence of stellar mass in the halo model is consistent with that of the SDSS, suggesting that the observed correlation with large-scale environment is primarily due to the correlation between halo mass and environment. Accordingly, it is interesting to extend our analysis by investigating the environmental dependence of metallicity.

The chemical properties of galaxies couple to their recent star formation histories and are anti-correlated with stellar age (*e.g.* Bernardi et al. 2005, Gallazzi et al. 2005), and metallicity has also been observed to be tightly correlated with stellar mass and luminosity (Tremonti et al. 2004). The relation between metallicity, mass, and age may simply be due to the fact that low-mass galaxies lose more of their metals because of galactic winds and their smaller potential wells. Of course, the real picture is likely much more complicated, as most galaxies experience both inflows and outflows of gas, and star formation and gas recycling are neither instantaneous nor completely efficient. We can complement research concerning these issues by measuring the dependence of galaxies' metallicity on large-scale environment. As we have modeled luminosity and stellar mass marks and measured their clustering, this allows us to model metallicity-marked clustering statistics and compare them to our measurements.

In practice, of course, there are multiple ways to quantify 'metallicity'. Following Tremonti et al. (2004), we use the ratio of the abundance of oxygen to hydrogen, O/H, because the emission lines involved have strong S/N and are free of uncertainties due to age and the relative abundances α-elements enhancement that affect absorption lines. Using their sample of star-forming SDSS galaxies, Tremonti et al. measure the following relation between stellar mass and gas-phase metallicity:

$$12 + \log_{10}(O/H) \;=\; -1.492 \;+\; 1.847(\log_{10}M_*) \;-\; 0.08026(\log_{10}M_*) \qquad (6.14)$$

We combine this with our relation between stellar mass and halo mass to determine a metallicity-halo mass relation, which is shown in Figure 6.7. As can be seen, the dependence on halo mass is relatively weak, which means that in order to measure a marked signal

strong enough to be inconsistent with unity (no signal at all), we need to use faint luminosity thresholds. Therefore, our metallicity-halo mass relation is the one we obtain from our stellar mass-halo mass relation using the Bell et al. (2003) mass-to-light ratio measurement, as opposed to the stellar mass-halo mass relation of Mandelbaum et al. (2006) which is unconstrained at low masses. The solar value of $12 + \log_{10}(O/H)$ is 8.69, so our range of O/H metallicities extends from slightly subsolar to less than thrice solar at the high-mass end.

For our faint luminosity thresholds we expect a significant satellite galaxy fraction, and we continue to distinguish between the central and satellite galaxy marks, although in the case of O/H metallicities there is not that much difference. To obtain the satellite metallicities, we assume the same stellar mass-metallicity scaling as we apply to the central galaxies, although there is no reason *a priori* that they should be the same. The $O/H(M_{halo})$ relation is slightly different for central and satellite galaxies, and this is because metallicity increases more with mass in central galaxies than does the halo occupation distribution of the satellites.

We also use the Tremonti et al. (2004) SDSS DR4 O/H metallicities for our clustering measurements. They estimated the metallicities with a model based on fits of all the most prominent emission lines ([OII], Hβ, [OIII], He I, [OI], Hα, [NII], [SII]). The total dataset consists of more than 100,000 metallicities, and we construct two volume-limited catalogs from them: $-23.5 < M_r < -18.5$, $0.015 < z < 0.5$, containing 10,999 galaxies; $-23.5 < M_r < -19.5$, $0.017 < z < 0.082$, containing 16,014 galaxies. Also, many of the galaxies with $12 + \log_{10}(O/H) < 8.5$ are at low redshift and lack a measured [OII], which leads to larger uncertainties in the derived metallicities, so we minimize this by excluding the galaxies with such low metallicities. For our redshift and luminosity cuts, however, this requirement only excludes an additional small number of galaxies.

Our O/H-marked projected correlation function results are shown in Figure 6.13 for both $M_r < -18.5$ and $M_r < -19.5$. The marked signals in both the observations and the halo model predictions are weak, and this is mostly due to the narrow range of metallicities involved: $7.74\times10^8 < O/H\,10^{12} < 1.36\times10^9$ for $M_r < -18.5$ and $1.00\times10^9 < O/H\,10^{12} < 1.36 \times 10^9$ for $M_r < -19.5$. The relatively large error bars, which were estimated using 30 jack-knife samples in each case, are due to the small sample sizes. In spite of the large

uncertainties here, it is encouraging that we have agreement between the halo model and SDSS observations for the environmental dependence of both stellar mass and metallicity. It will be interesting to follow-up this analysis with comparisons with larger samples of metallicities and with samples of the α-element abundance ratio α/Fe in the future.

Since metallicity and age are correlated, our results show that the stellar populations of the massive halos in dense environments tend to be older and metal rich, and galaxies in underdense environments tend to be younger and metal-poor. Our results seem to be consistent with those of Sheth et al. (2006) and Gallazzi et al. (2006), although both of their samples are more luminous than ours. Our marked statistics are somewhat better than the former and ours is a more direct indication of environmental dependence than the latter.

Because of our relatively faint sample selections, we have shown that the stellar metallicities of less massive galaxies increases in high-density regions. Since the anti-correlation between age and metallicity has been observed in both low- and high-density environments (*e.g.* Bernardi et al. 2005), this indicates that even low-mass galaxies in dense environments tend to be more evolved than their counterparts in low-density regions. Chemical enrichment and star formation timescales are related to each other, and consequently, we should expect more passive star formation rates in dense environments than in underdense regions for a large range of halo masses. We explore this in the next section.

6.5 STAR FORMATION RATE-MARKED CORRELATION FUNCTIONS: SDSS *VS.* MS

To complement our analyses of the correlations of stellar mass and metallicity with large-scale environment, we compare the SDSS and Millennium run simulation star formation rate-marked correlation functions in this section. Modeling star formation rate-marked statistics in the halo model is a difficult task: it involves modeling the relation between the SFR mark and halo mass, as well as modeling in an observationally-motivated way the SFR-dependent halo occupation distribution and SFR-dependent satellite galaxy density profile, which is important at small scales. It will also involve accounting for the position-dependence of

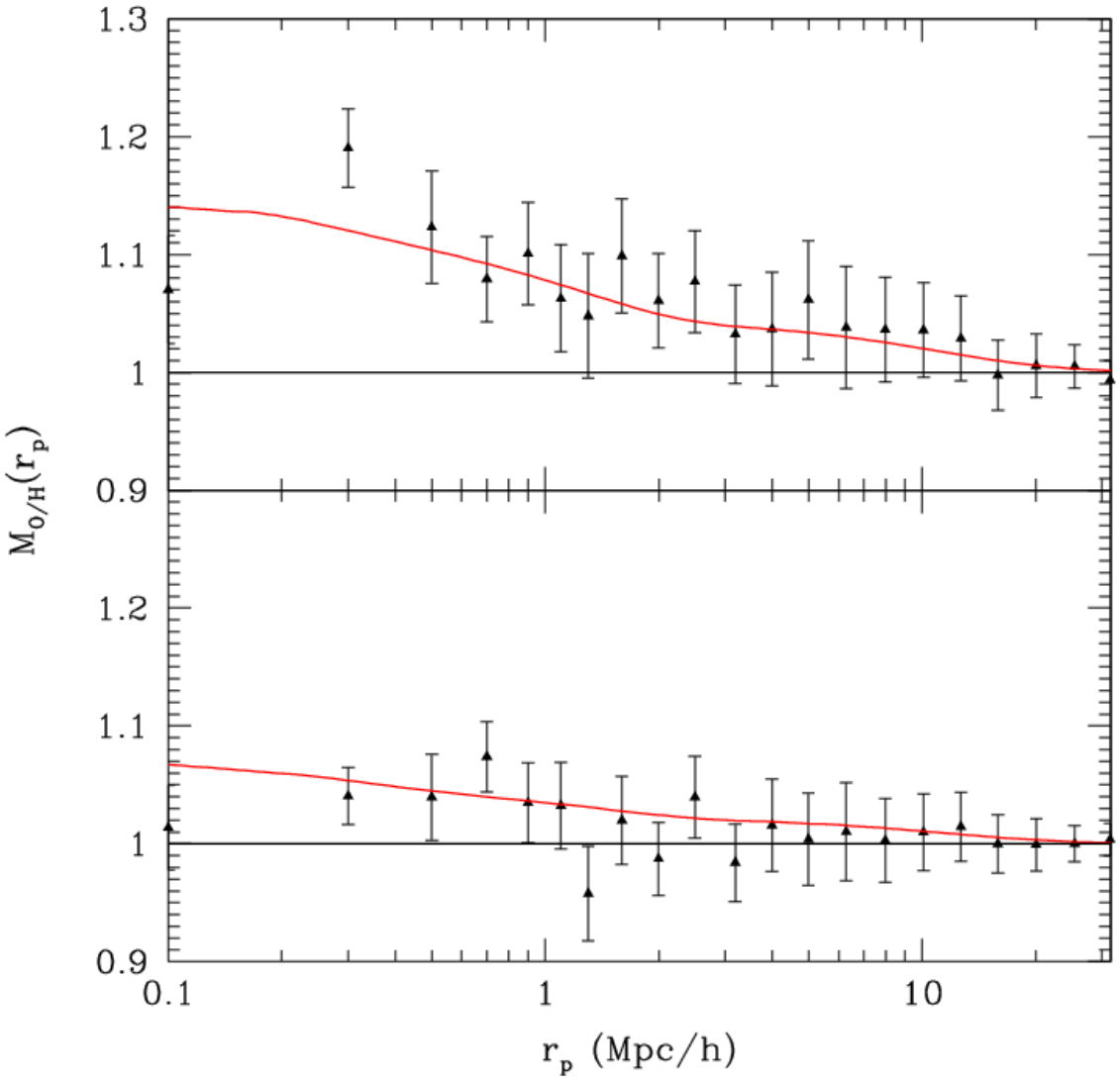

Figure 6.13: O/H metallicity-marked correlation functions of volume-limited SDSS catalogs $M_r < -18.5$ & $M_r < -19.5$, compared to halo model prediction using $L_r(M_h)$ relation + Bell et al. 2003 mass-to-light ratio + Tremonti et al. 2004 mass-metallicity relation.

SFR within halos, as satellite galaxies' SFRs have been observed to be strongly dependent on their positions within clusters and on their luminosities (Weinmann et al. 2006a). Our model will be somewhat similar to, but more sophisticated than, the conditional luminosity function model of Cooray (2005a). We will leave this for a future paper.

For the SDSS, we construct our volume-limited catalogs from the Brinchmann et al. (2004) DR4 data. We want to examine the environmental dependence of star formation rate, but it is not *a priori* clear *which* 'star formation rate' mark will be most useful. Firstly, there are many SFR indicators, and they have different advantages. For example, the Hα emission line probes star formation on a time-scale equal to the lifetime of a typical HII region ($\sim 10^7$ yr), while the H$\delta_{\rm A}$ absorption-line index peaks about 3×10^8 yr after an episode of star formation. In contrast, the λ4000 break increases monotonically with time and can be sensitive to decreasing star formation activity over relatively long time-scales (~ 1 Gyr) (Kauffmann et al. 2004). One can also use colors, such as $u-g$ or $u-r$ as indicators of SFR, but their marked correlation functions are weak and are difficult to interpret. Many of the observational difficulties of various SFR tracers, especially Hα versus [OII] luminosities, are described in detail in Hopkins et al. (2003) and Moustakas et al. (2006).

Brinchmann et al. (2004) have used multiple indicators to estimate the star formation rates of SDSS galaxies. We expect there to be less star formation in denser environments, but the environmental dependence of star formation rate is complicated by the fact that it tends to increase with stellar mass (see Figure 17 of Brinchmann et al.), while massive galaxies tend to reside in overdense regions (Figures 6.8, 6.9, and 6.12 above). The specific star formation rate, star formation per unit stellar mass SFR/M_*, is a useful quantity because it can account for this, allowing us to examine the environmental dependence of star formation activity relative to other galaxy formation processes, and it is insensitive to the initial mass function.

Secondly, we can use either the star formation rate inside the fiber or the estimated total star formation rate. The former are interesting because the central bulges of galaxies maybe affected by nearby galaxies and gas can be channeled down the potential well. On the other hand, the outer disks of spirals might be more affected by the large-scale environment (Brinchmann, private communication). In addition, the total star formation rates are even

more affected by the uncertainties involved in aperture corrections. At low redshifts in the SDSS the fiber aperture will only see a small part of a galaxy, and even at median redshifts the spectra sample only $\approx 1/3$ of the total galaxy light. Most galaxy properties have radial gradients, and hence estimates of the total quantities are substantially uncertain (Brinchmann et al. 2004). Furthermore, the classification of galaxies as 'star-forming', 'AGN', 'composites', etc., are themselves uncertain.

Based on all of these considerations, we have chosen to use the median SFR/M_* and SFR inside the fiber as our marks, and we compare to the median total SFR. We increase our minimum redshift cut in order to minimize the uncertainties in the aperture corrections, although this decreases the sample size, and we exclude AGN and composites. These are our two volume-limited catalogs: $-23.5 < M_r < -19.5$, $0.05 < z < 0.082$, 32,725 galaxies; $-23.5 < M_r < -20.5$, $0.06 < z < 0.125$, 34,801 galaxies.

The Millennium simulation outputs the star formation rates and stellar masses of each galaxy directly, so we construct our $M_r < -19.5$ and $M_r < -20.5$ catalogs the same way as before, for both SFR and specific SFR marks.

The SFR distributions are broadly similar, but there are notable differences. The distributions for $M_r < -19.5$ are compared in Figures 6.14 and 6.15. The SDSS unnormalized and specific SFRs appear similar, except for the fact that the specific SFR distributions have two peaks and the difference between the mean and median is larger by a factor of about 1.5. The inside-the-fiber and total SFR are remarkably similar, and for this reason their significantly different marked correlation functions for $z_{\rm min} = 0.02$ are initially surprising. This merely shows that similar mark distributions is not a sufficient condition for similar marked clustering statistics; the uncertainties in the individual marks are important too.

It turns out that the total SFR distribution in the SDSS is similar to the SFR distributions in the Millennium Run simulation, as can be seen in the figure. However, there is a very marked difference between them: while the low-SFR tail of the SDSS distribution drops off rapidly at $10^{-2} M_\odot$/yr, the significant low-SFR population in the MS persists down to and including 'zero' SFR. In fact, almost *half* of the simulation's galaxies have SFRs below this, compared to $< 5\%$ in the SDSS. Of course, in the simulation, zero SFR just means "really small", that is, negligible mass contribution to the galaxy. The star formation model

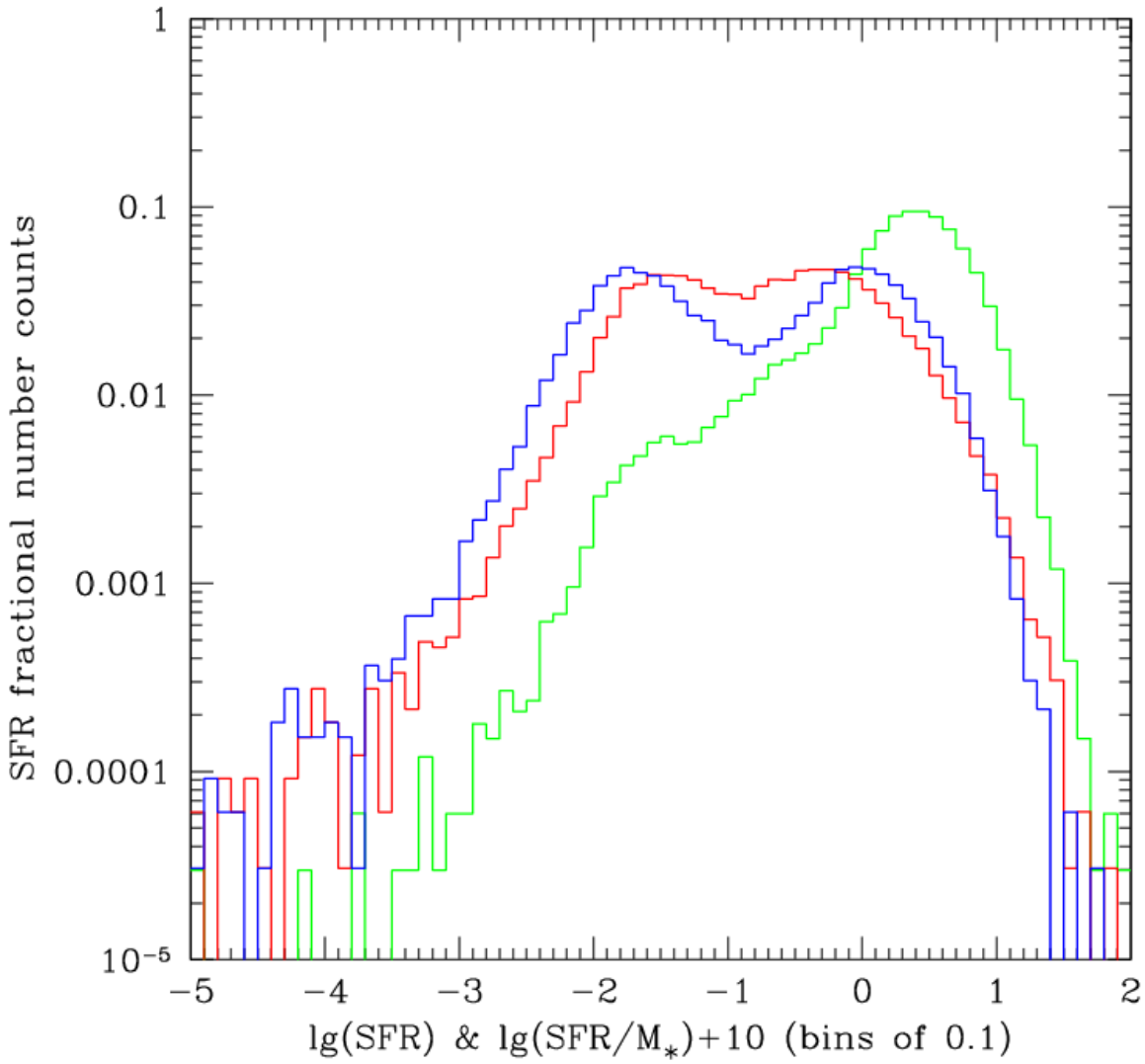

Figure 6.14: SDSS SFR distributions for $M_r < -19.5$. SFR (red) and SSFR (blue) inside the fiber, and total SFR (green).

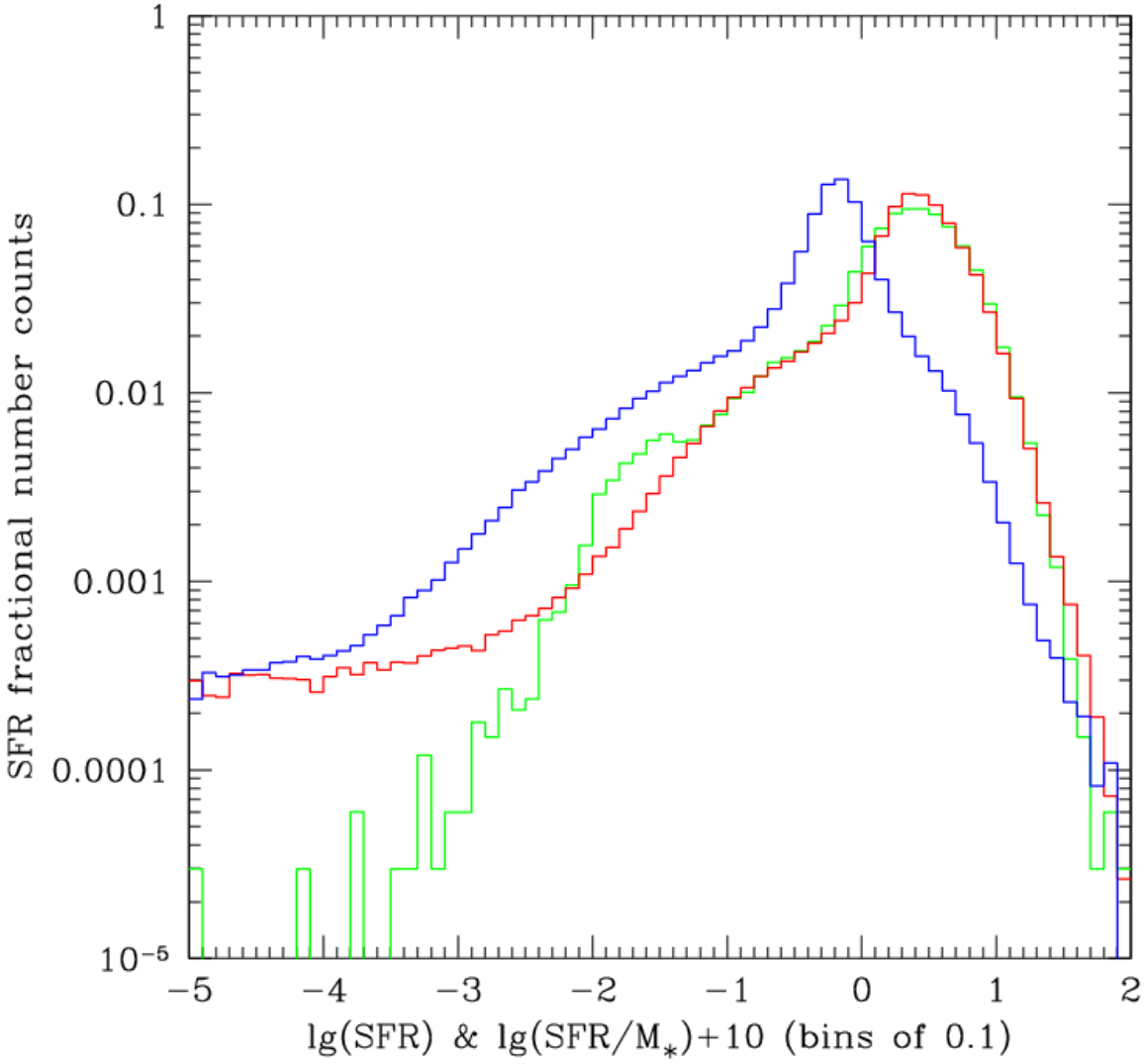

Figure 6.15: MS SFR distributions for $M_r < -19.5$. SFR (red) and SSFR (blue).

combined with the AGN model is not expected to resolve in detail very small specific SFRs (Darren Croton, private communication). Nonetheless this is very important for any analysis of the environmental dependence of SFR: although the simulation's distribution peaks similarly, the low and zero SFRs, most of which probably reside in denser environments, are extremely low relative to the mean SFR. Marked statistics are sensitive enough to be affected by this.

In Figures 6.16 and 6.17 we compare the marked projected correlation functions of unnormalized and specific SFR in the SDSS and MS for our pairs of catalogs $M_r < -19.5$ and $M_r < -20.5$. The SDSS error bars are jack-knife errors, and the MS error bars are estimates following the procedure outlined in the appendices in Sheth, Connolly & Skibba (2005). For $M_r < -20.5$, the SDSS SFR-marked uncertainties are similar to those of specific SFR, and so are left out, and the error bars are shown only up to 0.5 Mpc/h, for clarity; the error bars are similarly sized or larger at scales larger than this. The SDSS error bars are much larger than those of luminosity- and stellar mass-marked correlation functions of samples of similar size because of uncertainties in the SFR marks themselves.

As expected, the SFR-marked correlation functions are *much* below unity across a wide range of scales, indicating that passively star forming galaxies tend to reside in more dense regions than galaxies with active star formation. These are clearly very strong environmental trends, even stronger than the environmental signals we have observed with luminosity and stellar mass.

Firstly, it is interesting that the SFR- and SSFR-marked correlation functions are, on first appearance, fairly similar for both the real and simulated galaxies, in both plots. The environmental dependence of star formation rate in the Millennium Run simulation is generally not that different than what we observe, and this is encouraging. Our SDSS measurements are consistent, at least at large scales, with the star formation fraction-marked correlation functions of Sheth et al. (2006), who used star formation histories determined by the MOPED algorithm. Li et al. (2006) also observed similar trends in $w_p(r_p)$ as a function of $g-r$ color and $D_n(4000)$ for galaxy samples of different luminosity and stellar mass intervals.

Secondly, one can note that the total SFR-marked correlation function of the SDSS is

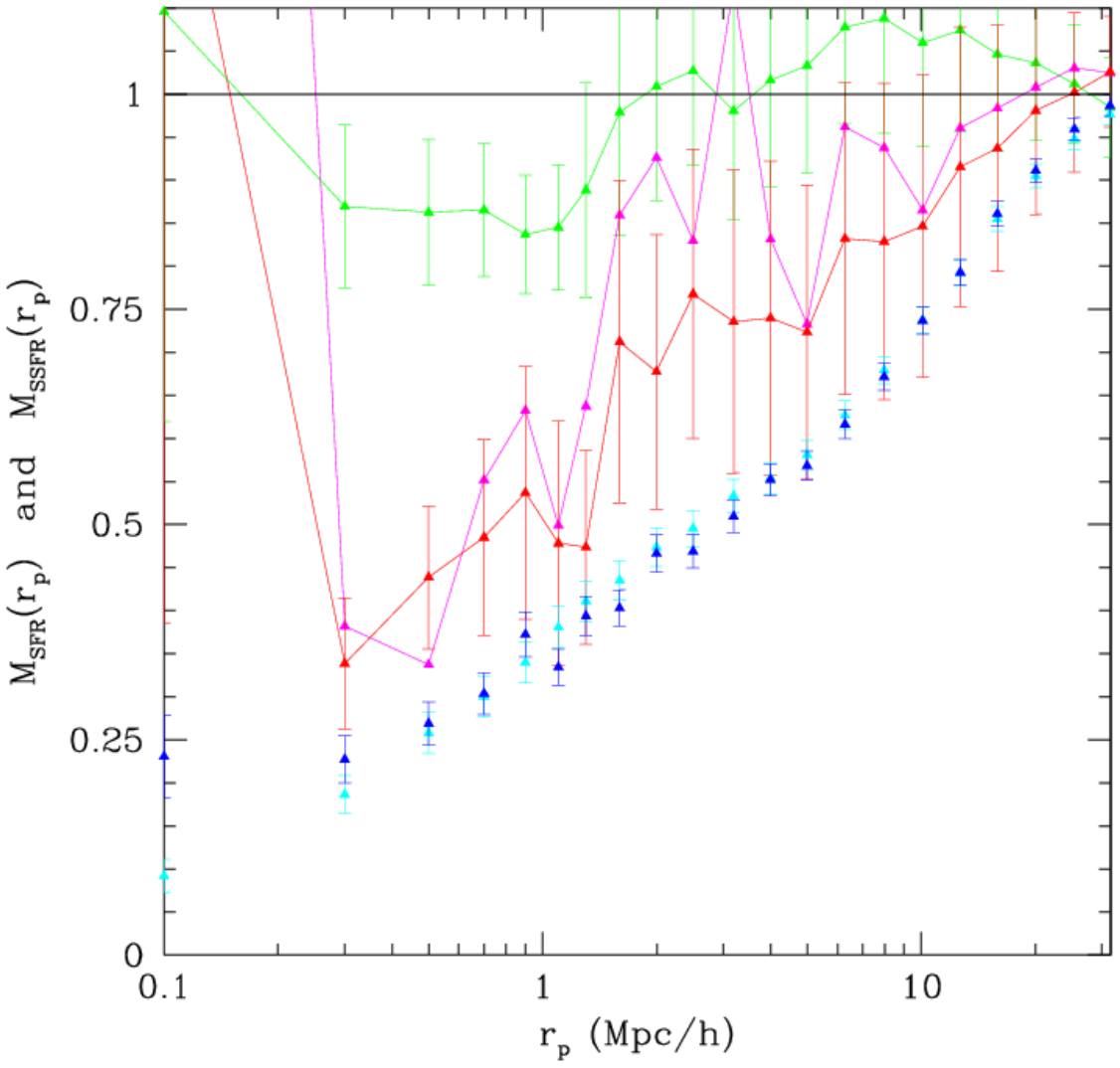

Figure 6.16: Star formation rate- and specific SFR-marked projected correlation functions for SDSS and MS $M_r < -19.5$ catalogs. SDSS measurements: SFR and SSFR inside the fiber (magenta and red, respectively) and total SFR (green). MS measurements: SFR (cyan) and SSFR (blue).

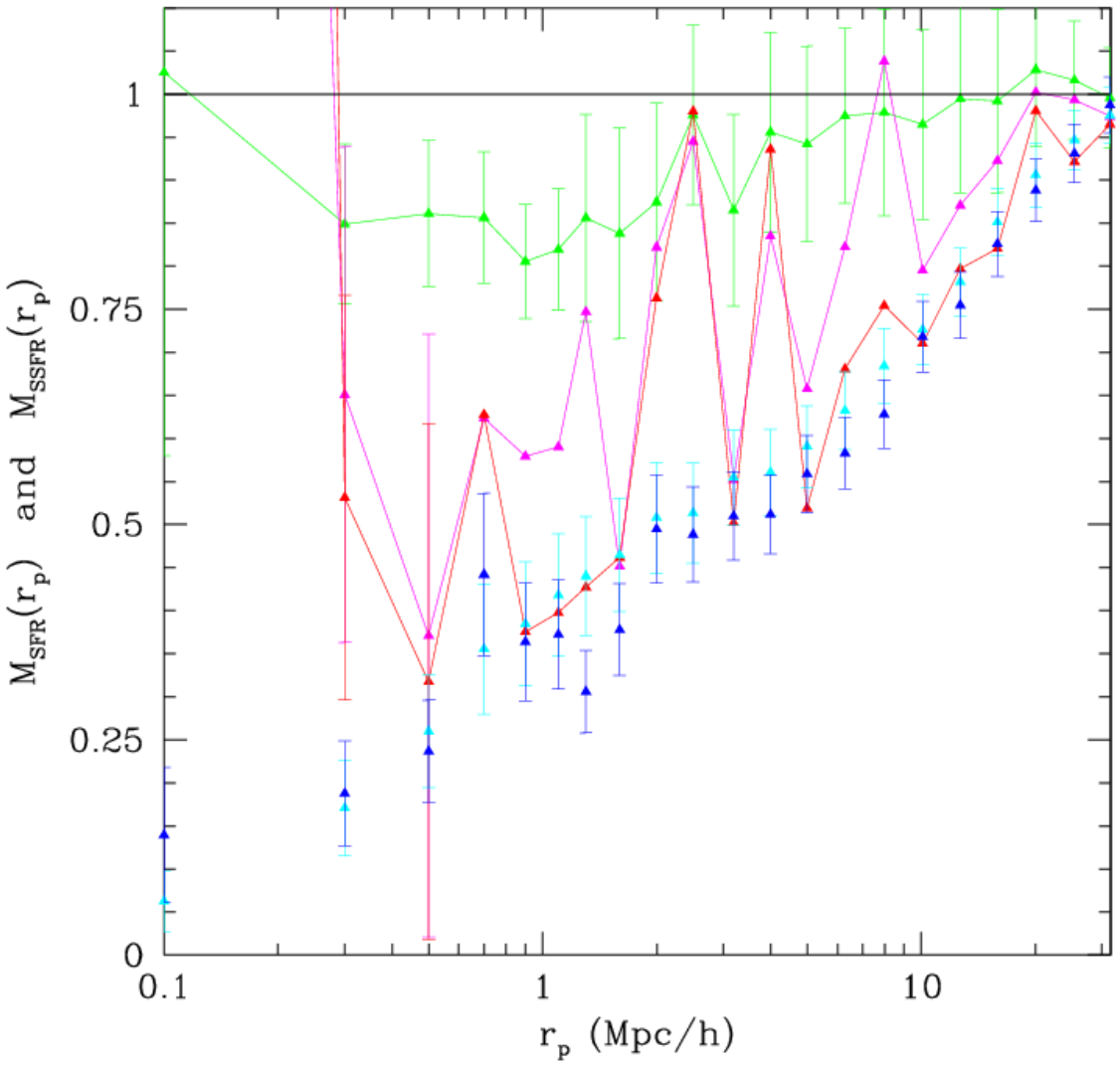

Figure 6.17: Star formation rate- and specific SFR-marked projected correlation functions for SDSS and MS $M_r < -20.5$ catalogs. SDSS measurements: SFR and SSFR inside the fiber (magenta and red, respectively) and total SFR (green). MS measurements: SFR (cyan) and SSFR (blue).

more weakly correlated with environment than both the SFRs inside the fiber and the SFRs in the simulation. This could merely be due to the uncertainties in the total SFRs, which are not reduced enough by our deeper minimum redshifts, or if it is physical, it would suggest that most of the environmental dependence of SFR is due to star formation activity in the central regions of galaxies.

Thirdly, the $M_{sfr}(r_p)$ for $M_r < -19.5$ and $M_r < -20.5$ are similar for $r_p > 400\,\mathrm{kpc}/h$. This is not altogether surprising, considering the distributions of star formation rate as a function of mass. In Figure 24 of Brinchmann et al. (2004), the galaxies' have a very narrow range of SFR/M_* for the stellar mass range of $10^9 M_\odot$ to $5 \times 10^{10} M_\odot$, and most of these galaxies have been forming their stars at a fairly constant rate. Feulner et al. (2006), using galaxy samples from FORS Deep Field, GOODS-S Field, and galaxy groups in Munich Near-Infrared Cluster Survey, have observed a similar distribution. Our luminosity thresholds of $M_r = -19.5$ and -20.5 correspond to minimum masses within this range, and since the specific SFRs of massive galaxies in both volume-limited ranges drop dramatically with mass at $M_* \geq 10^{11} M_\odot$, we should expect that both measurements of correlations with environment are somewhat similar. Kauffmann et al. (2006) attempt to explain the origin of the apparent universality of the average specific SFR within this range, and they argue that it is due to different gas consumption times: all of these galaxies in recent times have accreted approximately the same fraction of their mass in gas, resulting in similar bursty or smooth star formation histories.

Moreover, one should note that the correlation between star formation rate and environment, although it decreases with increasing projected separation, still persists out to large scales. Kauffmann et al. (2004) found larger values of $D_n(4000)$ in denser environments up to $r_p \sim 3\,\mathrm{Mpc}/h$, and we have found a statistically significant environmental dependence up to $10\,\mathrm{Mpc}/h$. In contrast, at such large scales luminosity and stellar mass do not exhibit such an environmental dependence. This suggests that the physical processes that cause the dramatic shutdown of star formation in massive galaxies in dense environments must have begun to operate fairly early on and are strongly related to the hierarchical mass accretion histories of the hosting dark matter halos. Menci et al. (2005) have argued as much: the observed partition of star formation histories is the final result of the interplay between the

biasing properties of the primordial density field from which the progenitors of local galaxies formed, and the dependence of feedback and star formation processes on the depth of the halo potential wells. This is of course related to arguments that have been made about the two major modes of halo and galaxy formation, by cold flows and shock-heating in low-mass and high-mass halos (Dekel & Birnboim 2006, Keres et al. 2005).

Fifthly, for $M_r < -19.5$ at least, there is a statistically significant difference between the SDSS and MS marked correlation functions ($\sim 1\sigma$) up to $r_p \approx 6\,\mathrm{Mpc}/h$. It is likely that with a larger sample of SDSS SFRs this discrepancy would not be resolved but would extend for $0.1\ \mathrm{Mpc}/h < r_p < 10\ \mathrm{Mpc}/h$. The discrepancy is almost certainly due to the large fraction of low SFRs in the Millennium simulation, which are probably the result of the extremely efficient AGN and supernova feedback models used in the simulation. Although both AGN and supernova are expected to suppress star formation to some extent, the star formation of faint and luminous simulated galaxies in overdense regions are apparently being almost entirely shut down, moreso than what we have observed in the SDSS.

Sixthly, although this was not our primary goal, it is very interesting that we have observed a sharp upturn at small scales in the SFR- and specific SFR-marked correlation functions in the SDSS. It is possible that this clustering result at such small scales is merely an artifact due to fiber collisions, but if not, then we have detected an *enhancement* of star formation at small scales. Star formation (and specific star formation) is enhanced even more in our brighter catalog: $M_{\mathrm{sfr}}(100\,\mathrm{kpc}/h) = 4.3 \pm 3.5$ and $M_{\mathrm{ssfr}}(100\,\mathrm{kpc}/h) = 1.4 \pm 1$ for $M_r < -19.5$, and $M_{\mathrm{sfr}}(100\,\mathrm{kpc}/h) = 6.6 \pm 3$ and $M_{\mathrm{ssfr}}(100\,\mathrm{kpc}/h) = 14.2 \pm 6.4$ for $M_r < -20.5$. Our results are consistent with those of Alonso et al. (2005), who proposed that this small-scale enhancement of star formation is due to galaxy interactions and mergers. They used SDSS DR2 specific SFRs, so their samples are smaller than ours, but they probed the environmental dependence of SFR with smaller small-scale bins. They also found an increase of star formation activity in pairs at small projected separations, but in high density regions galaxies have to be closer to show this enhancement. In low density environments, extremely star-forming galaxies tend to be in pairs, while in high density environments galaxies without a close companion (*i.e.*, recently accreted satellites) are more likely to contribute more SF to overall strongly star-forming systems.

Our results complement theirs by showing that the small-scale ($\sim$100 kpc/h) SF enhancement occurs in a variety of environments, while strongly star-forming galaxies are *much* more likely to reside in low-density regions when environment is measured at scales of 200 kpc/h or larger. These observations are consistent with the color and SFR dependence of the halo occupation distribution and satellite galaxy density profile. The slope of the satellite occupation distribution $\langle N_{\rm sat}|M\rangle$ increases with color (Zehavi et al. 2005) and hence decreases with SFR, and the density profile $\rho_{\rm gal}(r|M)$ is more concentrated for redder galaxies (Collister & Lahav 2005) with less star formation. Our picture of the situation then is this. In overdense environments halos tend to contain many satellites, tend to be more concentrated, and the central galaxy and most satellites tend to exhibit little star formation, compared to their counterparts in underdense regions. However, newly accreted satellites in the may interact with bright central galaxies or with their merging neighbors in the outskirts of the host halos, stimulating star formation in them. This occurs less often in the halos of bright central galaxies in low-density environments.

This observation is also seems consistent with Hogg et al. (2006), who find evidence of 'tidal-impulse triggering' of starburst events in galaxies, although they are more likely caused by galaxy-galaxy mergers. However, they also argue that most post-starburst galaxies are *not* in high-density environments. This would imply that, since most close pairs of star-forming galaxies in dense regions do not consist of post-starburst galaxies, the satellite galaxies of massive halos in these regions must have formed their stars over longer timescales and must continue to form stars subsequent to merging.

In any case, it is clear that there is not even a hint of these effects in the Millennium Run simulation. In conjunction with our results in the preceding section, these results suggest that at small scales in dense regions in the simulation, galaxies with large stellar masses are too highly clustered but these galaxies, whether they are centrals or satellites, are completely devoid of star formation. This is evidently not the case in the SDSS.

6.6 DISCUSSION

In this paper we showed how to describe luminosity-marked two-point correlation functions in the halo model (see Skibba et al. 2006), in which we determined the luminosity mark as a function of halo mass from the SDSS galaxy luminosity function, the halo mass function, and the halo occupation distribution. Measurements of clustering as a function of luminosity completely determine the halo-model description of the marked statistics. In our previous paper we compared luminosity-marked correlation functions to our measurements of volume-limited SDSS catalogs with $M_r < -19.5$ and $M_r < -20.5$. We have done the same with similar catalogs from the semi-analytical galaxy formation model of Croton et al. (2005), which was built upon the Millennium Run simulation of Springel et al. (2005).

We found previously that, when we account for the fact that central and satellite galaxies have distinctly different luminosities as a function of host halo mass, our halo-model luminosity-marked statistics can explain the SDSS clustering measurements, with good agreement down to small scales of a few hundred kpc. In this paper we performed a similar analysis with the simulated galaxies, although it was more difficult to do a direct comparison: our volume-limited catalogs of the Millennium Run had larger number densities than the corresponding catalogs of the SDSS, the simulation assumed a 2dF galaxy luminosity function while our luminosity mark and luminosity-dependent HOD are based on SDSS observations, and after accounting for both of these issues the minimum halo mass of the catalogs were still somewhat low compared to our expectations from the halo model. After correcting for these issues, which included rescaling the marks, we found that the luminosity-marked statistics of the simulation demonstrated much poorer agreement with the halo model than our SDSS measurements. We noted this discrepancy at multiple luminosity thresholds and the discrepancy was largest at small scales, $r \leq 1\,\mathrm{Mpc}/h$, where the 'one-halo term' of the two-point correlation function dominates.

Luminosities within halos are less clustered in the simulation than in our halo-model predictions. In other words, the environmental dependence of luminosity is too weak in the simulation, as opposed to the environmental dependence we expect given the correlation between halo abundance with large-scale environment. This discrepancy apparently increased

for fainter luminosity thresholds, even though our halo-model un-marked correlation function $\xi(r)$ was in excellent agreement with the simulation for thresholds of $M_r \sim -19.5$ and fainter, and even for our halo-model predictions that incorporated the halo occupation distribution ($\langle N_{\rm cen}|M\rangle$, $\langle N_{\rm sat}|M\rangle$, $\langle N_{\rm sat}(N_{\rm sat}-1)|M\rangle$) fitted to the simulated catalogs.

There are a number of reasons for this. Firstly, as Weinmann et al. (2006b) have noticed, the fraction of satellite galaxies is too large in the simulation, especially for the satellites of massive halos. Secondly, we have observed a 'kink' in the relation between central galaxy luminosity and host halo mass at $L_r \sim 2\times10^{10}L_\odot$ (or $M_r \sim -21$) and $M_{\rm host}$ between $2-4\times10^{12}M_\odot$, and forcing the luminosities to have a SDSS luminosity function does not correct for this. Thirdly, the scale dependence of satellite galaxy luminosities as a function of host halo mass is different than that of the SDSS or the halo model.

We also showed how to describe stellar mass-marked statistics in the halo model, using two independent stellar mass marks as a function of halo mass: obtained from the mass-to-light ratio $\langle M_*/L_r\rangle$ of Bell et al. (2003) combined with our luminosity mark, and the stellar mass-halo mass observed by Mandelbaum et al. (2006) from their weak lensing measurements. We used the SDSS stellar masses of Kauffmann et al. (2003) to construct two volume-limited catalogs for $M_r < -19.5$ and $M_r < -20.5$, and we measured the stellar mass-marked projected correlation function of them. We found excellent agreement between our halo model predictions and the measurements, and this is very encouraging. Our results suggest that the environmental dependence of stellar mass is driven by the correlation between halo mass and environment. In deriving the stellar mass marks we assumed no scatter in the mass-to-light ratio, so our agreement with the data suggests that the M/L scatter is not significantly correlated with the large-scale environment and is mostly due to small-scale processes.

We also calculated the halo model predictions of metallicity-marked statistics, and compared to measurements with SDSS volume-limited catalogs of galaxies with the O/H metallicities measured by Tremonti et al. (2004). We again found very good agreement between the halo model and the SDSS, although our small samples resulted in measurements with larger uncertainties. As with the stellar mass marks, we assumed no scatter in the relation between metallicity and mass, so our results again suggest that much of this scatter

is due to local processes. It will be very interesting to extend our halo-model analysis to color-marked correlation functions, and compare them similar SDSS measurements. Galaxy colors exhibit a strong bimodality, which affects how colors vary at fixed halo mass and the color-dependence of the halo occupation distribution, and we look forward to exploring how this bimodality is correlated with the large-scale environment.

We also compared our halo model predictions of stellar mass-marked correlation functions to measurements of our Millennium Run simulation catalogs. There was fairly good agreement with them, but *only when we rescaled the central and satellite galaxy marks*. Prior to rescaling, we measured the stellar masses of the simulated galaxies to be too strongly correlated with large-scale environment. This discrepancy is primarily due to the different stellar mass distributions of the Millennium Run and the SDSS. The discrepancy would *not* be resolved by using something other than a Salpeter initial mass function, which was assumed by Croton et al., because marked statistics are always normalized by the *mean* mark.

Finally, we compared measurements of star formation rate-marked correlation functions of the Millennium Run simulation and the SDSS, using volume-limited catalogs constructed from the SFRs of Brinchmann et al. (2004). There are many indicators of SFR, and we chose to focus on the estimated SFRs and specific SFRs of Brinchmann et al., while excluding AGN and composites, and increasing the minimum redshifts of our catalogs to minimize the uncertainties in the SFRs because of the aperture corrections. The measurements showed similar strong trends with environment, such that more passive SFRs are in denser regions and actively star-forming galaxies are in less dense regions. However, the simulated SFRs were much too strongly correlated with environment at small scales ($r_p < 400\,\mathrm{kpc}/h$) for $M_r < -20.5$ and at small and intermediate scales ($r_p < 10\,\mathrm{Mpc}/h$) for $M_r < -19.5$. The most striking discrepancy occurred in our smallest scale bins, $r_p < 200\,\mathrm{kpc}/h$, such that star formation was significantly enhanced in the SDSS, with the marked correlation function much greater than unity, while the environmental trend in the Millennium was not affected. The effect may be due to the merging of satellite galaxies and to interactions between them, but there is clearly no such small-scale star formation enhancement in the Millennium Run. As with the environmental dependence of stellar mass, the marked statistics, would *not* be significantly affected by modifying the IMF.

The similar environmental trends in the simulation, the SDSS, and in our halo model predictions is very encouraging, as it indicates that we are beginning to understand the major aspects of the environmental dependence of galaxy formation. Moreover, it is also encouraging that the modeling of feedback mechanisms in the simulation correctly and simultaneously predicts the flattening of the faint end of the luminosity function and the sharp cutoff at the bright end, which has been a longstanding issue in semi-analytic modeling (Benson et al. 2003). The inconsistent luminosity-marked correlation functions in the Millennium Run and halo model may turn out to be a relatively minor issue. Nonetheless, the extreme environmental dependence of stellar mass and star formation rate in the Croton et al. (2005) semi-analytic model may be an indication that their explanation of the 'downsizing' of star formation, which depends on their model of AGN feedback, needs to be improved or is incomplete. Their modeling of the suppression of gas cooling due to AGN feedback may be *too* efficient, resulting in too much shutdown of star formation in satellite galaxies. Indeed, our results are consistent with those of Weinmann et al. (2006b), who found that there are *much* too few blue satellites in the simulation compared to the SDSS. 'Down-sizing' is expected in hierarchical cold dark matter models (De Lucia et al. 2006, Cattaneo et al. 2006, Neistein et al. 2006), but it may simply be the case that the AGN feedback and supernova feedback models assumed in the Millennium Run are too efficient.

In order to shed more light on these issues, we look forward to modeling the environmental dependence of star formation rate in the halo model framework, and comparing the resulting SFR- and specific SFR-marked statistics to our SDSS observations and to the Millennium simulation. It has been shown that the star formation-density relation results from the prevalence of massive systems in high-density environments (*e.g.*, Mateus et al. 2006), and with an observationally-motivated model of the star formation-halo mass relation we can conduct such comparisons. It will also be very interesting to compare marked statistics with indicators of AGN activity, such as the [OIII]/Hβ and [NII]/Hα emission lines, as marks. These may be complemented with measurements of marked statistics in the simulations, with black hole mass or ejected gas mass as marks. We also look forward to performing similar analyses with data from the simulation of Bower et al. (2005).

6.7 ACKNOWLEDGEMENTS

We thank Darren Croton for some interesting discussions about the Millennium Run simulation data and for providing us with the center/satellite information of the galaxies and the host halo masses. We thank Jarle Brinchmann for help processing the SDSS DR4 stellar mass and SFR catalogs and for helpful discussions about them. We also thank Guinevere Kauffmann for helpful discussions about the SDSS stellar masses and how they compare to those of the Millennium Run. Finally, we thank Bob Nichol for interesting discussions about star formation rate-marked correlation functions in the SDSS, and we look forward to comparing our results with his in the future.

7.0 DEPENDENCE OF GALAXY CLUSTERING ON HALO FORMATION TIME AND ASSEMBLY HISTORY

7.1 INTRODUCTION

Hierarchical models of gravitational clustering and structure formation predict that dark matter halo formation should be determined by halo mass, and that halo formation should depend only weakly, if at all, on the large-scale environment. The excursion set model (Bond et al. 1991, Lacey & Cole 1993, Sheth, Mo & Tormen 2001) explains well the distributions of halo mass and formation times, as well as the abundance of halos as a function of mass and large scale environment. The dark matter halo model of halo and galaxy clustering (see Cooray & Sheth 2002) has correctly described the two-point correlation functions of galaxies (*e.g.*, Zehavi et al. 2005, Abbas & Sheth 2006, Skibba et al. 2006), thus apparently justifying the assumption that any correlation between halo formation time and environment or between halo substructure and environment is sufficiently weak to be negligible.

Lemson & Kauffmann (1999) seemingly confirmed as much by finding that halo formation times are independent of environment when averaging over the range of halo masses in a region. However, sensitive measurements of environmental correlations have shown that there is, in fact, a weak but significant correlation between formation time and the surrounding density field at fixed mass, with close pairs forming earlier than more distant pairs, independent of mass (Sheth & Tormen 2004). Gao et al. (2005) obtained a similar result, by measuring the two-point correlation function and halo bias factor as a function of formation time, and they found that the clustering of low-mass halos especially has a formation time dependence. Harker et al. (2006) and Wechsler et al. (2005), like Sheth & Tormen, employed 'marked' correlation functions with different formation time marks and confirmed

this result.

Additional unexpected correlations between halo attributes and environment have been observed. In contrast with Sheth & Tormen, a number of groups (Zentner et al. 2005, Avila-Reese et al. 2005, Wechsler et al. 2005, Wetzel et al. 2006) have measured an environmental dependence of halo concentration, especially for low-mass halos, and Wechsler et al. and Wetzel et al. argue that this dependence is *not* negligible to galaxy clustering. Related to this, Maulbetsch et al. 2006 and Wetzel et al. 2006 have found that the recent mass assembly and merger histories are correlated with environment, although the statistics are still rather poor and some of the effect is due to local processes. Furthermore, Zentner et al. (2005), Wechsler et al. (2005), and Zhu et al. (2006) have detected correlations between the halo occupation distribution and formation time, with earlier forming halos hosting fewer satellite galaxies.

The picture if further complicated by the observation that the effect of assembly bias on the clustering of simulated galaxies varies as a function of luminosity and color (Croton et al. 2006). If the effect is strong enough, analyses of the environmental dependence of luminosity and color of real galaxies (*e.g.*, Zehavi et al. 2005, Skibba et al. 2006) may need to be re-assessed. Hierarchical clustering models are consistent with the picture that the present day bimodal color distribution is the result of a partition in star formation histories that is driven by the divergent merging histories and mass accretion histories of the progenitors (Menci et al. 2005, Maulbetsch et al. 2006). If the color dependence of galaxy clustering were actually due to the clustering dependence on halo *age* (rather than halo mass), this would pose a problem for hierarchical clustering models. However, Reed et al. (2006) argue that the clustering age dependence is so much weaker that this is likely not the case.

In order to shed some light on these issues, we perform a simple test of assembly bias in this chapter. We take a galaxy sample with a given luminosity cut from the Millennium Run Simulation (Croton et al. 2005) and we measure its two-point correlation function. Then we fit the corresponding minimum halo mass and the first and second moments of the halo occupation distribution, and we use these to determine the halo model prediction of the correlation function. If the halo model and the Millennium Run are not consistent within the uncertainties then this is likely evidence of assembly bias. If they are consistent, then

assembly bias may still be important but is weak enough to be neglected for such types of clustering analyses.

This chapter is organized as follows. In the next section we lay out the details of our assembly bias test. We describe and show our fits to the minimum halo mass and the halo occupation distribution in Section 7.3. We show our result in Section 7.4, and discuss it in the final section.

7.2 REQUIREMENTS FOR A DIRECT TEST

For the halo model prediction of the two-point correlation function, we calculate it in the same manner as described in Chapter 6. The only difference–and it is an important difference–is that we *fit* the minimum halo mass (from the distribution of central galaxies) and the satellite galaxy halo occupation distribution, rather than basing them on other observations of the luminosity function and HOD. The halo model components required for the correlation function calculations are the mass function and bias factor (Sheth & Tormen 1999) and the density profile (Navarro, Frenk & White 1997). For consistency, we assume the same cosmological parameters ($\Omega_M = 0.25$, $\Omega_\Lambda = 0.75$, and $h = 1$), linear power spectrum (CMBFAST), and virialization density ($\Delta_{\rm vir} = 200$).

For Millennium Run galaxy catalog, we need to ensure that the number density is consistent with the SDSS number density for that luminosity threshold. In general, the Millennium has larger number densities than the SDSS. One way to correct for this is to rank-order the galaxies by their luminosities and then exclude the required number of faintest galaxies so that the catalog has the same number density as its SDSS counterpart. As in Chapter 6, this is our approach here. We choose a relatively faint luminosity threshold, $M_r < -19.5$, for our assembly bias test, for which this correction is not very important.

Our resulting $M_r < -19.5$ galaxy catalog consists of 1,268,751 galaxies, and we measure its halo occupation distribution and minimum mass in the next section.

7.3 FITTING THE HALO OCCUPATION DISTRIBUTION

The halo occupation distribution is the probability distribution of halos of mass M hosting N galaxies. We assume that each halo contains exactly one central galaxy and some number of satellite galaxies:

$$\langle N_{\rm gal}|M, L_{\rm min}\rangle \;=\; \langle N_{\rm cen}|M, L_{\rm min}\rangle \;+\; \langle N_{\rm sat}|M, L_{\rm min}\rangle\,. \tag{7.1}$$

In practice, for a given luminosity threshold there is scatter among the minimum halo masses, and so we fit an error function, rather than a step function, to the distribution of central galaxies as a function of mass:

$$\langle N_{\rm cen}\rangle \;=\; \frac{1}{2}\,{\rm erfc}\left(\frac{\log_{10}(M_{\rm min}(L_{\rm min})/M)}{\sqrt{2}\,\sigma_{\rm cen}}\right) \tag{7.2}$$

There are two free parameters we fit to in $N_{\rm cen}$: the minimum mass of halos hosting galaxies brighter than the luminosity threshold, $M_{\rm min}$ and the scatter $\sigma_{\rm cen}$.

The mean value of the satellite distribution increases with halo mass like a power law, and we allow for the possibility that it drops below a power law at low mass:

$$\langle N_{\rm sat}\rangle \;=\; \exp\left(-\beta\,\frac{M_{\rm min}(L_{\rm min})}{M}\right)\left(\frac{M}{\mu\,M_{\rm min}(L_{\rm min})}\right)^{\alpha} \tag{7.3}$$

Given the minimum mass from the central galaxy distribution, we fit for three free parameters in $N_{\rm sat}$: the mass required for a single satellite μ, the slope of the power law α, and the drop below the power law β. Since it is not guaranteed that the satellite galaxies will have a Poisson distribution, we also fit the same equation (7.3) to the square root of the second moment of the HOD, $\langle N_{\rm sat}(N_{\rm sat}-1)|M\rangle^{1/2}$.

We measured the mean and variance of $N_{\rm cen}$ and $N_{\rm sat}$ of the catalog as a function of host halo mass. In order to do this, we needed to know which galaxies are centrals and which are satellites, and we needed the halo masses, both of which were kindly provided to us by Darren Croton. We used a nonlinear least-squares fitting method, the Levenberg-Marquardt minimization algorithm, which uses the function as well as its derivatives as a function of the parameters to be fitted, and which accounts for the scatter in each bin. We fit the HOD

in logspace because of the wide range of masses and because otherwise the algorithm does not fit to the large-mass points (whose absolute uncertainties are large).

Our fit to the central galaxy distribution is shown in Figure 7.1. For illustrative purposes only, we also showed the SDSS result of Zehavi et al. (2005) (green line); one should not expect they be the same. From our relation between halo mass and galaxy luminosity (see Chapter 3), a luminosity threshold of $M_r < -19.5$ corresponds to a minimum halo mass of $\log_{10} M_{\rm min}(M_\odot) = 11.70 \pm 0.04$ if one assumes the Blanton et al. (2003a) luminosity function, and to $\log_{10} M_{\rm min}(M_\odot) = 11.54 \pm 0.06$ if one assumes the Blanton et al. (2001) LF. The former is only slightly lower than that of Zehavi et al., but we expected the result to be closer to the latter, as the older Blanton et al. luminosity function is more '2dF-like', with a brighter M_* and a steeper faint-end slope. Although the Millennium Run outputs SDSS *ugriz*-band luminosities, it assumed 2dF luminosity functions. Our fitted minimum mass is fairly low: $\log_{10} M_{\rm min}(M_\odot) = 11.485 \pm 0.092$, and the width of the error function is $\sigma_{\rm cen} = 0.13 \pm 0.06$.

Our fit to the satellite galaxy halo occupation distribution is shown in Figure 7.2. For illustrative purposes, we also show the SDSS result of Zehavi et al. (2005). Our results for the halo mass required to host a single satellite and for the power-law slope were both lower than our expectations: $\mu = 16.0 \pm 3.7$ and $\alpha = 0.98 \pm 0.05$. For the second moment of the HOD, the corresponding parameters are: $\mu' = 15.5 \pm 4.7$ and $\alpha' = 0.98 \pm 0.06$. It turns out that the first and second moments are slightly different at low masses, with exponential drop-offs characterized by $\beta = 4.1 \pm 5.0$ and $\beta' = 3.1 \pm 11$, respectively. There is a *lot* of scatter at such low masses (at $M < M_{\rm min}$), but the Millennium satellites do appear to deviate slightly from a Poisson distribution, although the difference is not statistically significant.

We use these HOD fits as input for the halo model, and compare the resultant two-point correlation function to our measurement with the Millennium Run catalog. It turns out that the difference between the first and second moments of the HOD affects the correlation function by as much as 4% at small scales, $200\,{\rm kpc}/h < r < 600\,{\rm kpc}/h$, where the satellite-satellite one-halo term is important.

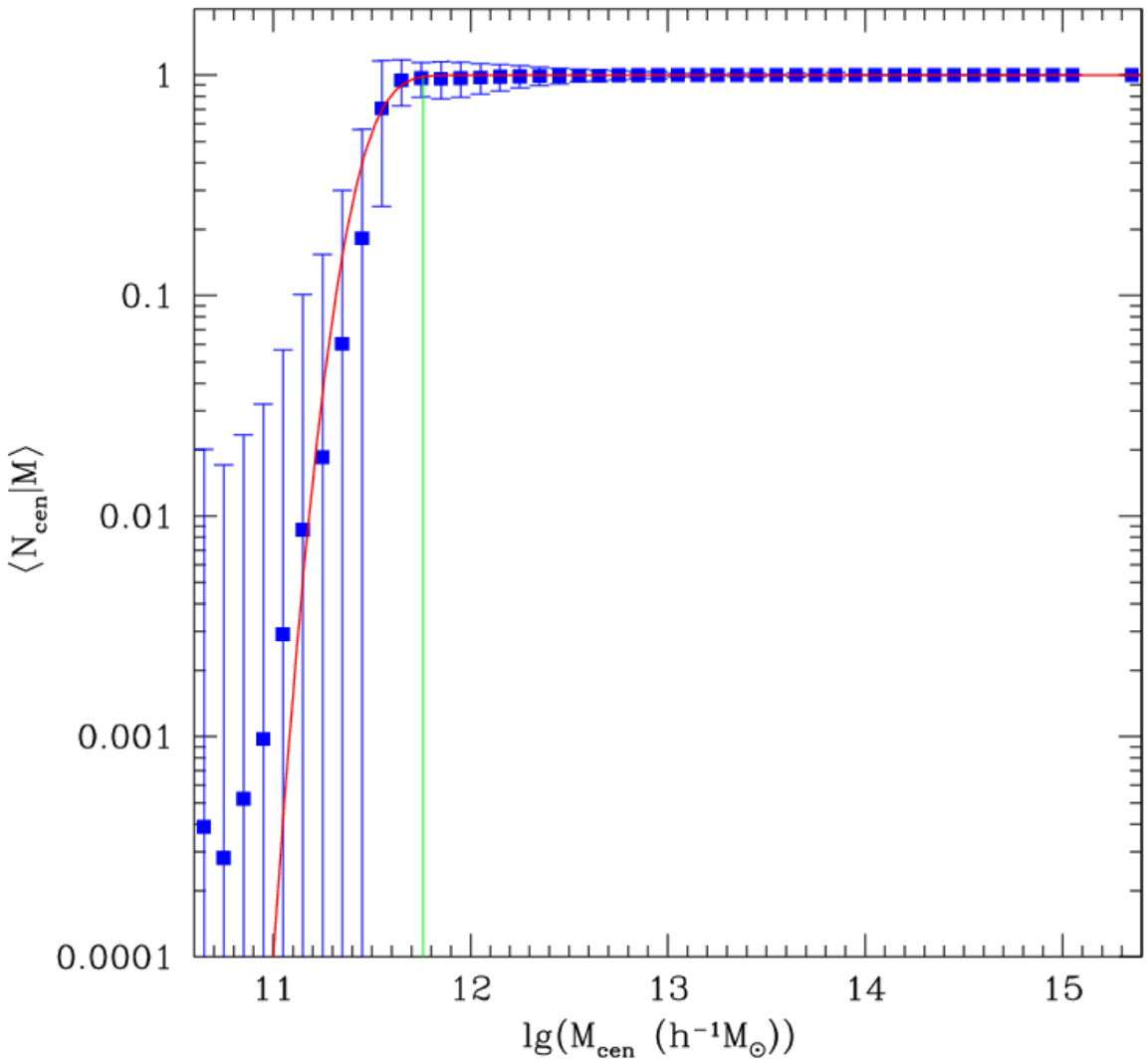

Figure 7.1: Millennium Run central galaxy distribution as a function of mass for $M_r < -19.5$. Our fit (red curve) is also compared to the SDSS result (green, Zehavi et al. 2005).

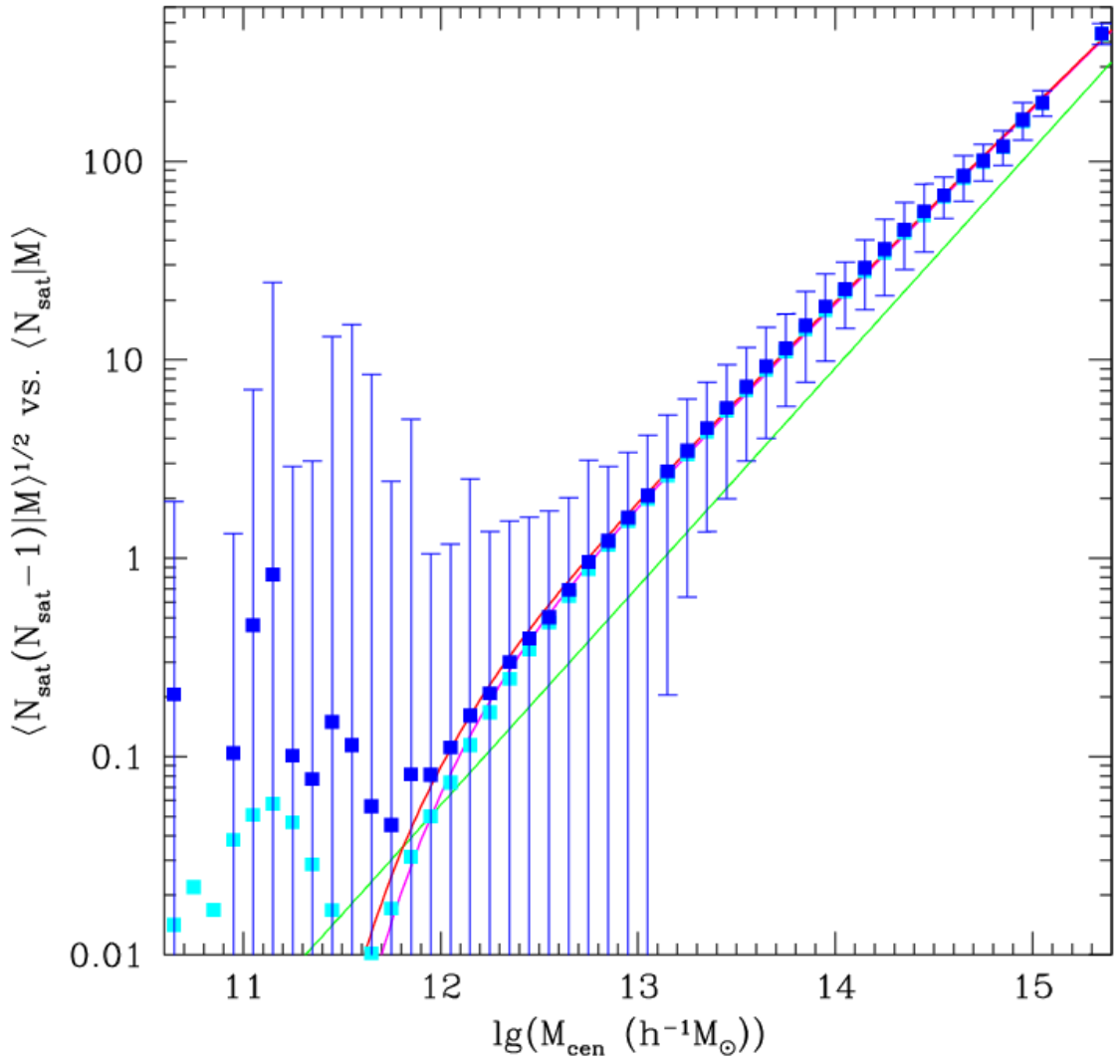

Figure 7.2: First moment (cyan) and second moment (blue) of satellite galaxy halo occupation distribution of Millennium Run for $M_r < -19.5$. Our fits are the magenta and red curves, respectively. The SDSS result (green, Zehavi et al. 2005) is shown for comparison.

7.4 HALO MODEL *VS.* MILLENNIUM RUN CORRELATION FUNCTION

Our halo model prediction of the two-point correlation function for $M_r < -19.5$, with the fitted halo occupation distribution, and our measurement of the Millennium Run catalog are shown in the top panel of Figure 7.3. The error bars on the measured points are estimated according to the procedure described in the appendix of Sheth, Connolly & Skibba (2005); they are consistent with jack-knife errors on scales less than a few Mpc and they tend to underestimate the errors at large scales. In any case, the enormous size of the simulated catalog makes for extremely small uncertainties.

The halo-model uncertainties are estimated according to a method proposed by Ryan Scranton. We sampled from the distributions of the fitted HOD parameters, assuming them to be Gaussian, and repeated the calculation 100 times. The error bars are the square root of the variance at each point. In order to avoid overestimating the uncertainties, and in order to account for the correlated HOD parameters, we used the randomly chosen $\log_{10} M_{\rm min}$ to re-fit the satellite HOD each time, then sampled from the distributions of those parameters. We then assumed that the second moment of the HOD is related to the first moment as usual, such that the exponential drop-off is always $\beta' \approx 3/4\beta$, for example. Finally, we required that $\beta \geq 0$, and of course, $\sigma_{\rm cen} \geq 0$.

Because $\xi(r)$ is shown in log-space, in the bottom panel we showed the ratio of the halo-model correlation function to the measured one (to which we fitted a spline). The largest discrepancy between the halo model and the Millennium simulation occurs at $r \sim 500\,{\rm kpc}/h$. If assembly bias is strong enough to cause a discrepancy, it would probably occur on such scales (Wechsler et al. 2005, Croton et al. 2006). However, the halo-model correlation function is most uncertain at these scales, as the uncertainties in the HOD parameters matter most in the satellite-satellite one-halo term. Nonetheless, the correlation functions are discrepant by as much as $\sim 20\%$, and this could be evidence of assembly bias. Ignoring assembly bias may result in slightly overestimating small-scale correlations.

There is also a large-scale discrepancy at $r \sim 10\,{\rm Mpc}/h$ (and one at $r \sim 30\,{\rm Mpc}/h$, but this is likely due to the largest-scale measured point). This initially seems like a cause for concern, but it may be due to our assumed bias factor (Sheth & Tormen 1999). Large-scale

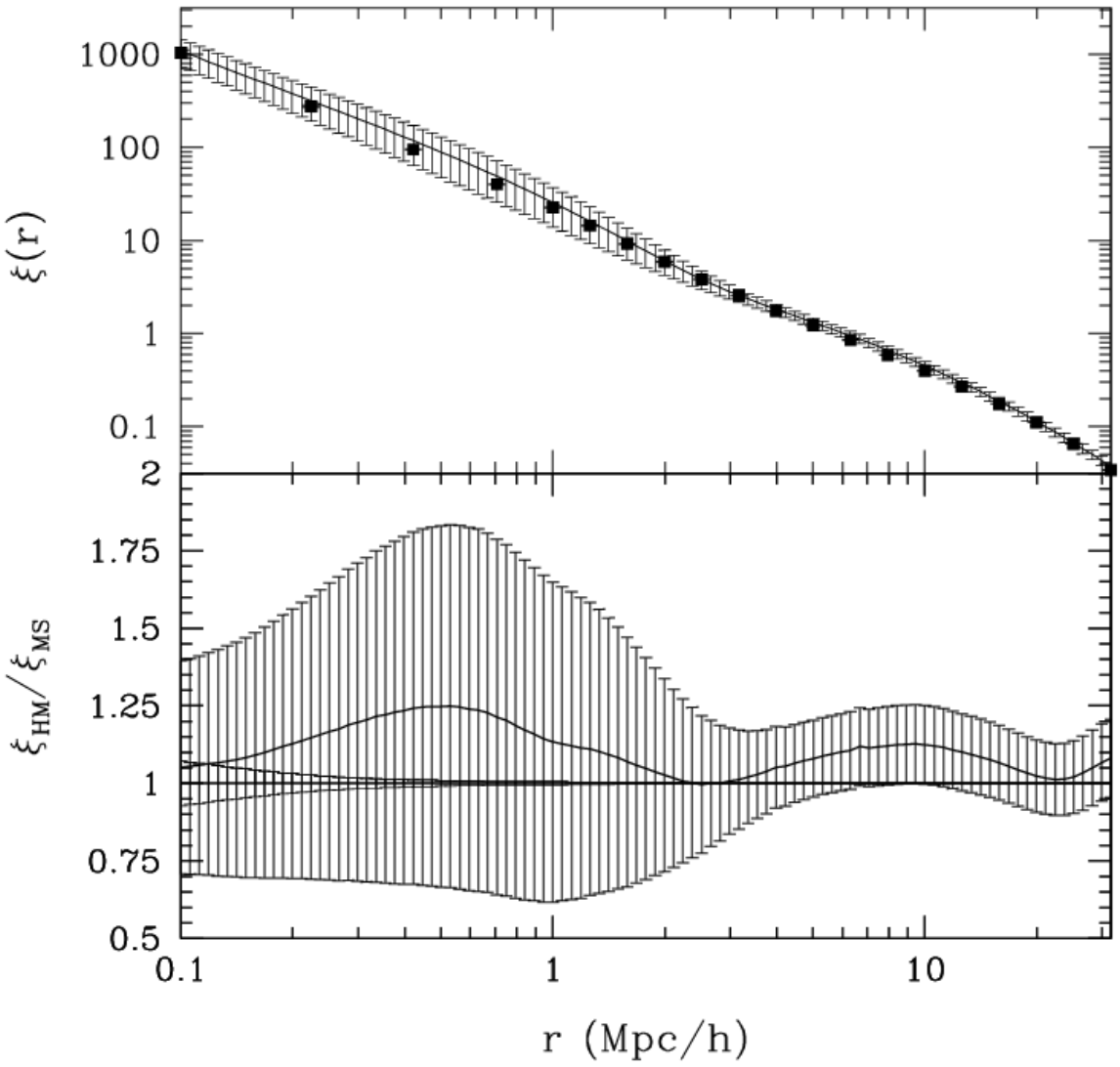

Figure 7.3: Halo-model correlation function for $M_r < -19.5$ compared to Millennium simulation measurement. Halo-model uncertainties are due to scatter in measured halo occupation distribution, Figures 7.1 and 7.2.

bias is scale-dependent (Sheth & Lemson 1999), and ours may overestimate it at these scales (see appendix of Tinker et al. 2005). Moreover, Seljak & Warren (2004) argue that Sheth & Tormen overpredict the bias factor at $M < M_*$, and the fitted minimum halo mass for $M_r < -19.5$ is $\approx 0.04\,M_*$. If we assume their bias as a function i of mass, our halo-model correlation function decreases by more than 10% at $10\,\mathrm{Mpc}/h$. However, Croton et al. (2006) showed that the effects of assembly bias may be important on large scales. It is possible that the discrepancy we see at $\sim 10\,\mathrm{Mpc}$ is not due to the bias factor but is actually the same one that Croton et al. found, and it is possible that it would have been worse had it not been for the discrepancy they found.

7.5 DISCUSSION

Many astrophysicists have found evidence for 'assembly bias' over the past few years, claiming that galaxy clustering is slightly affected by correlations between large-scale environment, halo formation time, halo concentration, and possibly the halo occupation distribution. The purpose of this paper is to perform a simple test of assembly bias: we took a galaxy sample with a particular luminosity cut ($M_r < -19.5$) from the Millennium Run Simulation and measured its two-point correlation function. We compared this to the halo-model prediction, using the halo occupation distribution fitted to the catalog.

Our results suggest that assembly bias is too weak to significantly affect two-point galaxy clustering. Assembly bias may have a $5-10\%$ effect, but the discrepancy we found between the halo model and the measurement was much smaller in the uncertainty due to the scatter in the HOD. The galaxies with the range of luminosities in our test correspond to halos with a very wide range of masses, and this range may have been sufficiently large to make assembly bias negligibly small.

Nonetheless, for galaxy samples defined by small ranges of luminosities, especially faint luminosities, the effect may not be negligible. For example, galaxies in samples with $-19 < M_r < -18$ are hosted by halos with masses of a few $\times 10^{11} M_\odot$, which are halo masses for which halo bias is relatively strong. However, galaxy samples over such ranges are still rela-

tively small, making it difficult to constrain models with statistical clustering measurements of them.

Interestingly, unlike $\xi(r)$, Figure 7.4 shows that the Millennium luminosity-*marked* correlation function is *not* entirely consistent with the halo model, even within the variance due to the scatter in the HOD parameters. The halo-model calculation was done with the same fitted $N_{\rm cen}$ and $N_{\rm sat}$, with the additional assumption of the luminosity function (Blanton et al. 2003a) and luminosity-dependent HOD, with an approximately constant value of $M_1/M_{\rm min}$ and a power-law slope of the mean satellite number that slightly increases with luminosity, according to our Millennium $\xi(r)$ measurements at various luminosity thresholds. The Millennium simulation exhibits too weak of a correlation between luminosity and environment at small scales ($r < 1\,{\rm Mpc}/h$). If one assumes a '2dF-like' luminosity function similar to that of Blanton et al. (2001), with a somewhat brighter L_* and slightly higher $\phi_a st$, the minimum luminosity corresponds to the fitted minimum mass (Section 7.3). Such an assumption results in a weaker halo-model prediction of luminosity-marked correlations, but for consistency the luminosities of the Millennium galaxies are similarly rescaled, and now we see a discrepancy at $r \sim 1\,{\rm Mpc}/h$. In short, the small-scale environmental dependence of luminosity in the Millennium Run is not consistent with the halo model or the SDSS (see Chapter 6), and this is not due to assembly bias.

A clearer comparison of marked correlation functions can be done, for the case of the luminosity mark, by measuring the luminosity function of the simulation directly and by measuring the luminosity dependence of the halo occupation distribution parameters for a range of luminosity thresholds (*i.e.*, the dependence of $\langle N_{\rm cen}|M\rangle$, $\langle N_{\rm sat}|M\rangle$, and $\langle N_{\rm sat}(N_{\rm sat}-1)|M\rangle$ on $M_{\rm min}(L_{\rm min})$). Then these measurements can be applied to the halo model calculation, rather than assuming SDSS or 2dF-like observations of the luminosity function and luminosity-dependent HOD. The halo-model prediction of the luminosity-marked correlation function can then be compared to the Millennium simulation measurement, without any need to rescale the marks. We will perform such an analysis in the very near future.

Although most current galaxy clustering analyses may justifiably ignore effects of assembly bias, datasets of galaxies with very narrow ranges of masses, luminosities or colors, and upcoming larger datasets (especially with the LSST), will require theoretical models that do

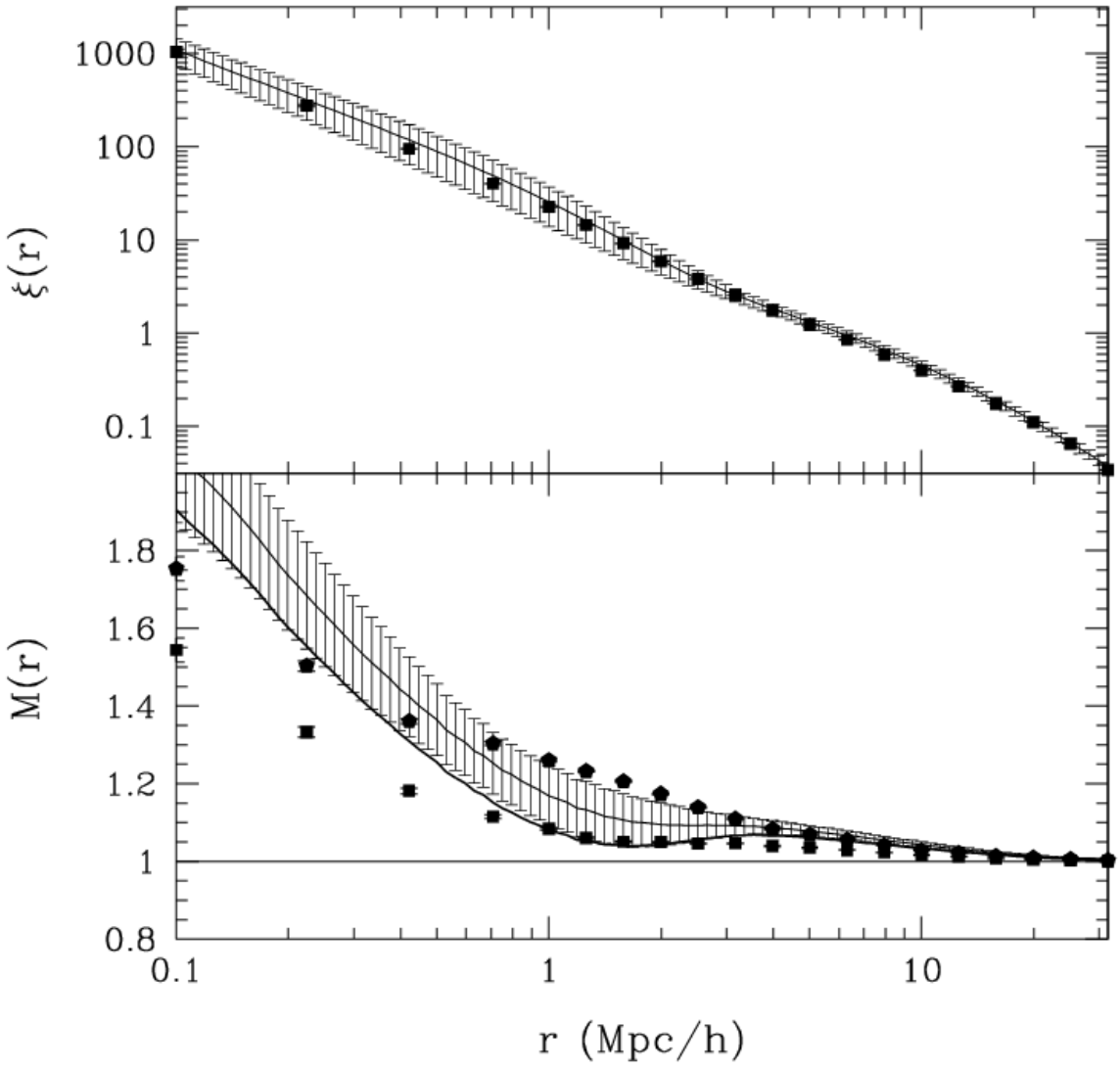

Figure 7.4: r-band luminosity-marked correlation function for $M_r < -19.5$ (lower panel). Squares show Millennium simulation measurement with rescaled central and satellite luminosities (see Chapter 6), and these are to be compared to the solid curve with error bars, the halo-model prediction with fitted HOD and Blanton et al. 2003a LF. Thick solid curve without error bars is halo-model prediction assuming LF similar to Blanton et al. 2001, to be compared to measurement with luminosities rescaled assuming same LF (pentagons).

account for assembly bias. Perhaps, as suggested by Wechsler et al. (2005), the halo occupation distribution and density profile applied to satellite galaxies can be modified to account for some of the correlations between formation time, concentration, $N_{\rm sat}$, and environment. Croton et al. (2006) argue that the environmental dependences of halo formation cannot easily be included in HOD models, and in any case, our attempts to model assembly bias effects will be instructive only insofar as we can constrain them with observations.

8.0 LUMINOSITY-MARKED PROJECTED CORRELATIONS OF SDSS GALAXY GROUPS & CLUSTERS

8.1 INTRODUCTION

Galaxies live in a variety of environments and they form and evolve in a variety of ways. In hierarchical models of gravitational clustering and structure formation, dark matter that originated around peaks in the primordial density field eventually collapse into dense 'halos', within which galaxies reside. In this paradigm, one naturally expects that the environmental dependence of galaxy formation and evolution is largely determined by correlations between the mass and formation histories of halos and their large-scale environment. Based on marked clustering analyses of samples of local galaxies, we have shown that the environmental dependence of particular galaxy attributes, namely, luminosity, color, stellar mass and metallicity, indeed appears to be driven by the environmental dependence of halo mass (Chapters 5 and 6).

Galaxy groups and clusters are in particularly dense environments, and they consist of galaxies that tend to have brighter luminosities, redder colors, lower star formation rates, higher stellar masses, and more elliptical morphologies, than galaxies in underdense regions. In order to better understand galaxy formation and evolution, it is important to address the question as to what extent the distributions of galaxy properties depends on halo formation, and to what extent they are established by the myriad of *group-specific* processes (*i.e.*, processes in the local environment), such as ram-pressure stripping, galaxy 'strangulation', and galaxy-galaxy interactions and 'harassment' (described in Weinmann et al. 2006a).

In the dark matter halo model (see Cooray & Sheth 2002 for a review), massive halos host luminous galaxies at their center, and previously accreted less massive subhalos contain

'satellite' galaxies, which tend to be fainter and bluer than the 'central' galaxy. Although some satellite galaxies merge with the central galaxy, massive halos in dense regions often accrete new subhalos, so that the largest halos may contain dozens, even hundreds, of galaxies. The goal of this paper is to study the clustering of the properties of galaxies that reside in such groups, in order to explore the environmental dependence of those properties.

We do so with marked projected correlation functions in the halo model framework (see Chapter 4). We will first focus on galaxy luminosity as our mark, as it is very strongly correlated with halo mass (Chapter 3), and then we will use a color mark. In order to calculate the correlation between a particular galaxy mark and environment in the halo model, the relation between that mark and halo mass is required, as well as the dependence of the halo occupation distribution on that mark. Once these are modeled in an observationally-motivated way, one can compare the marked correlation functions predicted by the halo model to measurements of galaxy catalogs, and thus gain information about the environmental dependence of the galaxy mark and the extent to which it is driven by halo mass and halo formation.

We use two of the Sloan Digital Sky Survey (SDSS) group catalogs identified by Berlind et al. (2006) for our measurements. The galaxy groups and clusters in these catalogs were identified using a redshift-space friends-of-friends algorithm, in which the linking lengths were chosen so that the identified groups inhabit the same dark matter halos, with limited contamination by galaxies in different halos. This makes these catalogs ideal for marked statistics analyses with the halo model.

The research in this chapter is still a work-in-progress. The chapter is organized as follows. In the next section we show some properties of the galaxies in the groups and compare them to the halo-model predictions. In Section 8.3, we show the luminosity-marked projected correlation functions of galaxy groups, comparing our halo-model predictions with measurements of the SDSS catalogs and our SDSS-based mock catalogs. We note a discrepancy with the data at small scales, which we attribute to group-specific processes affecting the positions of the central and satellite galaxies, and we explore them in Section 8.4. Since galaxies in groups compose a minority of all galaxies, these group-specific processes have not been significant in previous clustering analyses of datasets containing galaxies in a wider range of

environments (*e.g.*, Zehavi et al. 2005, Skibba et al. 2006, Li et al. 2006, Chapter 6). In the final section, we summarize the current state of our results, and our goals for continuing our investigation into the competing roles of small-scale and large-scale environmental processes in galaxy formation.

8.2 PROPERTIES OF GALAXY GROUPS: SDSS *VS.* HALO MODEL

In this section, we compare the multiplicity function, central galaxy luminosities, and satellite galaxy luminosities of the group catalogs and the halo model. Each of these are calculated as a function of richness N, the number of galaxies per group. Based on clustering constraints on the halo occupation distribution, the richness scales like a power-law with halo mass at large N.

We use the volume-limited $M_r < -19$ ($0.015 \leq z \leq 0.068$) and $M_r < -19.9$ ($0.015 \leq z \leq 0.100$) SDSS group member catalogs described in Berlind et al. (2006). The catalogs contain 15,413 and 21,301 galaxies, respectively. *Only the galaxies in groups of three or members are included*, and they constitute 40.7% and 37.2% of the total number of galaxies in the volume-limited ranges. Isolated galaxies and galaxies in pairs are excluded.

In order to calculate the group multiplicity function, central and satellite galaxy luminosities, and marked and un-marked galaxy correlation functions in the halo model, an observationally-motivated model of the halo occupation distribution is required. One also requires the minimum halo mass corresponding to the luminosity threshold, but it is almost completely determined by the luminosity function (Blanton et al. 2003a; *cf.* Zandivarez, Martínez & Merchán 2006), following the approach described in Chapter 3. The minimum masses we obtained for the two catalogs are approximately $\log_{10} M_{\rm min}(M_\odot) = 11.86$ and 12.22, respectively. Based on our initial correlation function analysis, we determined approximately what halo occupation distribution parameterization results in a good fit to the large-scale projected correlation function, $w_p(r_p)$. We assumed the following model of the

HOD:

$$\langle N_{\rm gal}|M, M_{\rm min}\rangle \;=\; \langle N_{\rm cen}|M, M_{\rm min}\rangle \;+\; \langle N_{\rm sat}|M, M_{\rm min}\rangle \;\approx\; 1 \;+\; \left(\frac{M}{\mu\, M_{\rm min}(L_{\rm min})}\right)^{\alpha}. \qquad (8.1)$$

Our fiducial HOD consists of: $\mu \equiv M_1/M_{\rm min} = 8$ and $\alpha = 1.12$ for $M_r < -19$, $\mu = 8$ and $\alpha = 1.19$ for $M_r < -19.9$. No parameterization fit well with the small-scale correlation function, as discussed in the next section, except for an extreme HOD with $\mu = 4$, but this meant *many* satellite galaxies in halos, especially massive halos, resulting in a *much* stronger luminosity-marked correlation function than measured. As we show below, the different HODs also result in different multiplicity functions and mean central and satellite galaxy luminosities.

The multiplicity function is the space density of groups as a function of richness. For the halo model, it is completely determined by the abundance of halos and the halo occupation distribution, and hence can be used to constrain HOD parameters. It is important to compare the halo-model group multiplicity function to that of the SDSS catalogs because one of the main requirements of the group-finding algorithm was that it produce a multiplicity function that is unbiased with respect to the halo multiplicity function of mock catalogs (Berlind et al. 2006). The group-finding algorithm and mock catalog tests are described in detail in their paper, as are the effects of the survey edges and fiber collisions on the multiplicity function..

For the halo model, the group multiplicity function, $n_{\rm grp}(N)$, is completely determined by the abundance of the halos and the halo occupation distribution, and hence can be used to constrain HOD parameters. It is obtained by summing over the halo mass function and the probability distribution that a halo of mass M contains N galaxies (1 central plus $N_{\rm sat}$ satellites):

$$n_{\rm grp}(N) \;=\; \int_{M_{\rm min}} {\rm d}M\, \frac{{\rm d}n(M)}{{\rm d}M}\, P(N_{\rm sat}|M), \qquad (8.2)$$

where we assume that $P(N_{\rm sat}|M)$ is a Poisson distribution. Our halo-model group multiplicity function for $M_r < -20$ is compared to the measured one in Figure 8.1. It is encouraging that they are in agreement. The multiplicity function resulting from applying the HOD of all SDSS galaxies for such volume-limited cuts (Zehavi et al. 2005) in $P(N_{\rm sat}|M)$ is also shown

for comparison. Since this HOD parameterization characterizes halos with fewer satellites as a function of halo mass, it is not surprising that it results in a multiplicity function that is too low.

One of our main goals is to analyze the luminosity-marked projected correlation function of galaxy groups, analogous to the analysis in Skibba et al. (2006). Accordingly, it is useful to compare the halo-model and measured mean central and satellite luminosities as a function of multiplicity. Similar to equation (8.2), the halo model description of the mean luminosities involves an integral over the mass function and the HOD:

$$\bar{L}_{\rm cen}(N) \;=\; \int_{M_{\rm min}} {\rm d}M \, \frac{{\rm d}n(M)}{{\rm d}M} \, P(N_{\rm sat}|M) \, L_{\rm cen}(M) \tag{8.3}$$

and

$$\bar{L}_{\rm sat}(N) \;=\; \int_{M_{\rm min}} {\rm d}M \, \frac{{\rm d}n(M)}{{\rm d}M} \, P(N_{\rm sat}|M) \, \langle L_{\rm sat}|M, M_{\rm min}\rangle, \tag{8.4}$$

where $L_{\rm cen}(M)$ and $\langle L_{\rm sat}|M, M_{\rm min}\rangle$ are determined by the halo mass function, the galaxy luminosity function, and the halo occupation distribution (Chapter 3 here; Skibba et al. 2006).

Our halo-model central galaxy and satellite galaxy luminosities for $M_r < -20$ are compared to the measured ones in Figures 8.2 and 8.3. The halo-model prediction of $L_{\rm cen}(N)$ appears to be offset from the measurement by a small and approximately constant amount, although it is still consistent within the uncertainties. The offset may be due to the luminosity function we assumed; the group luminosity function may be significantly different than the overall LF (Zandivarez et al. 2006), and central galaxy luminosity as a function of richness is dependent on the characteristic luminosity L_* and number density ϕ_*.

The halo-model prediction of $L_{\rm sat}(N)$ agrees well with the measurement. Interestingly, the mean satellite luminosity in both the SDSS and the halo model is almost independent of richness (except for the extreme HOD, magenta curve), just as it is almost independent of halo mass (Skibba et al. 2006).

This is not inconsistent with the fact the normalization of the *luminosity function* of satellite galaxies increases with cluster richness and the characteristic luminosity becomes brighter (Hansen et al. 2005). We have shown that *given some luminosity threshold*, the mean satellite luminosity does not strongly depend on richness; however, as richness increases

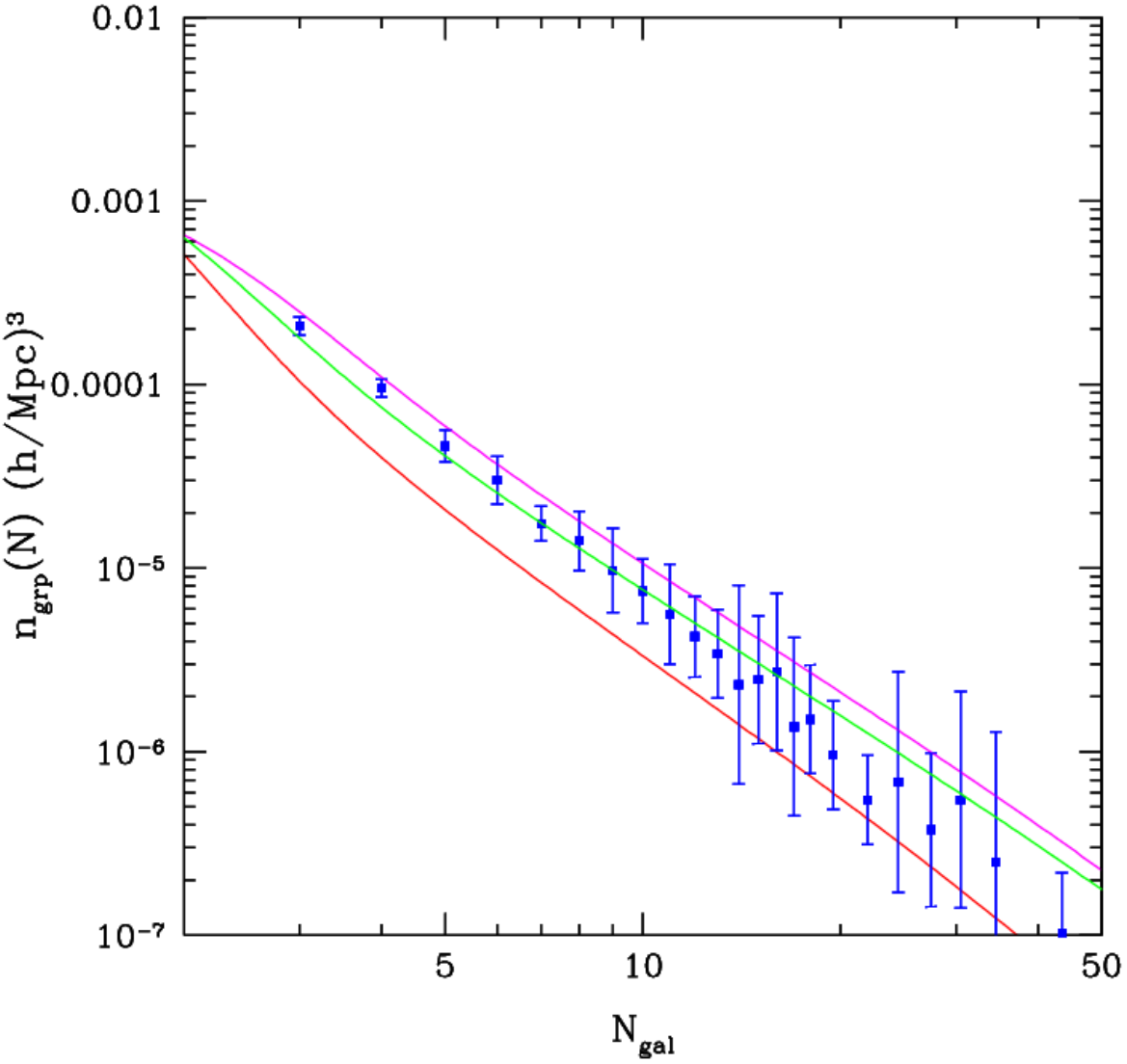

Figure 8.1: Group multiplicity function for $M_r < -20$. Blue points show the measurement from the SDSS group catalog. Halo-model predictions are shown for our fiducial HOD ($\mu = 8$ and $\alpha = 1.19$, green curve), the extreme HOD that 'fits' the small-scale correlation function ($\mu = 4$ and $\alpha = 1.1$, magenta curve), and for the HOD of galaxies of all environments (Zehavi et al. 2005: $\mu = 23$ and $\alpha = 1.15$, red curve).

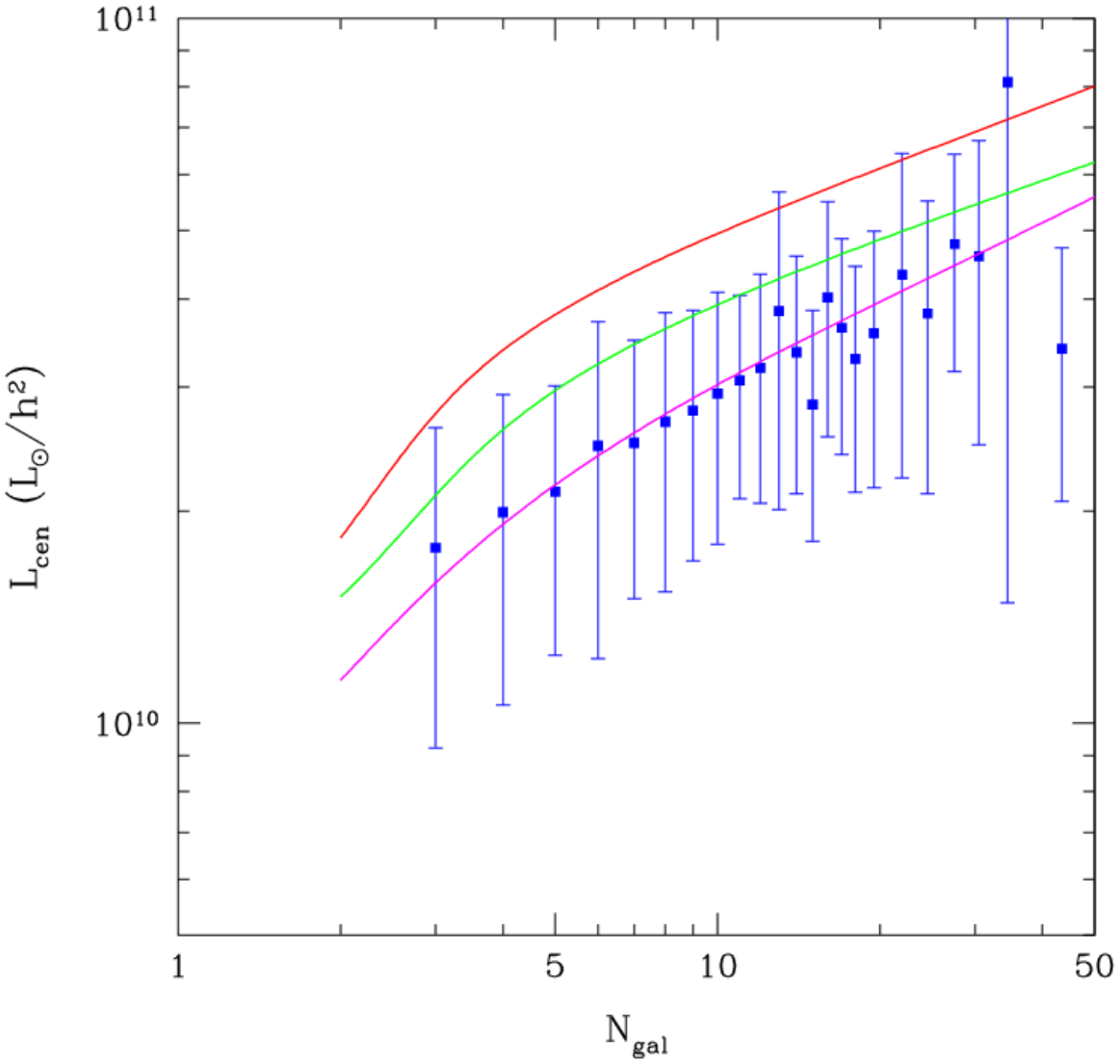

Figure 8.2: Central galaxy luminosity as a function of richness for $M_r < -20$. Blue points show the measurement from the SDSS group catalog. Halo-model predictions are shown for our fiducial HOD ($\mu = 8$ and $\alpha = 1.19$, green curve), the extreme HOD that 'fits' the small-scale correlation function ($\mu = 4$ and $\alpha = 1.1$, magenta curve), and for the HOD of galaxies of all environments (Zehavi et al. 2005: $\mu = 23$ and $\alpha = 1.15$, red curve).

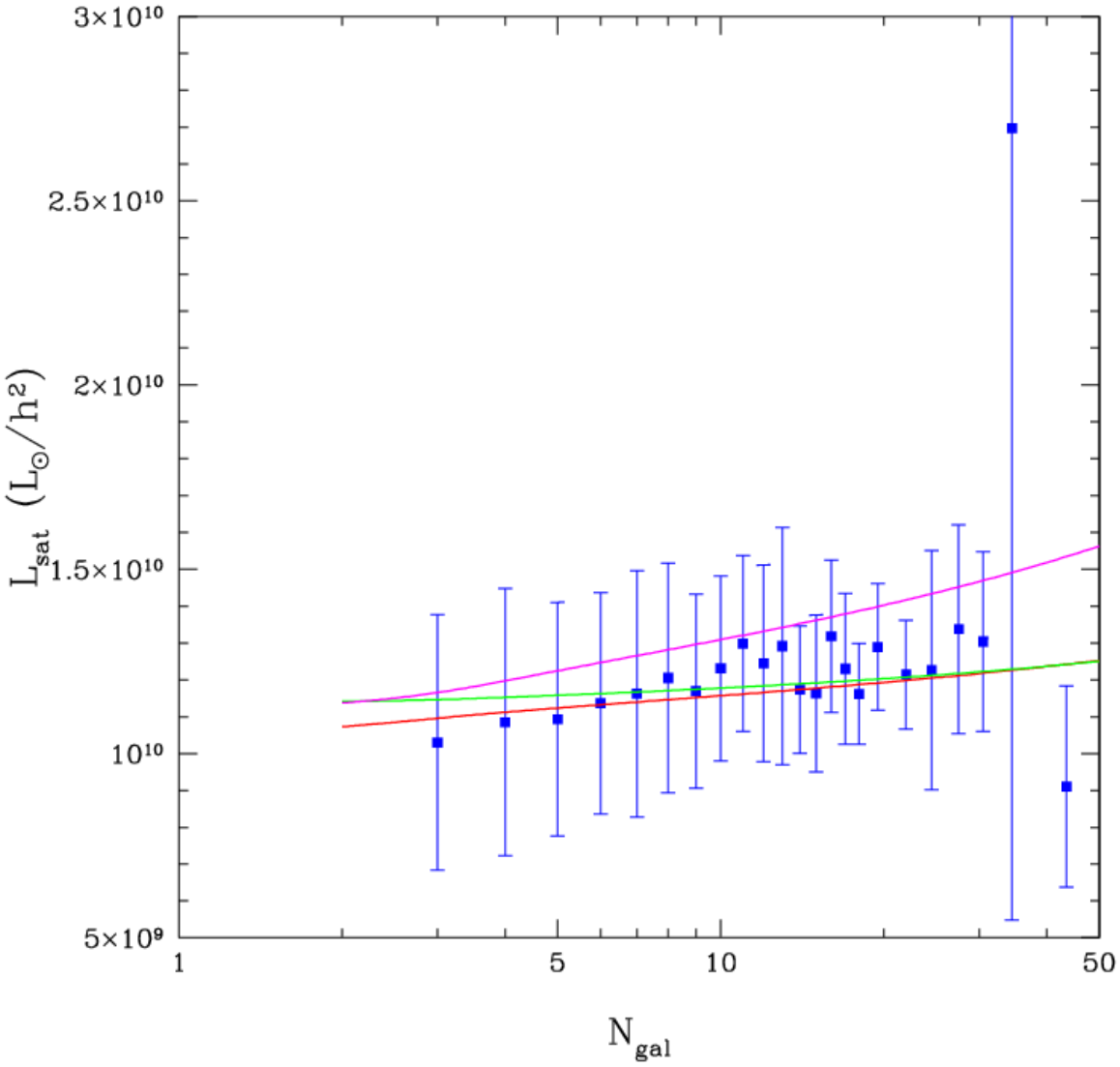

Figure 8.3: Satellite galaxy luminosity as a function of richness for $M_r < -20$. Blue points show the measurement from the SDSS group catalog. Halo-model predictions are shown for our fiducial HOD ($\mu = 8$ and $\alpha = 1.19$, green curve), the extreme HOD that 'fits' the small-scale correlation function ($\mu = 4$ and $\alpha = 1.1$, magenta curve), and for the HOD of galaxies of all environments (Zehavi et al. 2005: $\mu = 23$ and $\alpha = 1.15$, red curve).

it is expected that central galaxy luminosities, and their associated satellite luminosities, also increase. Our results are also consistent with those of Lin & Mohr (2004), who argued that the luminosity growth rates of BCGs are slower than the luminosity growth rates of clusters because BCGs account for a decreasing fraction of the total light in galaxies in clusters. This simply occurs because, in the most massive halos, central galaxy luminosity increases with mass at a slower rate than the satellite galaxy halo occupation distribution.

Although the halo model describes the group multiplicity function well, as well as the central and satellite galaxy luminosities as a function of multiplicity, the next section shows that these are necessary but not sufficient to correctly describe the small-scale two-point correlation function and luminosity-marked correlation function.

8.3 MARKED PROJECTED CORRELATIONS OF SDSS GALAXY GROUPS

We show preliminary results in this section, comparing the marked and un-marked projected correlation functions of the halo model, SDSS group catalogs, and SDSS-based mock catalogs. We show the results for $M_r < -20$; we have found the same (in)consistencies in our $M_r < -19$ comparisons.

The halo-model description of marked projected correlation functions are described in Chapter 4. In order to account for the fact that only $N \geq 3$ groups are included, which means groups with two or more satellites, one could naïvely raise the minimum halo mass in the number density and correlation function integrals such that $\langle N_{\rm sat}|M\rangle \geq 2$, or approximately $M_{\rm min} \rightarrow 2\,M_1$.

The more accurate approach, which we employ here, accounts for the distribution of satellite occupation numbers. We sum over the distribution, but starting at $N_{\rm sat} = 2$. In the case of the number density of galaxies (equation 4.6), it becomes

$$\bar{\rho}_{\rm gal} = \int_{M_{\rm min}} {\rm d}M \, \frac{{\rm d}n(M)}{{\rm d}M} \sum_{N_{\rm sat}=2}^{\infty} P(N_{\rm sat}|M) \langle N_{\rm gal}|M, M_{\rm min}\rangle, \tag{8.5}$$

where, assuming the satellites have a Poisson distribution, the sum is the following (in which $g \equiv \langle N_{\rm gal}|M, M_{\rm min}\rangle$):

$$\begin{aligned}
&\sum_{N_{\rm sat}=2}^{\infty} \left(\frac{g^{N_{\rm sat}}}{N_{\rm sat}!}\, e^{-g} \right) (\langle N_{\rm cen}|M\rangle \, + \, \langle N_{\rm sat}|M\rangle) \qquad (8.6)\\
&= \langle N_{\rm cen}|M\rangle \, + \, \langle N_{\rm sat}|M\rangle \, - \, e^{-N_{\rm sat}} (\langle N_{\rm cen}|M\rangle + \langle N_{\rm cen}|M\rangle \langle N_{\rm sat}|M\rangle + \langle N_{\rm sat}|M\rangle)\\
&= 1 \, + \, \langle N_{\rm sat}|M\rangle \, - \, e^{-N_{\rm sat}} (1 \, + \, 2\langle N_{\rm sat}|M\rangle).
\end{aligned}$$

The exponential correction term only matters at low N.

The power spectrum is a sum of two terms, one term due to pairs of objects in the same halo (the 'one-halo term'), and another due to pairs in separate halos (the 'two-halo term'). Similar to equation (8.5), in the integral over halo mass for the two-halo term of the power spectrum, $P_{2h}(k)$, we sum over the product of the halo occupation distribution and the density profile:

$$\begin{aligned}
\langle N_{\rm gal}|M\rangle \, u_{\rm gal}(k|M) \quad &\rightarrow \quad \sum_{N_{\rm sat}=2}^{\infty} P(N_{\rm sat}|M) \, (\langle N_{\rm cen}|M\rangle \, + \, \langle N_{\rm sat}|M\rangle u_{\rm gal}(k|M)) \qquad (8.7)\\
&= \quad 1 \, + \, \langle N_{\rm sat}|M\rangle u_{\rm gal}(k|M) \, - \, e^{-N_{\rm sat}} ((1 + \langle N_{\rm sat}|M\rangle) \, + \, \langle N_{\rm sat}|M\rangle u_{\rm gal}(k|M)).
\end{aligned}$$

We similarly modify the two-halo term of the weighted power spectrum, $W_{2h}(k)$, in which the central and satellite galaxies are weighted by the respective central galaxy and satellite galaxy weights:

$$\begin{aligned}
&\sum_{N_{\rm sat}=2}^{\infty} P(N_{\rm sat}|M) \, (W_{\rm cen} N_{\rm cen} \, + \, \langle W_{\rm sat}\rangle \langle N_{\rm sat}\rangle u_{\rm gal}(k|M)) \qquad (8.8)\\
&\quad = W_{\rm cen} \, + \, \langle W_{\rm sat}\rangle \langle N_{\rm sat}\rangle u_{\rm gal}(k|M)\\
&\qquad - \, e^{-N_{\rm sat}} (W_{\rm cen}(1 + \langle N_{\rm sat}\rangle) \, + \, \langle W_{\rm sat}\rangle \langle N_{\rm sat}\rangle u_{\rm gal}(k|M)),
\end{aligned}$$

where, in the case of the luminosity mark, $W_{\rm cen}(M) = L_{\rm cen}(M)$ and $\langle W_{\rm sat}|M, M_{\rm min}\rangle = \langle L_{\rm sat}|M, M_{\rm min}\rangle$, described in the previous section. The un-weighted and weighted one-halo term of the power spectrum, which consists of the center-satellite and satellite-satellite term, is modified analogously.

For the SDSS group catalogs, Berlind et al. (2006) discuss the effects of fiber collisions and of the survey edges, and these could potentially affect our correlation function

measurements as well. Following Zehavi et al. (2005), Berlind et al. correct for the effect of fiber collisions by including each collided galaxy at the redshift of its nearest neighbor. This correction results in boosting the small-scale ($r \sim 100\,\mathrm{kpc}/h$) redshift-space correlation function, $\xi(s)$, but we focus on *projected* correlations, which are not significantly affected.

As shown in Chapter 4, the projected two-point correlation function involves summing over the redshift-distorted correlation function along the line-of-sight:

$$w_p(r_p) \;=\; \int \mathrm{d}r\, \xi(r_p, \pi) \;=\; 2 \int_{r_p}^{\infty} \mathrm{d}r\, \frac{r\,\xi(r)}{\sqrt{r^2 - {r_p}^2}}. \tag{8.9}$$

The marked projected correlation function is simply the ratio of the weighted and and unweighted projected correlation functions:

$$M_p(r_p) \;=\; \frac{1 \,+\, W_p(r_p)/r_p}{1 \,+\, w_p(r_p)/r_p}. \tag{8.10}$$

The difficulty about accounting for edge effects is that, although groups straddling the edges must be excluded, conservatively rejecting many groups within some distance of the edges may unnecessarily reduce the sample size. Using mock group catalogs, Berlind et al. (2006) found that a minimum distance of 500 kpc in the tangential direction and 500 km/s in the redshift direction eliminates most groups affected by the edges. For such low redshifts, the buffer along the line-of-sight amounts to up to distances up to 5 Mpc/h. To ensure that our random catalogs cover the same volume in the sky as the edge-corrected group catalogs, we use the same redshift distributions when generating them.

First, we compare in Figure 8.4 the projected correlation function and luminosity-marked projected correlation function of the halo model to our SDSS-based mock galaxy catalog for $M_r < -20$. The mock catalog is constructed so that the galaxies have the same real-space two-point correlation function as measured in the SDSS and the same luminosity function. The satellite galaxies are generated so that they have the same halo occupation distribution as in our halo model calculations, and the luminosities are generated so that on average they have the same relation with halo mass. Finally, the redshift-distorted coordinates are assigned by assuming that a galaxy's velocity is the sum of the velocity of the parent halo plus a virial motion contribution that is drawn from a Maxwell-Boltzmann distribution with mass-dependent dispersion (see Skibba et al. 2006).

It is important to note that the number of galaxies in a halo, the spatial distribution of galaxies within a halo and the assignment of luminosities all depend only on halo mass, and not on the surrounding large-scale structure. Therefore, the mock catalog includes only those environmental effects which arise from the environmental dependence of halo abundances.

The top panel of Figure 8.4 shows that the halo-model prediction of the correlation function is in excellent agreement with the measurement from the mock catalog. The agreement is not perfect, and this is likely due to the redshift distortions imposed on the mock galaxies. Note that the amplitude of $w_p(r_p)$ is approximately an order of magnitude larger than correlation functions of catalogs that include isolated and paired galaxies.

The bottom panel compares two halo-model predictions of the luminosity-marked correlation function to the measurement. The dashed curve is the result of weighting central and satellite galaxies the same way, by the mean luminosity of all galaxies in a halo. The solid curve is the result of weighting the central galaxies by $L_{\rm cen}(M)$ and the satellite galaxies by $\langle L_{\rm sat}|M\rangle$. The agreement between this and the measurement is reassuring, since the mock catalog does treat central galaxies differently than the satellites.

The luminosity-marked correlation function clearly does not monotonically decrease with scale, as one might expect. When isolated and paired galaxies are not excluded, the marked correlation function is always greater than unity, indicating that bright luminosities tend to reside in denser environments than faint luminosities, regardless of the scale at which the environmental dependence is measured (Skibba et al. 2006). Since isolated and paired galaxies are excluded here and the focus is on galaxies in denser regions, the mean luminosity mark is large and this strongly affects marked correlations, which are relative to the mean mark. The dip below unity in Figure 8.4 occurs on scales at which the transition between the satellite-satellite term and two-halo term occurs. The two-halo term at these scales also consists mostly of satellite galaxies, as well as central galaxies in small halos. Hence, the marked correlation function for $N \geq 3$ galaxy groups at $r \sim 1\,{\rm Mpc}/h$ is dominated by satellite galaxies with faint luminosities, fainter than the mean. In contrast, the excluded galaxies consisted mostly of single central galaxies ($\sim 70\%$) and center-satellite pairs, which would contribute pair counts more luminous than the mean, thus producing the luminosity-marked correlation functions in Skibba et al. that are continually greater than unity.

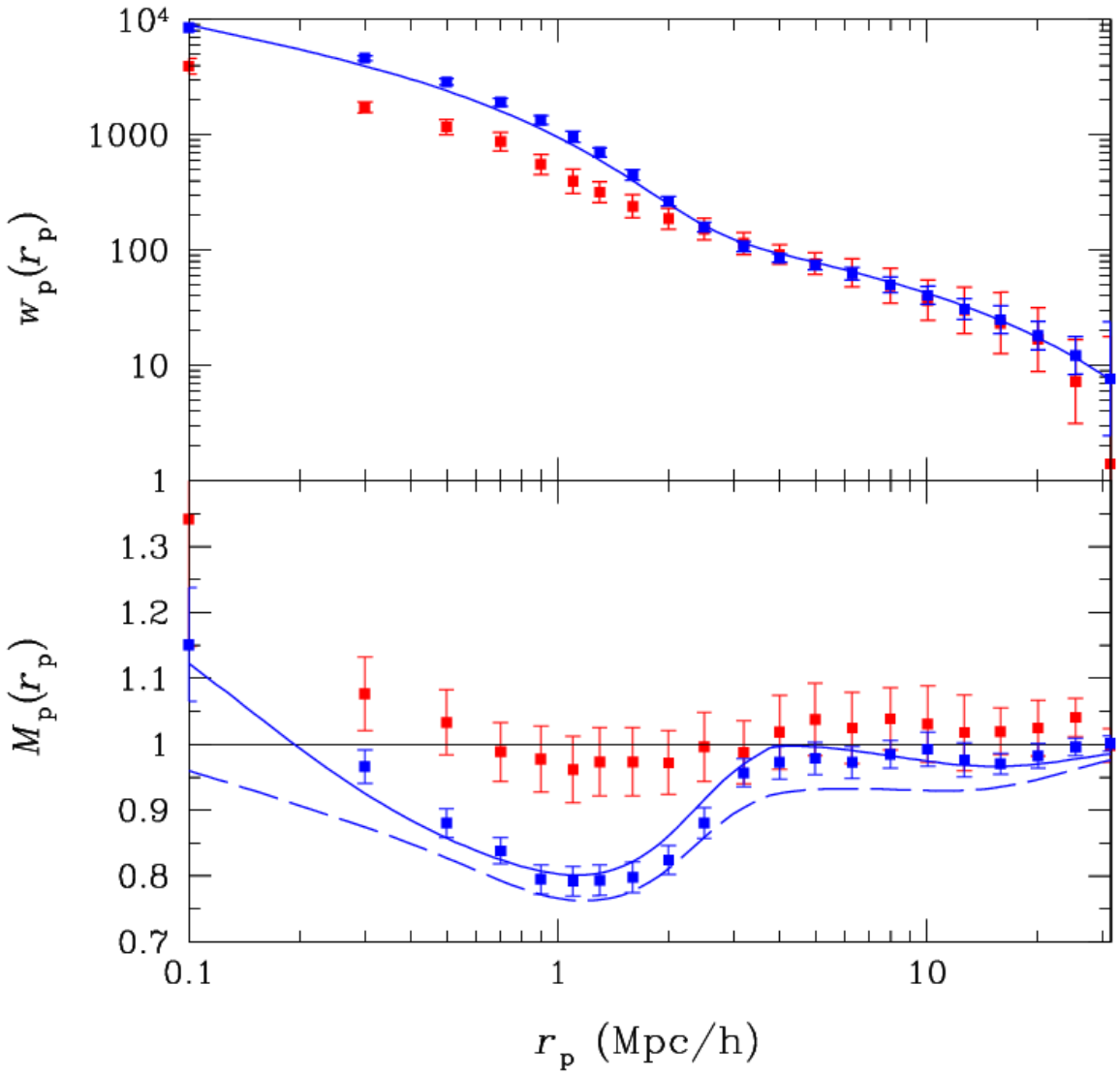

Figure 8.4: Projected correlation function and luminosity-marked projected correlation function members of galaxy groups for $M_r < -20$. The halo-model prediction (blue curves) is compared to the mock catalog measurement (blue points). In the bottom panel, the solid curve shows the halo-model prediction when the luminosity of the central galaxy is assumed to be different than the others, and the dashed curve shows the prediction when the central galaxy is not special. The SDSS measurement (red points) is also shown, for comparison. Error bars are jack-knife estimates.

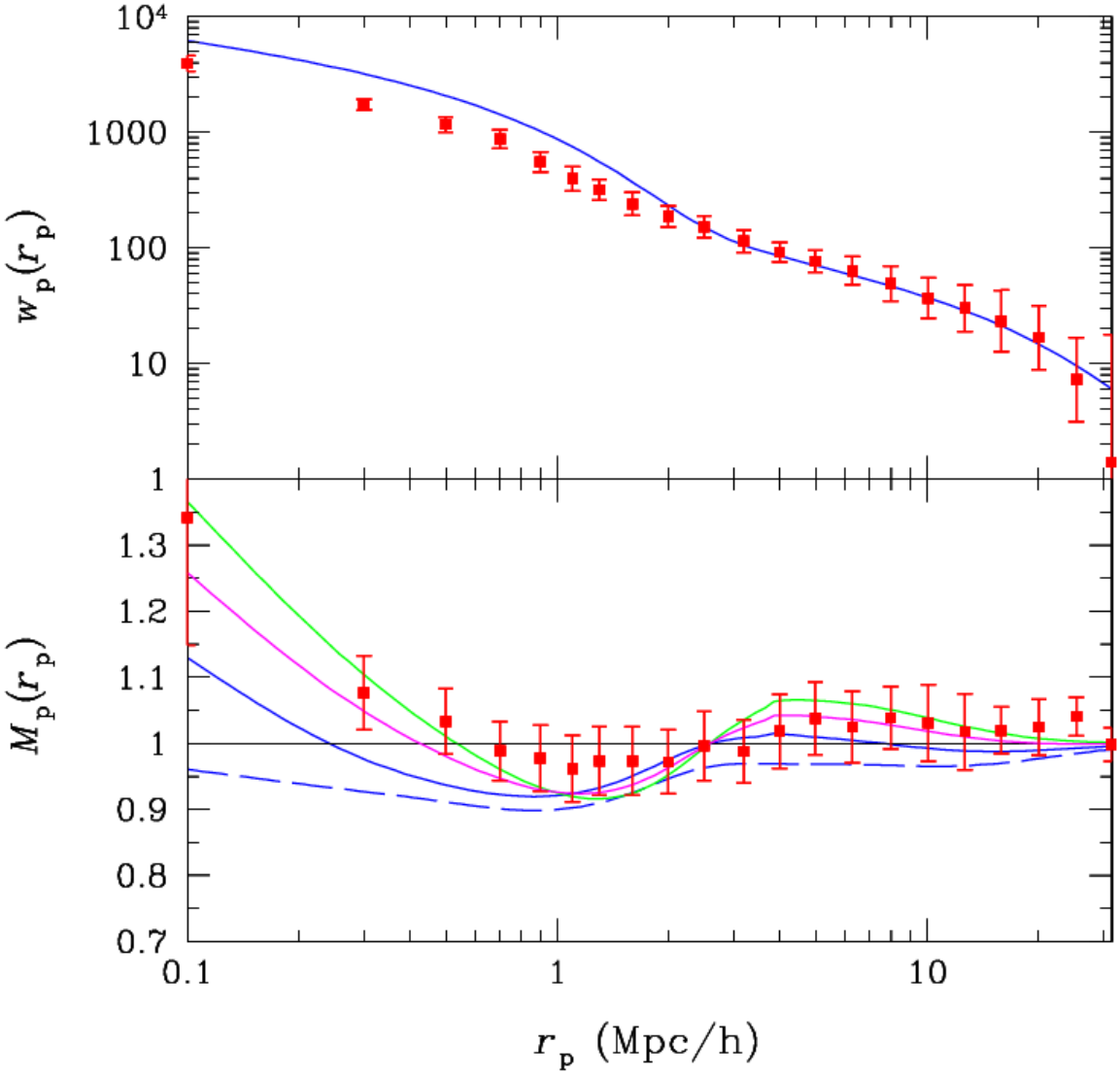

Figure 8.5: Projected correlation function and luminosity-marked projected correlation function members of galaxy groups for $M_r < -20$. The halo-model prediction (blue curves) is compared to the SDSS group catalog measurement (red points). In the bottom panel, the magenta and green curves are the halo-model predictions when the satellite luminosities are allowed to vary with the positions within the host halo (see Section 4).

For comparison, we also show the measurement of the $M_r < -20$ SDSS group catalog. On large scales ($r_p > 3\,\mathrm{Mpc}/h$), the projected correlations agree with the mock catalog. On small scales, where the one-halo term dominates, there is a vast discrepancy in both the correlation function and the luminosity-marked projected correlations. Only a small fraction of this discrepancy is due to the fact that the mock and SDSS group catalogs are at different redshifts. The discrepancy is *not* due to our assumed cosmological parameters. Measurements of the first year of WMAP plus 2dF (Sánchez et al. 2006) and of the third year of WMAP (Spergel et al. 2006) show that $(\Omega_m,\ \sigma_8) \approx (0.24,\ 0.75)$ rather than our assumed (0.3, 0.9). However, although the lower cosmological parameters slightly narrow the small-scale gap in $w_p(r_p)$, they create a larger discrepancy on large scales that was not there before. This might not seem surprising, since cosmological measurements of WMAP have been found to be somewhat inconsistent with those of the SDSS, from which the group catalogs were constructed.

We compare the halo model (at $z = \bar{z}$ of the catalog) to the SDSS measurement directly in Figure 8.5. The small-scale $w_p(r_p)$ are discrepant by as much as a factor of two or more up to $r \sim 2\,\mathrm{Mpc}/h$. A similar discrepancy occurs for $M_r < -19$ up to $r \sim 1.5\,\mathrm{Mpc}/h$. In the bottom panel, we relax the assumption that satellite luminosities are independent of their position within the halo (see Section 4), and the resulting luminosity-marked correlation functions of the halo model are in better agreement. Satellites with brighter luminosities and redder colors have been observed closer to the cluster center than satellites in the outskirts, and since larger luminosities and colors tend to be found in denser regions, this results in stronger small-scale luminosity- and color-marked correlations of group galaxies.

However, it is premature to interpret the marked correlation function until the un-marked correlation function can be explained. We speculate that the small-scale discrepancy with the data is due to various group-specific processes which are not currently accounted for in the halo model. In particular, this appears to be due to the combined effect of central galaxies having velocity dispersion relative to the satellite galaxies, resulting in them not being located at the halo center, and of satellite galaxies having a less concentrated density distribution than that of dark matter, due to dynamical friction, tidal stripping, and other processes. The extent to which these effects reduce the strength of the small-scale clustering of galaxies is

not yet known. We discuss these two issues, as well as that of satellite luminosities depending on their positions within the halo, in the next section.

8.4 GROUP-SPECIFIC PROCESSES AFFECTING CENTRALS & SATELLITES

In this section, we discuss some possible explanations for the small-scale discrepancy between the halo-model and SDSS marked and un-marked projected correlation functions shown in the previous section. The discrepancy occurs on scales at which the one-halo term dominates, and since our halo occupation distribution is consistent with the group multiplicity function and satellite luminosities as a function of multiplicity, this suggests that group-specific physical processes may be involved.

8.4.1 Central and Satellite Galaxy Positions

The small-scale projected correlation function of SDSS group galaxies is low compared to the halo model, which means that galaxies in groups are less clustered within halos than expected. This may be due to the combined effect of two processes: many central galaxies have some velocity dispersion with respect to the satellites, resulting in them not being located at the halo center (van den Bosch et al. 2005b), and satellite galaxies tend to have a less concentrated distribution than dark matter, due to dynamical friction, tidal stripping, and other processes (Lin et al. 2004, Nagai & Kravtsov 2005).

Since groups have a wide range of richnesses and occupy halos of a wide range of sizes, it is necessary to examine the central and satellite galaxy positions relative to a distance scale that increases with richness. The virial radius of the group halos is such a scale, and we estimate it by taking the luminosity of the central galaxy (assumed to be the brightest in each group) and obtaining the corresponding host halo mass (Figure 8.2), and then converting that to $r_{\rm vir}$.

Figure 8.6 shows the distribution of the projected satellite distances from the group

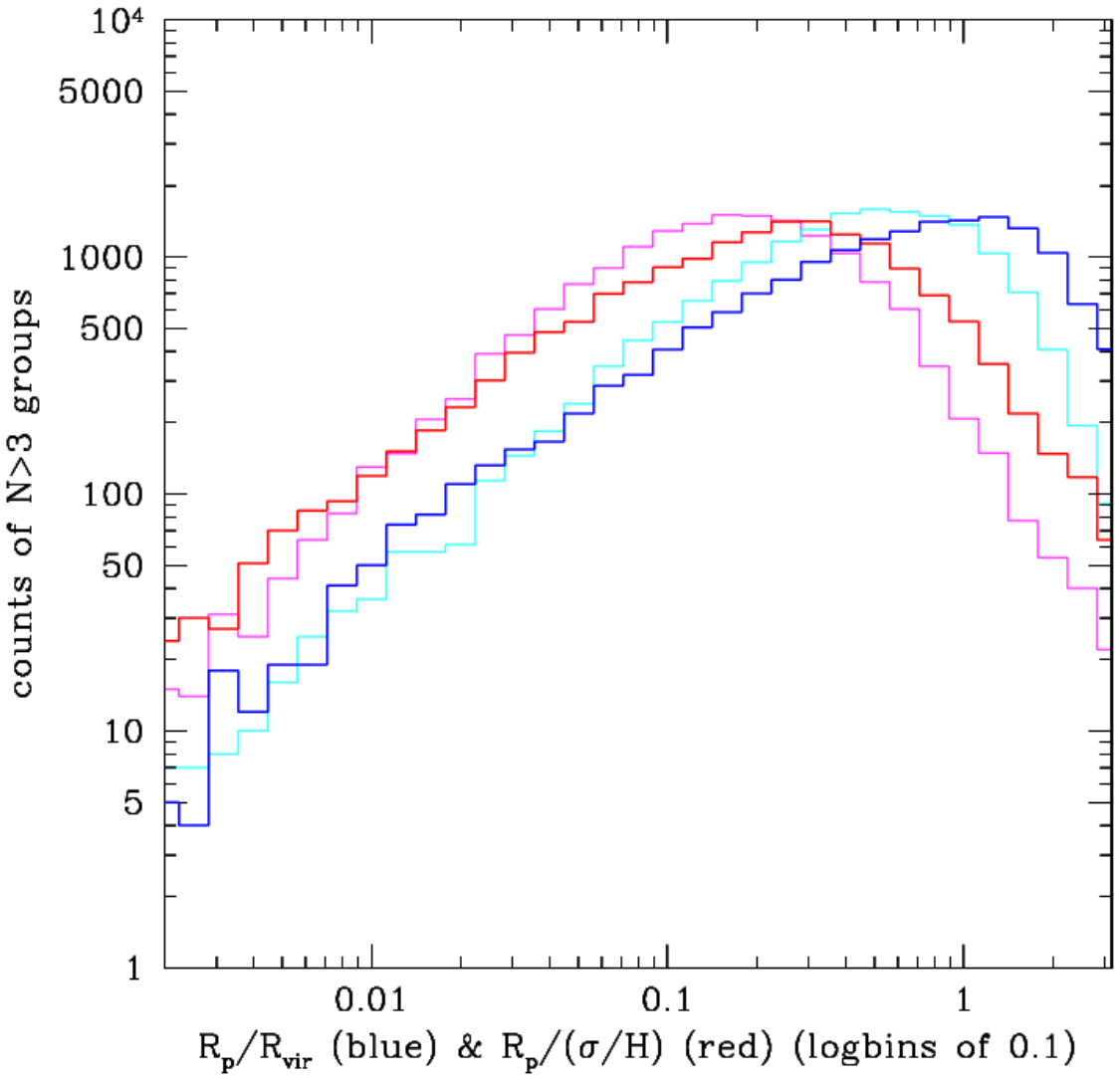

Figure 8.6: Distributions of projected distances of satellite galaxies from group center normalized by group's velocity dispersion (red and magenta) and by host halo's virial radius (blue and cyan) for $M_r < -20$. Distributions of distances from central (brightest) galaxy (red and blue) are compared to distances from group centroid (magenta and cyan).

center, normalized by the estimated virial radius of each group (blue and cyan histograms), for the $M_r < -20$ group member catalog. For comparison, we show the distribution of projected satellite distances normalized by the group's velocity dispersion (red and magenta histograms). The distributions are offset by approximately a factor of three, with $r_{\rm sat}/r_{\rm vir}$ indicating a more diffuse distribution. Note that we have not corrected for the offset in $L_{\rm cen}$ in Figure 8.2, which means that the estimated $r_{\rm vir}(M_{\rm vir}(L_{\rm cen}))$ are somewhat low and the distances normalized by $r_{\rm vir}$ are slightly overestimated. This only amounts to a small net offset, however.

We have used two measures of 'group center', and the difference between them yields information about the central galaxies. We have shown the satellite distances from the unweighted group *centroid*, and from the *central galaxy*, which we assume to be the brightest galaxy in the group. From the peaks of the distributions in Figure 8.6, it is clear that satellite galaxies are generally further away from central galaxy than from the centroid.

We have also examined the relative satellite distances for different ranges of richness, and we have found no N-dependence in the shape of the distributions.

We further explore the central and satellite galaxy positions by examining the dependence of the relative satellite distances on host halo mass. First, we compare the mean projected satellite distances from the group centroid and central galaxy for our $M_r < -20$ mock group catalog, in Figure 8.7. The distances from the group centroid are less than from the central galaxy in low mass groups, but they are the same on average for $M_{\rm host} \geq 10^{14} M_\odot$, meaning that the central galaxy settled at the group centroid in these massive groups. Also, mean satellite distances from the central galaxy are almost independent of host halo mass, between $0.35 \leq \langle r_{\rm sat}/r_{\rm vir} \rangle \leq 0.4$. Our mock catalog assumes that the galaxies are distributed with Navarro, Frenk & White (1997) profiles with Bullock et al. (2001a) concentrations, and we have also shown the analytic prediction for the mean projected distances:

$$\langle r_p/r_{\rm vir}(M) \rangle \;=\; \frac{\int {\rm d}r'_p \, 2\pi r'_p \, \rho_{\rm gal}(r'_p|M) \, r'_p}{\int {\rm d}r'_p \, 2\pi r'_p \, \rho_{\rm gal}(r'_p|M)},$$

where $r'_p \equiv r_p/r_{\rm vir}$, and $\rho(r_p)$ is an integral over the density profile along the line-of-sight. As expected, the mean satellite distance as a function of mass according to the above equation, assuming an NFW profile, is the same as that of the mock catalog.

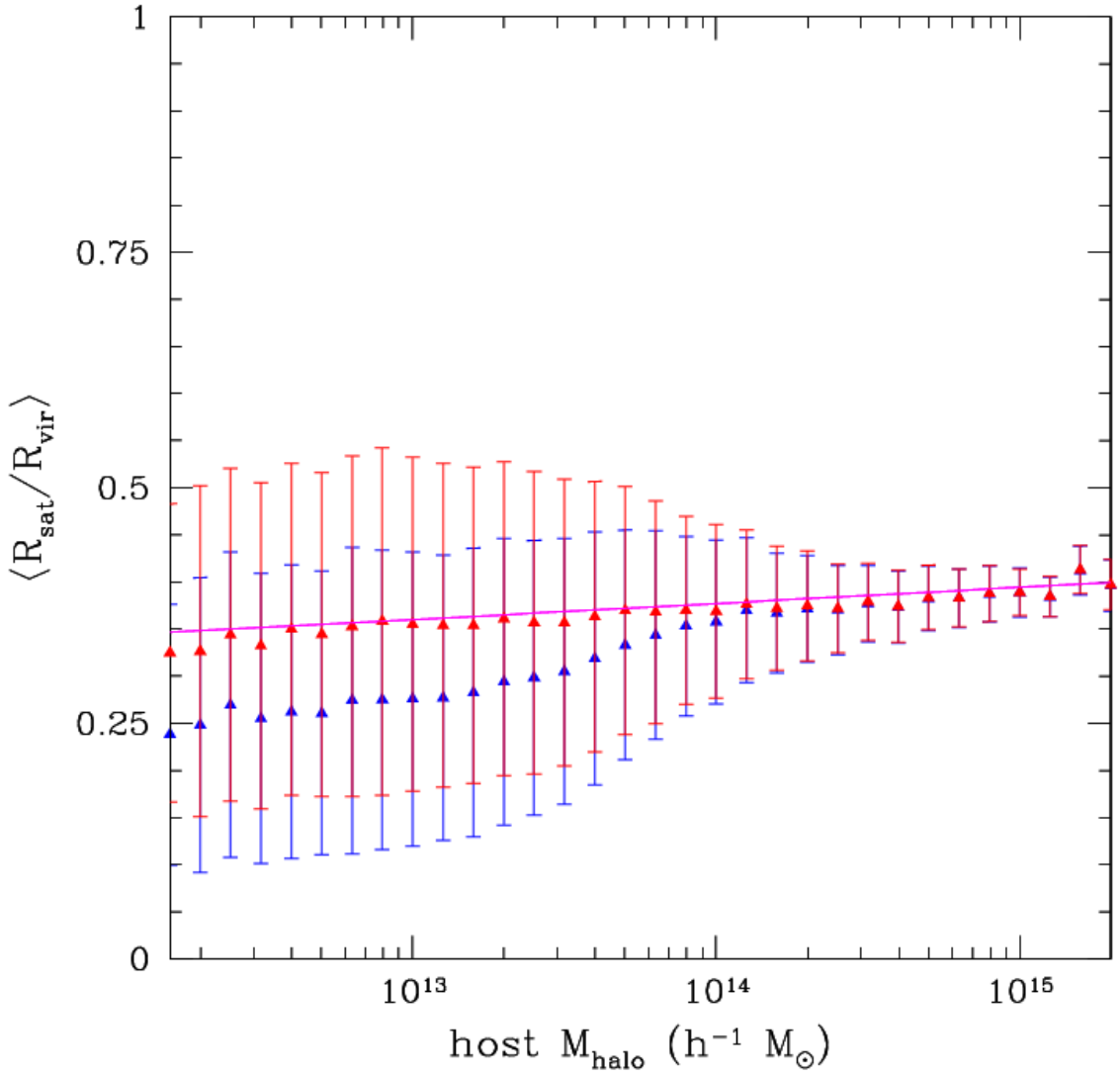

Figure 8.7: Mean projected satellite distances from group centroid (blue points) and from central galaxy (red points) as a function of host halo mass for $M_r < -20$ mock catalog. Magenta line shows analytic calculation, assuming a Navarro, Frenk & White 1997 density profile with Bullock et al. 2001a concentration, as was assumed in the mock.

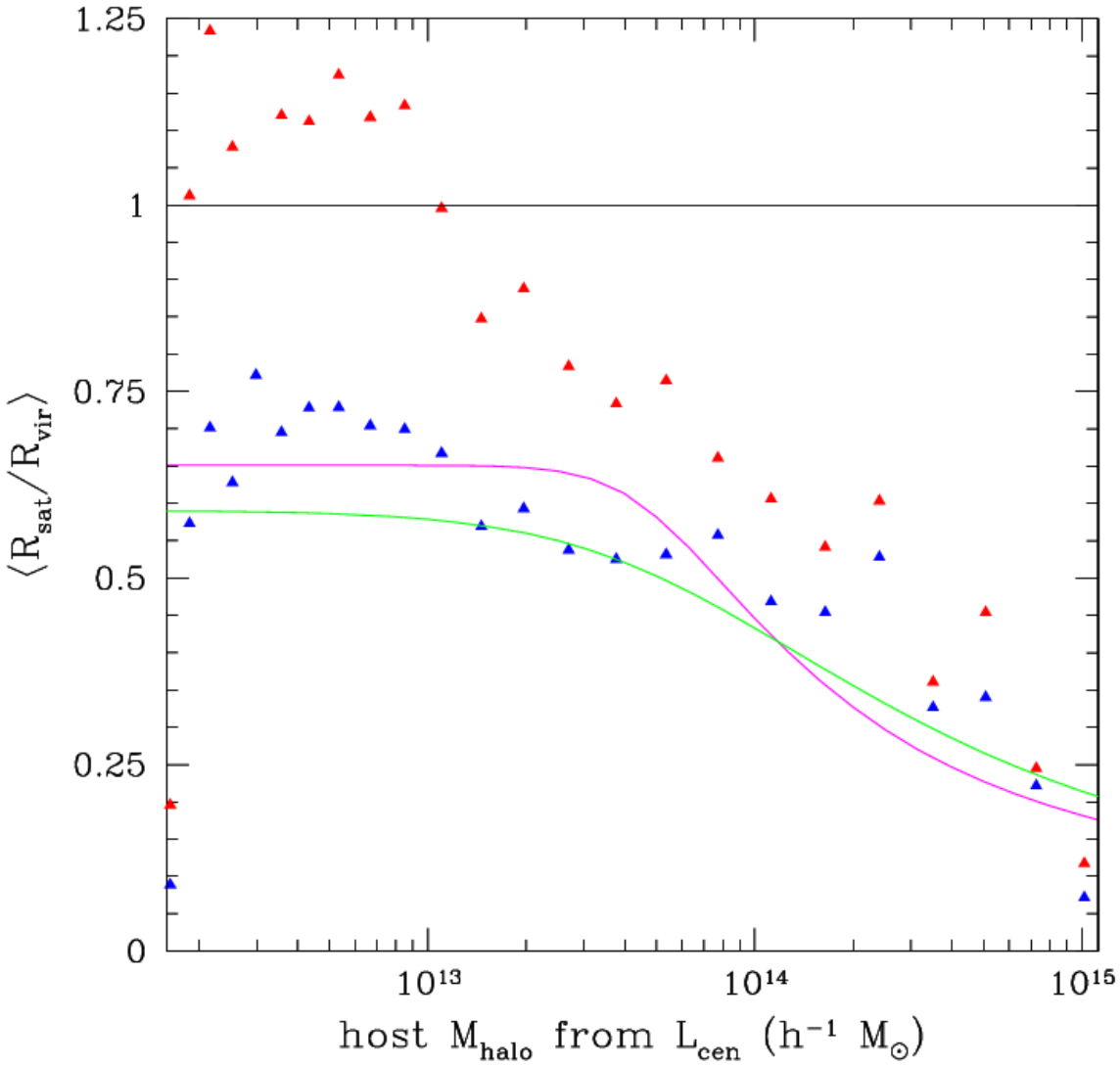

Figure 8.8: Mean projected satellite distances from group centroid (blue points) and from central galaxy (red points) as a function of host halo mass for $M_r < -20$ SDSS catalog. Scatter in each bin is almost as large as the means themselves, and is not shown for clarity. Two analytic calculations are shown, with an NFW profile with low concentration *increasing* with mass (green), and with same concentration but for shallower inner density profile (magenta).

Figure 8.8 shows the mean projected satellite distances from the group centroid and central galaxy for the $M_r < -20$ SDSS group catalog. The host halo masses and corresponding virial radii were estimated from the central galaxy luminosities (Figure 8.2), and as mentioned above, they are slightly offset, but the offsets in $r_p/r_{\rm vir}$ and $M_{\rm vir}$ partially cancel. The SDSS measurement is strikingly different than the mock catalog result, in a number of ways. Firstly, although there is a lot of scatter, the central galaxy and centroid positions are very different at low masses and do not merge until *very* large halo masses, with the difference between them dropping below $20\% r_{\rm vir}$ only for $M > 3 \times 10^{14} M_\odot$. However, the relative offset between the central and centroid positions is interestingly similar to that of the mock catalog. Secondly, the satellite galaxies in low-mass, low-N, halos are spread out more than in the mock catalog, such that for $M < 10^{13} M_\odot$ they are at least a virial radius away from the central galaxy *on average*. Many of these halos contain only a few galaxies, and the central galaxy is often a significant distance from the centroid, suggesting that some of the groups are experiencing mergers and are still forming.

Thirdly, the mean satellite distances are much larger those of our mock catalog for host halo masses up to $4 \times 10^{14} M_\odot$, and then they are *more* concentrated in richer groups. The rapidly decreasing satellite distances suggest that concentration is *increasing* significantly with halo mass and richness, which may be surprising. This is contrary to other results that there is no significant variation of concentration with mass (Lin, Mohr & Stanford 2004; Yang et al. 2005c), or that it decreases slightly with mass (Hansen et al. 2005). The latter's $c_{\rm gal}(M_{\rm vir})$ measurement results in a mean satellite distance that increases with mass and is always less than half the virial radius. For illustrative purposes, we have shown two analytic calculations in Figure 8.8, both with low concentrations that increase with halo mass, but with different inner density profiles. However, our results are consistent with others in the sense that satellite concentrations lower than DM concentrations are required to produce large values of $r_{\rm sat}/r_{\rm vir}$.

In any case, these results show that the issue of central galaxy velocity dispersion must be addressed prior to analyzing the satellite galaxy distributions. Our preliminary analysis of the SDSS group catalogs along the lines of van den Bosch et al. (2005b) yields evidence for velocity bias in the central galaxies relative to the satellite galaxies, but the value of

the velocity bias as a function of mass is not yet clear, and it will be necessary for accurately modeling central galaxy distributions in our halo-model predictions of group galaxy clustering. We will need to disentangle the effects of central galaxy velocity bias and diffuse satellite galaxy distributions, both of which will likely decrease to some extent the small-scale clustering predicted by the halo model, although their relative importance will likely vary with scale.

8.4.2 Satellite Galaxy Position-Dependent Marks

In addition to the velocity dispersion of central galaxies and the diffuse distributions of satellite galaxies, we must consider the possibility that the marks of satellite galaxies vary *within their host halo*. That is, although most galaxy properties vary with halo mass and richness, we have heretofore assumed, in our halo model and mock catalog, that the mean satellite marks do not vary with their distance from the halo center.

For some galaxy properties, such as luminosity, color, and stellar mass, high values of the marks are more clustered in denser environments (*i.e.*, brighter, redder, more massive galaxies tend to be in overdense regions). Marked statistics with such galaxy properties are consequently generally above unity (Chapters 5 and 6), except when the mean mark is high (Section 2). The mean value of these galaxy properties within halos can be estimated from the halo mass, but if the galaxy properties are also clustered within the halos, with brighter, redder, more massive satellite galaxies concentrated near the halo center and fainter bluer satellites in the halo outskirts, this generally will boost the small-scale marked statistics of these properties. In this context, the relatively strong luminosity-marked correlation function for $r_p < 1\,\mathrm{Mpc}$ observed in Figure 8.5 suggests that this is the case for satellite luminosities in groups.

In the halo model, the position-dependence of a galaxy mark is modeled with a weighted density profile. Following Sheth (2005), the Fourier transform of the density profile in equations (8.8) and (8.8) then becomes

$$u_{\mathrm{gal}}(k|M) \;\rightarrow\; w(k|M) \;\equiv\; \frac{\int \mathrm{d}r'\, 4\pi r'^2\, w(r'|M)\, \rho_{\mathrm{gal}}(r'|M)\, \sin(kr')/kr'}{\int \mathrm{d}r'\, 4\pi r'^2\, w(r'|M)\, \rho_{\mathrm{gal}}(r'|M)}, \tag{8.11}$$

where $r' \equiv r/r_{\mathrm{vir}}(M)$ and $w(r'|M)$ is our model of the position dependence of the mark.

Hansen et al. (2005) found strong evidence that the luminosity function of satellite galaxies in groups ($N \geq 5$) decreases is one moves from the cluster center to the virial radius and beyond. Based on such evidence, we tested a simple model in which satellite luminosity decreases like $1/(r/r_{\rm vir})$ at the outer halo radii:

$$w_L(r'|M) \propto \frac{1}{r' + r'_s}, \tag{8.12}$$

where r'_s is a scale radius, for which a smaller value means a stronger position dependence. The luminosity-marked projected correlation functions with $r'_s = 1.5$ and 3 are shown in Figure 8.5 (green and magenta curves).

Figure 8.9 shows the measured mean satellite luminosity as a function of projected distance from the central galaxy, relative to the virial radius. There is a weak position-dependence of satellite luminosity in the SDSS groups, although luminosity decreases less rapidly than our model with $r'_s = 1.5$. We have found a similarly weak position-dependence of satellite luminosity as a function of distance from group centroid, and the shape of the position-dependence does not vary with group richness.

For $g-r$ color, in contrast, there appears to be even stronger position-dependence, with redder galaxies often located closer to the halo center (Weinmann et al. 2006a, Martínez & Muriel 2006). We will need to account for this effect when analyzing color-marked correlation functions in groups. However, we must clearly explain the distributions of central and satellite galaxies discussed in the previous section prior to exploring the small-scale environmental dependence of luminosity and color.

8.5 DISCUSSION

The environmental dependence of galaxy formation is due to many large-scale and small-scale environmental effects. We argue that these are the combined effects of the correlations between the mass and formation histories of halos and large-scale environment, on the one hand, and on the other, the local dynamical and physical processes in groups and clusters,

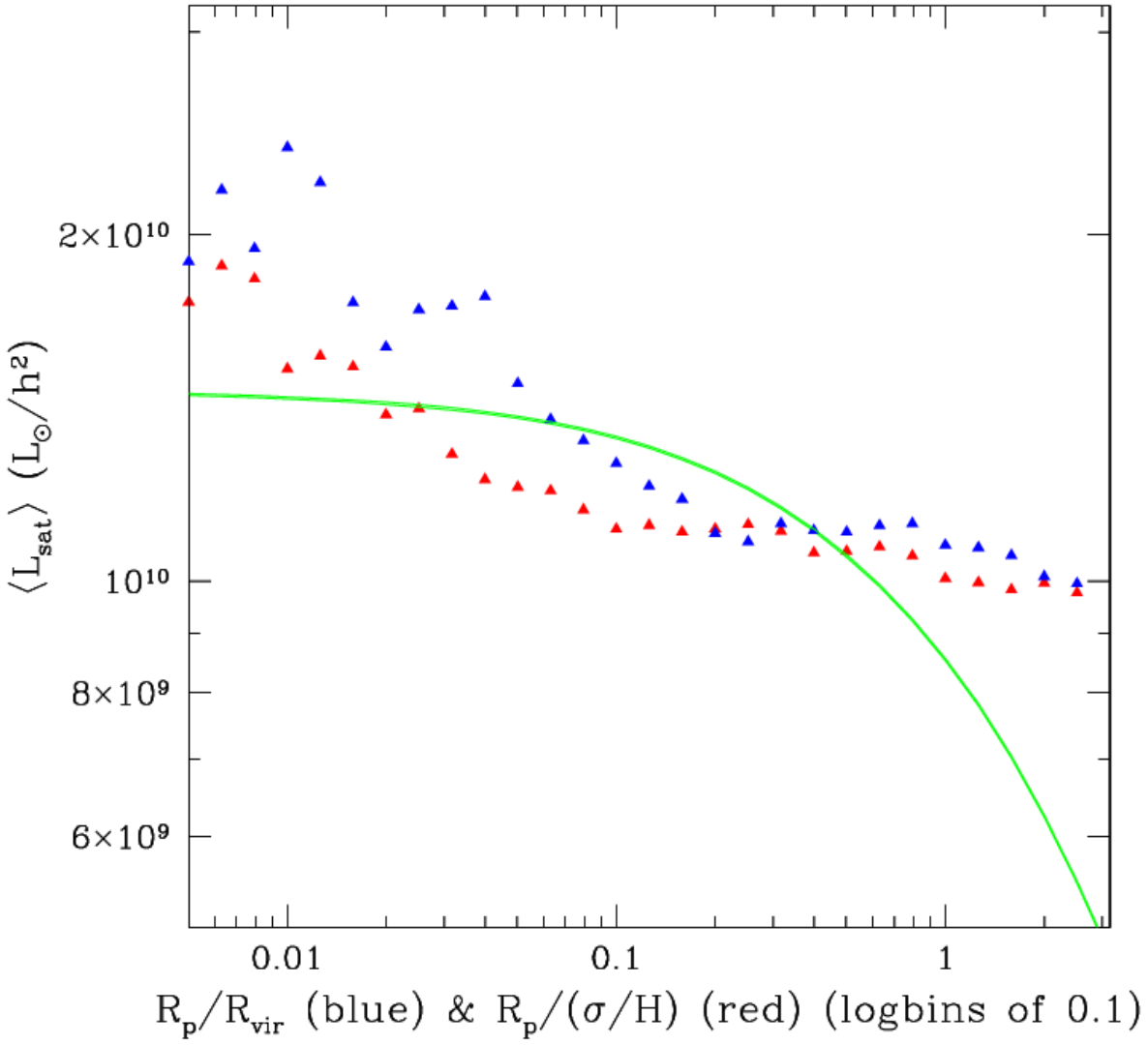

Figure 8.9: Mean satellite galaxy r-band luminosity as a function of projected distance to central galaxy for $M_r < -20$. The distances are normalized by the virial radius (blue) and by the group velocity dispersion (red). Also shown is the analytic estimate with the simple model that satellite luminosity decreases according to $(r/r_{\rm vir} + 1.5)^{-1}$.

such as tidal stripping, dynamical friction, galaxy-galaxy interactions, and the quenching of star formation in accreted satellite galaxies.

The purpose of this chapter has been to explore the environmental dependence of galaxy luminosity in groups and clusters using marked projected two-point correlation functions. This is still a work in progress.

Our challenge is to disentangle the various processes involved that affect central and satellite galaxies. In particular, the clustering of galaxies in groups is reduced by the velocity dispersion of central galaxies and the radial number density distribution of satellite galaxies. We are in the process of modeling these in an observationally-motivated way, and it will be very interesting to investigate how they each affect small-scale galaxy clustering.

In order to study the small-scale correlations between luminosity and environment, we also need to account for the fact that satellite galaxy luminosities vary with cluster-centric distance, with fainter galaxies at outer radii. We are also modeling this effect, which tends to result in a stronger clustering of luminosities at small scales.

The environmental dependence of galaxy color will be the focus of ongoing research. We also look forward to analyzing the clustering of group galaxies and their properties in the Millennium Run, and comparing to our results with the halo model and SDSS groups discussed here.

8.6 ACKNOWLEDGEMENTS

We thank Andreas Berlind for many useful discussions about the SDSS group catalogs.

9.0 LUMINOSITY-MARKED CLUSTERING OF HIGH-REDSHIFT GALAXIES

9.1 INTRODUCTION

In hierarchical models of galaxy formation, the spatial clustering of galaxies as a function of their properties is primarily driven by the clustering of dark matter halos as a function of mass. The environmental dependence of galaxy formation, indicated by the marked clustering of galaxy properties, has been explained as being driven by correlations between the masses and formation histories of halos and the large-scale environment. We have previously shown that galaxy luminosity, color, stellar mass, metallicity, and star formation rate are correlated with environment, such that bright, red, massive, metal-rich, passively star forming galaxies tend to reside in denser regions. Good agreement with halo-model predictions of marked statistics with these properties is strong evidence in favor of the general hierarchical picture.

There has been recent progress in determining the clustering of galaxies at higher redshifts (*e.g.*, Ouchi et al. 2005, Coil et al. 2006b), and some have interpreted this clustering in the halo-model framework, in the context of evolving halo abundances and the halo occupation distribution (*e.g.*, Hamana et al. 2004, Lee et al. 2005, Cooray & Ouchi 2006). By comparing measurements at a range of redshifts, astrophysicists are beginning to explore the evolution of galaxy properties (*e.g.*, Bell et al. 2005, McIntosh et al. 2005) and the evolution of their environmental dependence (*e.g.*, Cooper et al. 2006, Ilbert et al. 2006). The goal of such research is to better understand galaxy evolution processes, and how they are related to halo evolution, merger histories, and the large-scale environment.

Our goal is to contribute to this research with marked statistics analysis at $z \sim 1$.

We are currently examining angular correlation functions and luminosity-marked correlation functions of a variety of galaxy catalogs from the GOODS, GEMS, and COMBO-17 surveys. Our research has just begun, and the purpose of this brief chapter is to describe the types of measurements we are doing and how we are beginning to analyze them.

This chapter is organized as follows. In Section 9.2 we describe our datasets and how we measure angular correlation functions. We also show a few examples of measurements we have recently made. In Section 9.3 we discuss some theoretical issues involved in interpreting marked angular correlation functions with the halo model. We conclude by outlining our ongoing research plans.

9.2 MEASURING ANGULAR CORRELATION FUNCTIONS

We quantify the clustering properties of high-redshift galaxies by measuring the angular two-point correlation function $w(\theta)$. Because of uncertainties in the redshift estimates and the limited survey area, these are more reliable than other clustering measures. As with the real-space and projected correlation functions, $w(\theta)$ decreases with increasing separation, and over a range of angles it can be approximated by a power-law. When the angular correlation function is sufficiently precise, one can discern the 'one-halo term' and 'two-halo term', which are characterized by pairs of galaxies occupying the same halo and pairs in separate halos. Although the redshifts of galaxies in the deeper surveys are often highly uncertain, many galaxy properties, such as luminosity, color and stellar mass, have been reliably determined, allowing us to study the *marked* angular correlation function by weighting by these properties. By probing small-scale and large-scale angular clustering at higher redshifts, we can constrain the evolution of the halo occupation distributions of the galaxies, and with marked angular clustering we hope to constrain the evolution of the environmental dependence of particular galaxy attributes.

As with the real-space and projected correlation functions, we measure $w(\theta)$ with the Landy-Szalay estimator (1993), using large random catalogs with same geometry as the data. In addition, the limited survey size results in a slight underestimate of $w(\theta)$ by a known

constant called the 'integral constraint', which is determined from the size and geometry of the survey and the slope of the galaxy power spectrum.

We define the 'marked' angular correlation function as the ratio of the weighted to the unweighted angular correlation function:

$$M(\theta) \equiv \frac{1 + W(\theta)}{1 + w(\theta)} \tag{9.1}$$

In general, volume-limited catalogs are required to obtain reliable measurements of marked correlation functions.

As an example of marked and unmarked angular correlation functions, we show our measurements for a mock catalog in Figure 9.1. The mock catalog was constructed to have the same real-space correlation function as the SDSS, and it contains galaxies with $M_r < -20.5$. The catalog is a box with sides of length 479 Mpc/h, and we moved it 2 Gpc/h away ($z \sim 0.85$). This is the same $z = 0$ catalog used in Chapter 5; it would be even more useful to have placed a box with $z = 0.85$ particle positions, velocities and halo occupation distribution. Our catalog consists of 256,741 galaxies, and so the uncertainties in the measurement are likely very small, except for the two largest-scale bins.

The angular correlation function clearly exhibits power-law behavior, and its slope is ≈ -0.85. The lower panel shows the r-band luminosity-marked angular correlation function, and the fact that it is above unity at small scales means that the bright galaxy luminosities in the catalog are clustered more than the faint ones. The one-halo and two-halo terms are also visible in the marked correlation function, with the transition occurring at $\theta \sim 0.01$ deg. This translates into a projected separation of $r_p \sim 350\,\mathrm{kpc}/h$, which is a smaller scale than the 1-halo-2-halo transition of the projected correlation function of galaxies at low redshifts (*cf.* Figure 5.7). This measurement is encouraging because it shows that we should be able to detect the environmental dependence of luminosity with marked angular correlation functions.

We show measurements for GOODS catalogs with different i-band magnitude ranges in Figure 9.2. Our samples two samples are: $24 \leq i \leq 26$, $\bar{z} = 0.68$, with 3,266 galaxies; $22 \leq i \leq 24$, $\bar{z} = 0.72$ with 1,098 galaxies. These measurements are much more uncertain than those of the mock catalog because of the small sample size and because the GOODS

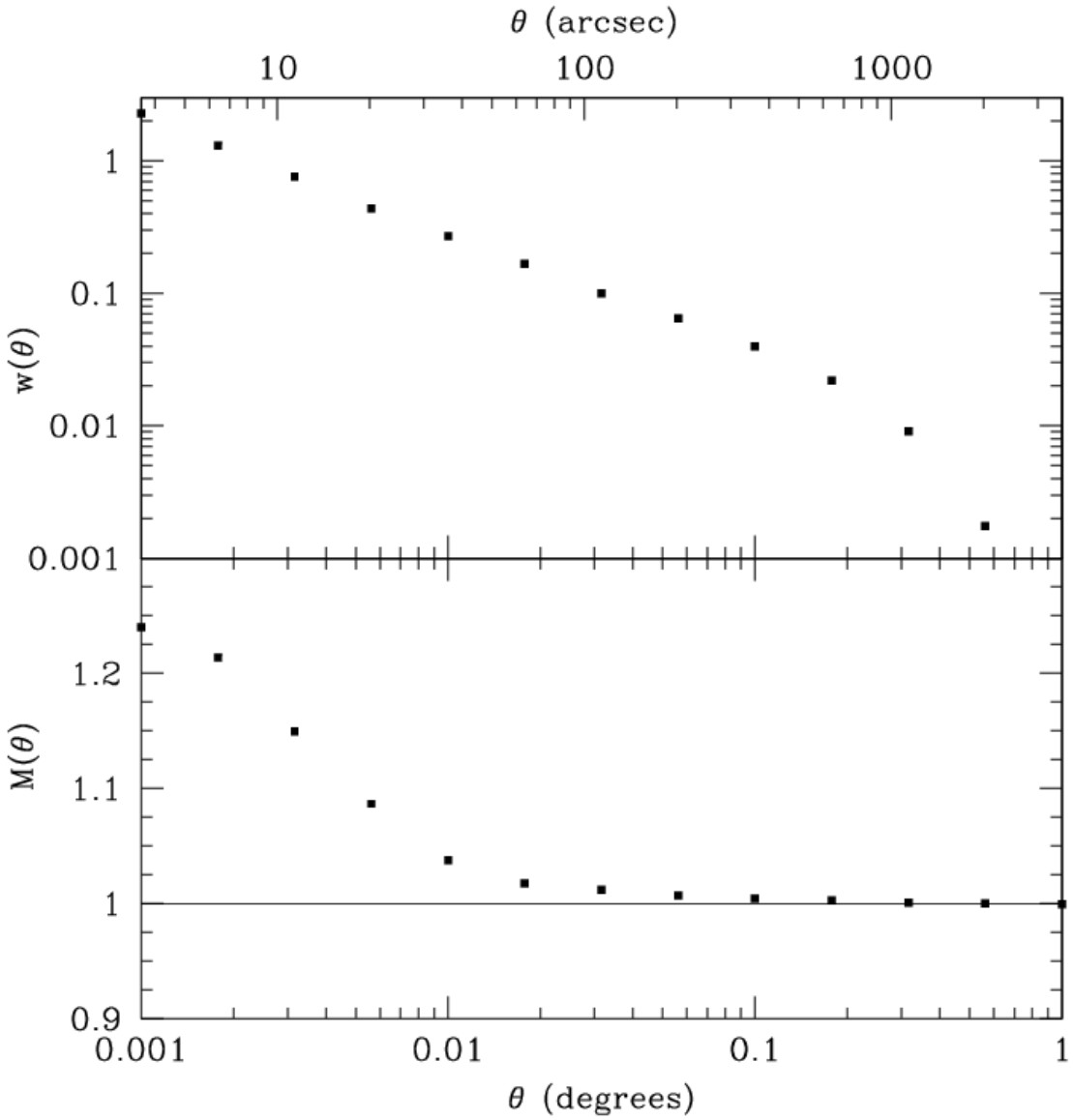

Figure 9.1: Angular correlation function and r-band luminosity-marked angular correlation function for $M_r < -20.5$ mock catalog.

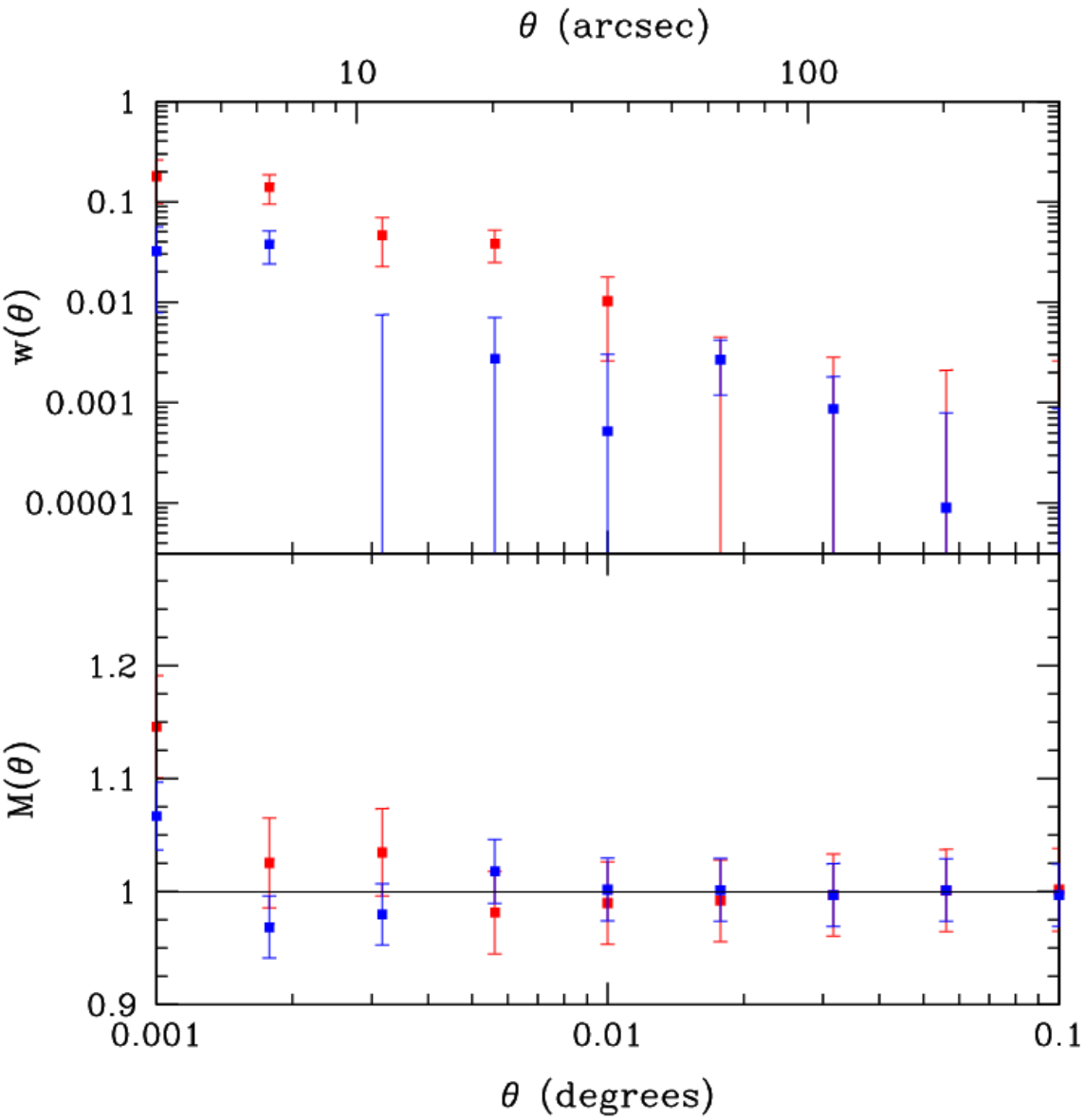

Figure 9.2: GOODS-South angular correlation functions at $z \sim 0.7$. Measurements for $24 \leq i \leq 26$ (blue points) and $22 \leq i \leq 24$ (red points) are shown. Bottom panel shows V-band "luminosity"-marked angular correlation function, with a crude magnitude conversion: "M_V" $= m_V - 5\log_{10}(d_{\rm com}(z)/10{\rm pc}) - 2.5\log_{10}(1+z)$.

survey is relatively small, with a total area coverage of about 0.1 square degree. The marked correlation function in the lower panel was not done with real K-corrected luminosities, so it is not a reliable indicator of the environmental dependence of the luminosities of GOODS galaxies.

However, the amplitude of the angular correlation functions are significantly different, showing that more luminous galaxies are indeed more strongly clustered at $z \sim 0.7$. A more accurate measurement with larger DEEP2 samples was done by Coil et al. (2006b), who detected a statistically significant luminosity dependence of clustering at $z \sim 1$.

We show the evolution of the clustering of high-redshift galaxies in Figure 9.3. Here we use samples from GEMS, in which the redshifts and V-band magnitudes came from COMBO-17. Our samples have the following redshift ranges: $0.30 \leq z \leq 0.55$, with 1,370 galaxies; $0.55 \leq z \leq 0.80$, with 3,319 galaxies; $0.80 \leq z \leq 1.05$, with 2,107 galaxies. All three samples are for $M_V < -18$, and we have also made the cut $R < 24$ because the redshifts are increasingly inaccurate for progressively fainter galaxies.

Angular separations correspond to larger physical separations at higher redshifts in the following way: $\theta = 2\tan(r_p/2d_C(\bar{z})) \approx r_p/d_C$. In order to account for this effect, in Figure 9.3 we have used $\theta \rightarrow \theta\, d_C$ normalized by the comoving distance for $z = 0.7$, which is $\approx 1.75\,\mathrm{Gpc}/h$.

With this correction, even if the uncertainties in the clustering measurements are very small, there does *not* appear to be statistically significant evidence of clustering evolution. The amplitude of $w(\theta)$ increases at lower redshifts only because the angles correspond to shorter distances, and by accounting for this in the figure, the effect disappears. The slope and amplitude of the angular correlation functions both do not appear to change significantly with redshift. This raises the issue of whether clustering is constant in physical or comoving coordinates; because of our $\theta \rightarrow \theta\, d_C$ correction, our result appears to be consistent with no evolution of clustering in comoving coordinates. This suggests that most of the clustering of galaxies observed at low redshifts ($z \sim 0.3$) took place prior to $z \sim 1$. In addition, the strength of the luminosity-marked angular correlations also do not appear to evolve with time over the redshift range, suggesting that the environmental dependence of luminosity was already in place by $z \sim 1$.

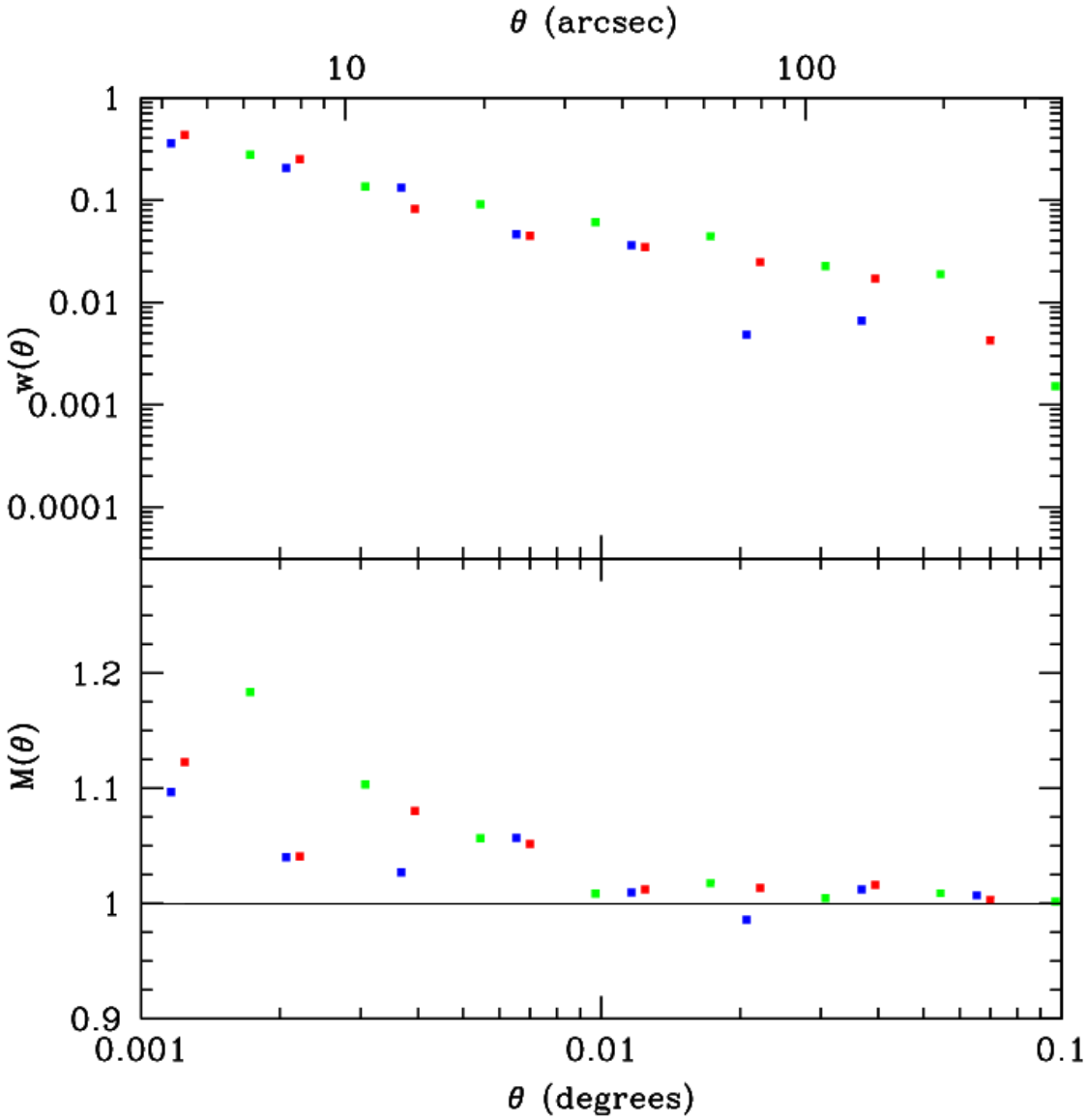

Figure 9.3: Evolution of GEMS/C17 angular correlation function for $M_V < -18$ and $R < 24$: $0.3 \leq z \leq 0.55$ (blue), $0.55 \leq z \leq 0.8$ (green) and $0.8 \leq z \leq 1.05$ (red). Lower panel: V-band luminosity-marked angular correlation functions.

9.3 THEORETICAL ISSUES

In this section we discuss some issues involved in describing the clustering and marked clustering of high-redshift galaxies with the halo model.

The power spectrum and weighted power spectrum at a particular redshift are simply the sums of the one-halo and two-halo terms:

$$P(k|z) \,=\, P_{1h}(k|z) \,+\, P_{2h}(k|z) \qquad \text{and} \qquad W(k|z) \,=\, W_{1h}(k|z) \,+\, W_{2h}(k|z). \tag{9.2}$$

The weighted and unweighted galaxy power spectra are determined by the number density of halos as a function of mass, the halo occupation distribution, the galaxy density profile, the bias factor, and the linear power spectrum (see Section 6.2 and Sheth 2005 for details).

We will compare the halo-model predictions to measurements such as those presented in the previous section. Given the redshift distribution of the galaxy sample to which we are comparing, we can obtain the angular correlation function with the Limber projection (Hamana et al. 2004, Cooray & Ouchi 2006):

$$w(\theta|z) \,=\, \int \mathrm{d}r\, n^2(r) \int \frac{\mathrm{d}k}{2\pi}\, k\, P(k|z)\, J_0(k d_\mathrm{A}\theta) \tag{9.3}$$

and

$$W(\theta|z) \,=\, \int \mathrm{d}r\, n^2(r) \int \frac{\mathrm{d}k}{2\pi}\, k\, W(k|z)\, J_0(k d_\mathrm{A}\theta) \tag{9.4}$$

where $n(r)$ is the normalized radial selection function, $J_0(x)$ is the zeroth-order Bessel function of the first kind, and $d_\mathrm{A}(z)$ is the comoving angular diameter distance.

We also need to average over the redshift distribution for the mean galaxy number density, which the power spectra and weighted power spectra are normalized by:

$$\bar{n}_\mathrm{gal} \,=\, \frac{\int \mathrm{d}z\, [\mathrm{d}V(r)/\mathrm{d}z] n(r)\, \bar{n}_\mathrm{gal}(z)}{\int \mathrm{d}z\, [\mathrm{d}V(r)/\mathrm{d}z] n(r)}, \tag{9.5}$$

where $\mathrm{d}V(r)/\mathrm{d}z$ is the comoving volume element per unit solid angle, and $\bar{n}_\mathrm{gal}(z)$ is the galaxy number density as a function of redshift, given by

$$\bar{n}_\mathrm{gal}(z) \,=\, \int_{M_\mathrm{min}} \mathrm{d}M\, \frac{\mathrm{d}n(M,z)}{\mathrm{d}M}\, \langle N_\mathrm{gal}|M, M_\mathrm{min}, z\rangle. \tag{9.6}$$

Of course, most of the halo model components involved in these calculations evolve significantly with redshift. The evolution of the halo mass function and bias factor are accurately known. The evolving galaxy density distribution, $\rho_{\rm gal}(r|M,z)$, and halo occupation distribution, $\langle N_{\rm gal}|M, M_{\rm min}, z\rangle$, are not. Given a realistic model, the latter can be somewhat constrained by $w(\theta)$ measurements at high redshifts (*e.g.*, Hamana et al. 2004). We will initially assume that the HOD consists of a single central galaxy and a number of 'satellite' galaxies, the mean number of which increases with halo mass as a power-law.

The galaxy density profile $\rho_{\rm gal}(r)$ is likely different from the dark matter density profile (see chapter 8), and may even be significantly non-NFW. However, fairly precise small-scale correlation function measurements will be required to constrain it; considering the current lack of precision in such measurements, assuming an NFW profile will likely be sufficient.

In addition, most galaxy samples are defined with a luminosity cut and knowledge of the corresponding minimum halo mass is also necessary. One can relate luminosity to halo mass by matching the number density of galaxies from the luminosity function with the number density of halos from the halo mass function (Chapter 3). Information about the evolution of the luminosity function up to $z \sim 1$ will be required (Ilbert et al. 2006, Cimatti et al. 2006), and the evolving LF parameters need to be fairly accurately known, as the luminosity-mass relation is highly dependent on them.

Finally, for marked correlation functions, a relation between the mark and halo mass is required, as is knowledge about the dependence of the HOD and galaxy density distribution on the mark. The evolving luminosity-mass relation is used for the luminosity mark, and for the $B-V$ color mark, it is necessary to model the evolving luminosity function in both bands. Given a stellar mass catalog and information about how the mass-to-light ratio evolves (Borch et al. 2006), one can also model stellar mass-marked correlations (Chapter 6).

9.4 DISCUSSION

A variety of recent studies have shown that, not only do galaxy properties evolve with redshift, but the environmental dependence of those properties evolve as well. Astrophysicists

have detected the evolving environmental dependence of luminosity and color (Cooper et al. 2006, Ilbert et al. 2006, Cucciati et al. 2006) and of star formation rate and metallicity (Sheth et al. 2006). The goal of our present work is to explore the environmental dependence of galaxy evolution with marked angular two-point correlation functions, which are sensitive tools for identifying and quantifying correlations between galaxy properties ('marks') and the large-scale environment.

Using datasets from GOODS, GEMS, and COMBO-17, we are currently measuring V-band luminosity-marked angular correlation functions, and we plan to do the same with the $B-V$ color mark. We look forward to the possibility of extending our analysis by investigating the environmental dependence of stellar mass and morphology (with a Sérsic index mark) as well.

The halo model of marked galaxy clustering is based on the assumption that any correlations between halo and galaxy formation and large-scale environment is due to the environmental dependence of halo abundances, such that massive halos tend to be located in denser environments. Therefore, if halo-model predictions of marked statistics are consistent with measurements, it suggests that the observed correlation with environment is driven by the environmental dependence of halo mass. To the extent that the halo model and observations are not consistent, other physical processes may be involved in the environmental correlations. Our goal is to model luminosity-marked angular correlation functions at $0.5 < z < 1$ in order to study the evolution of the environmental dependence of galaxy luminosities. This analysis can be extended to other galaxy marks. Such halo-model analyses require knowledge about the relation between the mark and halo mass, the evolution of the galaxy density distribution and the halo occupation distribution, and their dependence on the mark. It will be an exciting challenge to explore these issues as we attempt to better understand the drivers of galaxy evolution processes.

9.5 ACKNOWLEDGEMENTS

We thank Dan McIntosh for some useful discussions about the GEMS catalogs.

10.0 CONCLUSION

A wide range of studies using a variety of different methods have made it increasingly clear that the clustering of galaxies varies as a function of galaxy properties. Galaxies with bright luminosities, red colors, low star formation rates, high stellar masses, and elliptical morphologies tend to be much more clustered than their fainter, bluer, actively star-forming, less massive, and more disc-like counterparts. Since galaxy properties significantly vary depending on how dense the environments are in which they are located, this strongly suggests that galaxy formation and evolution are correlated with the environment as well.

The primary goal of this thesis has been to use marked clustering statistics to explore the environmental dependence of galaxy formation and evolution. Marked statistics are sensitive tools that are ideal for identifying and quantifying environmental correlations, especially given the vast size of current datasets. We have used a diversity of galaxy 'marks', namely, luminosity, color, stellar mass, metallicity, and star formation rate, with marked two-point correlation functions, and we have shown how they are correlated with the large-scale environment and the scale dependence of these environmental correlations.

We have interpreted our measurements of marked clustering statistics in the framework of the dark matter halo model. One of the major assumptions of the halo model is that only halo abundances are correlated with the large-scale environment. That is, since halos formed around the peaks of the primordial density field, all environmental correlations are due to the fact that massive halos tend to reside in dense regions and less massive halos tend to reside in underdense regions. Since all matter, including all galaxies, is assumed to exist inside halos, the halo model predicts that the environmental dependence of galaxy formation is driven by the environmental dependence of halo formation. In this context, we have modeled the marked correlation functions of galaxy luminosity, color, and stellar mass

with the halo model, and we have obtained very interesting results.

Our main conclusions of this thesis are the following:

- Luminous galaxies are more strongly clustered than faint galaxies, especially on scales smaller than the size of massive halos and especially for luminosities in the redder bands. By modeling the luminosity-marked correlation function in the halo model, while analyzing the luminosities of central and satellite galaxies separately, we have shown that the environmental dependence of luminosity in the SDSS is driven by the environmental dependence of halo mass. Our preliminary analysis of color-marked correlations suggests that this is also the case with the environmental dependence of color.
- Massive galaxies are more strongly clustered than less massive galaxies, and the environmental dependence is even stronger than that of luminosity. Our halo-model analysis provides strong evidence that the environmental dependence of stellar mass in the SDSS is also driven by the correlations between halo mass and large-scale environment. This also appears to be the case for metallicity, but more precise measurements with larger datasets are required.
- Our luminosity-marked correlation analysis of the Millennium Run Simulation shows that the environmental dependence of luminosity is weaker than predicted by the halo model and observed in the SDSS when we directly compare them. This is due to multiple factors, including the satellite galaxy fractions in the simulations.
- Our stellar mass-marked correlation analysis of the Millennium Run Simulation shows that the environmental dependence of stellar mass is stronger than predicted by the halo model and observed in the SDSS. This may be partly due to the IMF assumed in the simulation, but it is probably also tied to the modeling of the shutdown of star formation by AGN feedback.
- Star formation rate-marked correlation analysis in the SDSS and Millennium Run show that passively star formation galaxies are more strongly clustered than actively star forming galaxies. This environmental dependence increases at small scales, and our measurements with fainter galaxies suggest that the environmental dependence is too strong in the simulation on scales of $300\,\mathrm{kpc}/h < r_p < 3\,\mathrm{Mpc}/h$. In addition, at very small scales ($r_p < 200\,\mathrm{kpc}/h$), *actively* star forming galaxies are observed to be strongly clustered in

dense regions in the SDSS, possibly due to galaxy-galaxy interactions stimulating star formation. There is absolutely no signal of such an effect in the Millennium Simulation.

- Our test of the effect of halo assembly bias and formation time on galaxy clustering suggests that it is too weak to significantly affect the two-point correlations of galaxies with a wide range of luminosities, corresponding to a wide range of halo masses. More limited and fainter galaxy samples may be more significantly affected. As measurements become more precise, however, it will probably become necessary to account for the effects of assembly bias.

- Our luminosity-marked correlation analysis of SDSS galaxies in groups and clusters suggests that the large-scale environmental dependence of luminosity is driven by that of halo mass, while the small-scale environmental dependence appears to be affected by multiple group-specific processes to varying degrees. Firstly, 'central' galaxies often have a significant velocity bias with respect to the satellite galaxies. Secondly, while dark matter density profiles are tied to the matter accretion rate, satellite galaxies have a much more diffuse density distribution. Thirdly, more luminous satellites tend to be located near the halo center while fainter satellites are more often located at outer radii.

Finally, there is more exciting potential research along these lines We are currently studying the evolution of the environmental dependence of luminosity up to redshifts of $z \sim 1$, and we hope to extend this to other galaxy marks, including color, stellar mass, and Sérsic index. We look forward to using the halo model to study the color- and star formation-marked correlation functions. One can also examine the environmental dependence of AGN activity with [OIII] luminosity-marked correlation functions, or with marked cross correlations with other galaxy properties. There are other useful marked statistics as well, such as the mark variance and covariance. We look forward to continuing to use marked statistics with the halo model to explore the environmental dependence of galaxy formation and evolution.

10.1 ACKNOWLEDGEMENTS

Chapters 4-9: We thank the Pittsburgh Computational Astrostatistics group (PiCA) for the NPT code which was used to measure the correlation functions presented.

Chapters 4-6 and 8: Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, Cambridge University, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPA), the Max-Planck-Institute for Astrophysics (MPIA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

Chapters 6 and 7: The Millennium Run simulation used was carried out by the Virgo Supercomputing Consortium at the Computing Centre of the Max-Planck Society in Garching. The semi-analytic galaxy catalogue is publicly available at
http://www.mpa-garching.mpg.de/galform/agnpaper.

Chapter 9: Support for the GEMS project was provided by NASA through grant number GO-9500 from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc. for NASA, under contract NAS5-26555.

BIBLIOGRAPHY

Abbas U., Sheth R. K., 2005, MNRAS, 364, 1327

Abbas U., Sheth R. K., 2006, astro-ph/0601407

Adelman-McCarthy J. K. et al., 2006, ApJS, 162, 38

Allgood B., Flores R. A., Primack J. R., Kravtsov A. V., Wechsler R. H., Faltenbacher A., Bullock J. S., 2006, MNRAS, 367, 1781

Alonso M. S., Lambas D. G., Tissera P. B., Coldwell G., 2006, MNRAS, 367, 1029

Andreon S., 2004, A&A, 416, 865

Ascasibar Y., Yepes G., Gottlöber S., Müller V., 2004, MNRAS, 352, 1109

Austin C. G., Williams L. L. R., Barnes E. I., Babul A., Dalcanton J. J., 2005, ApJ, 634, 756

Avila-Reese V., Firmani C., Klypin A., Kravtsov A. V., 1999, MNRAS, 310, 527

Avila-Reese V., Colín P., Gottlöber S., Firmani C., Maulbetsch C., 2005, ApJ, 634, 51

Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Ž., Lupton R. H., Nichol R. C., Szalay A. S., 2004, ApJ, 600, 681

Baldry I. K. et al., 2005, MNRAS, 358, 441

Bardeen J. M., Bond J. R., Kaiser N., Szalay A. A., 1986, ApJ, 304, 15

Barnes E. I., Williams L. L. R., Babul A., Dalcanton J. J., 2005, astro-ph/0510332

Beisbart C., Kerscher M., 2000, ApJ, 545, 6

Beisbart C., Kerscher M., Mecke K., 2002, physics/0201069

Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, ApJS, 149, 289

Bell E. F. et al., 2005, ApJ, 625, 23

Benson A. J., Bower R. G., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2003, ApJ, 599, 38

Berlind A. A. et al., 2003, ApJ, 593, 1

Berlind A. A., Blanton M. R., Hogg D. W., Weinberg D. H., Davé R., Eisenstein D. J., Katz N., 2005, ApJ, 629, 625

Berlind A. A. et al., 2006, astro-ph/0601346

Bernardi M. et al., 2003, AJ, 125, 1866

Bernardi M., Nichol R. C., Sheth R. K., Miller C. J., Brinkmann J., 2005, astro-ph/0509360

Bernardi M., Hyde J. B., Sheth R. K., Miller C. J., Nichol R. C., 2006, astro-ph/0607117

Bertschinger E., 1985, ApJS, 58, 39

Best P. N., Kauffmann G., Heckman T. M., Brinchmann J., Charlot S., Ivezić Ž., White S. D. M., 2005, MNRAS, 362, 25

Birnboim Y., Dekel A., 2003, MNRAS, 345, 349

Blanton, M. R. et al., 2001, AJ, 121, 2358

Blanton, M. R. et al., 2003a, ApJ, 592, 819

Blanton, M. R. et al., 2003b, AJ, 125, 2348

Blanton M. R., Eisenstein D., Hogg D. W., Schlegel D. J., Brinkmann J., 2005, 629, 143

Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440

Borch A. et al., 2006, astro-ph/0604405

Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2005, astro-ph/0511338

Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS, 351, 1151

Bryan G. L., Norman M. L., 1998, ApJ, 495, 80

Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001a, MNRAS, 321, 559

Bullock J. S., Dekel A., Kolatt T. S., Kravtsov A. V., Klypin A. A., Porciani C., Primack J. R., 2001b, ApJ, 555, 240

Cattaneo A., Dekel A., Devriendt J., Guiderdoni B., Blaizot J., 2006, astro-ph/0601295

Cimatti A., Daddi E., Renzini A., 2006, astro-ph/0605353

Coil A. L., Newman J. A., Cooper M. C., Davis M., Faber S. M., Koo D. C., Willmer C. N. A., 2006, ApJ, 644, 671

Collister A. A., Lahav O., 2005, MNRAS, 361, 415

Connolly A. J., Szalay A. S., Brunner R. J., 1998, ApJ, 499, L125

Conroy C., Wechsler R. H., Kravtsov A. V., 2005, astro-ph/0512234

Cooper M. C. et al., 2006, astro-ph/0603177

Cooray A., Ouchi M., 2006, MNRAS, 369, 1869

Cooray A., Sheth R., 2002, Phys. Rep., 372, 1

Cooray A., 2005a, MNRAS, 363, 337

Cooray A., 2005b, MNRAS, 364, 303

Cooray A., 2006, MNRAS, 365, 842

Croton D. J. et al., 2006, MNRAS, 365, 11

Croton D. J., Gao L., White S. D. M., 2006, astro-ph/0605636

Cucciati O. et al., 2006, astro-ph/0603202

Davis M., Peebles P. J. E., 1983, ApJ, 267, 465

De Lucia G., Springel V., White S. D. M., Croton D., Kauffmann G., 2006, MNRAS, 366, 499

De Lucia G., Blaizot J., 2006, astro-ph/0606519

Dekel A., Birnboim Y., 2006, MNRAS, 368, 2

Diaferio A., Kauffmann G., Balogh M. L., White S. D. M., Schade D., Ellingson E., 2001, MNRAS, 323, 999

Diemand J., Moore B., Stadel J., 2004, MNRAS, 353, 624

Efstathiou G., Bond J. R., White S. D. M., 1992, MNRAS, 258, 1

Eke V. R., Navarro J. F., Steinmetz M., 2001, ApJ, 554, 114

Fall S. M., Efstathiou G., 1980, MNRAS, 193, 189

Faltenbacher A., Gottlöber S., Kerscher M., Müller V., 2002, A&A, 395, 1

Feulner G., Hopp U., Botzler C. S., 2006, astro-ph/0603616

Fillmore, J. A., Goldreich P., 1984, ApJ, 281, 1

Fisher K. B., 1995, ApJ, 448, 494

Gallazzi A., Charlot S, Brinchmann J., White S. D. M., Tremonti C., 2005, MNRAS, 362, 41

Gallazzi A., Charlot S., Brinchmann J., White S. D. M., 2006, astro-ph/0605300

Gao L., Springel V., White S. D. M., 2005, MNRAS, 363, L66

Gnedin O. Y., Kravtsov A. V., Klypin A. A., Nagai D., 2004, ApJ, 616, 16

Gott J. R. III, 1975, ApJ, 201, 296

Gottlöber S., Kerscher M., Kravtsov A. V., Faltenbacher A., Klypin A., Müller V., 2002, A&A, 387, 778

Gunn J. E., 1977, ApJ, 218, 592

Hamana T., Ouchi M., Shimasaku K., Kayo I., Suto Y., 2004, MNRAS, 347, 813

Hamilton A. J. S., 1992, ApJ, 385, L5

Hansen S. M., McKay T. A., Wechsler R. H., Annis J., Sheldon E. S., Kimball A., 2005, ApJ, 633, 122

Harker G., Cole S., Helly J., Frenk C., Jenkins A., 2006, MNRAS, 367, 1039

Hernquist L., 1990, ApJ, 356, 359

Hiotelis N., 2002, A&A, 382, 84

Hoffman Y., Shaham J., 1985, ApJ, 297, 16

Hogg D. W., Masjedi M., Berlind A. A., Blanton M. R., Quintero A. D., Brinkmann J., 2006, astro-ph/0606599

Hopkins A. et al., 2003, ApJ, 599, 971

Ilbert O. et al., 2006, astro-ph/0602329

Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, MNRAS, 321, 372

Jing Y. P, Suto Y, 2002, ApJ, 574, 538

Jing Y. P., Börner G., 2004, ApJ, 617, 782

Kaiser N., 1987, MNRAS, 227, 1

Kasun S. F., Evrard A. E., 2005, ApJ, 629, 781

Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999, MNRAS, 303, 188

Kauffmann G. et al., 2003, MNRAS, 341, 33

Kauffmann G., White S. D. M., Heckman T. M., Ménard B., Brinchmann J., Charlot S., Tremonti C., Brinkmann J., 2004, MNRAS, 353, 713

Kauffmann G., Heckman T. M., De Lucia G., Brinchmann J., Charlot S., Tremonti C., White S. D. M., Brinkmann J., 2006, MNRAS, 367, 1394

Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2

Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, ApJ, 609, 35

Kuehn F., Ryden B. S., 2005, ApJ, 634, 1032

Lacey C., Cole S., 1993, MNRAS, 262, 627

Landy S. D., Szalay A. S., 1993, ApJ, 412, 64

Lee K., Giavelisco M., Gnedin O. Y., Somerville R. S., Ferguson H. C., Dickinson M., Ouchi M., 2005, astro-ph/0508090

Lemson G., Kauffmann G., 1999, MNRAS, 302, 111

Li C., Kauffmann G., Jing Y. P., White S. D. M., Börner G., Cheng F. Z., 2006, MNRAS, 368, 21

Lin H. et al., 1999, ApJ, 518, 533

Lin Y. T., Mohr J. J., Stanford S. A., 2004, ApJ, 610, 745

Lin Y. T., Mohr J. J., 2004, ApJ, 617, 879

Łokas E. L., 2000, MNRAS, 311, 423

Loveday J., 2004, MNRAS, 347, 601

Lu Y., Mo H. J., Katz N., Weinberg M. D., 2006, MNRAS, 368, 1931

MacMillan J. D., Widrow L. M., Henriksen R. N., 2006, astro-ph/0604418

Mandelbaum R., Seljak U., Kauffmann G., Hirata C. M., Brinkmann J., 2006, MNRAS, 368, 715

Manrique A., Raig A., Salvador-Solé E., Sanchis T., Solanes J., 2003, ApJ, 593, 26

Martínez H. J., Muriel H., 2006, astro-ph/0605264

Mateus A., Sodré L., Fernandes R. C., Stasińska G., 2006, astro-ph/0604063

Maulbetsch C., Avila-Reese V., Colín P., Gottlöber S., Khalatyan A., Steinmetz M., 2006, astro-ph/0606360

McIntosh et al., 2005, ApJ, 632, 191

Menci N., Fontana A., Giallongo E., Salimbeni S., 2005, ApJ, 632, 49

Mo H. J., White S. D. M., 1996, MNRAS, 282, 347

Mo H. J., White S. D. M., 2002, MNRAS, 336, 112

Mo H. J., Yang X., van den Bosch F. C., Jing Y. P., 2004, MNRAS, 349, 205

Moore B., Quinn T., Governato F., Stadel J., Lake G., 1999, MNRAS, 310, 1147

Moustakas J., Kennicutt R. C., Tremonti C. A., 2006, ApJ, 642, 775

Nagai D., Kravtsov A. V., 2005, ApJ, 618, 557

Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563

Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493

Navarro J. F. et al., 2004, MNRAS, 349, 1039

Neistein E., van den Bosch F. C., Dekel A., 2006, astro-ph/0605045

Norberg P., Cole S., Baugh C. M., et al., 2002, MNRAS, 336, 907

Nusser A., 2001, MNRAS, 325, 1397

Nusser A., Sheth R. K., 1999, MNRAS, 303, 685

Ouchi M. et al., 2005, ApJ, 635, L117

Peacock J. A., Dodds S. J., 1994, MNRAS, 267, 1020

Peebles P. J. E., 1969, ApJ, 155, 393

Peebles P. J. E., 1980, The Large-Scale Structure of the Universe. Princeton Univ. Press, Princeton, NJ

Press W. H., Schechter P., 1974, ApJ, 187, 425

Reed D. S., Governato F., Quinn T., Stadel J., Lake G., 2006, astro-ph/0602003

Salvador-Solé E., Manrique A., Solanes J., 2005, MNRAS, 358, 901

Sánchez A. G., Baugh C. M., Percival W. J., Peacock J. A., Padilla N. D., Cole S., Frenk C. S., Norberg P., 2006, MNRAS, 366, 189

Scherrer R. J., Bertschinger E., 1991, ApJ, 381, 349

Scoccimarro R., 2004, PhysRevD, 70, 083007

Scoccimarro R., Sheth R. K., Hui L., Jain B., 2001, ApJ, 546, 20

Scoccimarro R., Sheth R. K., 2002, MNRAS, 329, 629

Scranton R. et al., 2002, ApJ, 579, 48

Scranton R., 2003, MNRAS, 339, 410

Seljak U., 2001, MNRAS, 325, 1359

Seljak U., Warren M. S., 2004, MNRAS, 355, 129

Sellwood J. A., Kosowsky A., 2002, ASPC, 273, 243

Sellwood J. A., McGaugh S. S., 2005, ApJ, 634, 70

Shankar F., Lapi A., Salucci P., De Zotti G., Danese L., 2006, ApJ, 643, 14

Sheth R. K., 1996, MNRAS, 279, 1310

Sheth R. K., 2005, MNRAS, 364, 796

Sheth R. K., Abbas U., Skibba R. A., 2004, in Diaferio A., ed., Proc. IAU Coll. 195, Outskirts of Galaxy Clusters: Intense Life in the Suburbs, CUP, Cambridge, p. 349

Sheth R. K., Connolly A. J., Skibba R., 2005, astro-ph/0511773

Sheth R. K., Diaferio A., 2001, MNRAS, 322, 901

Sheth R. K., Jain B., 1997, MNRAS, 285, 231

Sheth R. K., Jimenez R., Panter B., Heavens A. F., 2006, astro-ph/0604581

Sheth R. K., Lemson G., 1999, MNRAS, 304, 767

Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1

Sheth R. K., Tormen G., 1999, MNRAS, 308, 119

Sheth R. K., Tormen G., 2002, MNRAS, 329, 61

Sheth R. K., Tormen G., 2004, MNRAS, 350, 1385

Sheth R. K., Hui L., Diaferio A., Scoccimarro R., 2001a, MNRAS, 325, 1288

Sheth R. K., Diaferio A., Hui L., Scoccimarro R., 2001b, MNRAS, 326, 463

Sikivie P., Tkachev I. I., Wang Y., 1997, PhRvD, 56, 1863

Skibba R. A., Sheth R. K., Connolly A. J., Scranton R., 2006, MNRAS, 369, 68

Spergel D. N. et al., 2003, ApJS, 148, 175

Spergel D. N. et al., 2006, astro-ph/0603449

Springel V. et al., 2005, Nature, 435, 629

Stoyan D., Stoyan H., 1994, Fractals, Random Shapes, and Point Fields. Wiley, Chichester

Tasitsiomi A., Kravtsov A. V., Gottlöber S., Klypin A. A., 2004, ApJ, 607, 125

Tinker J. L., 2006, astro-ph/0604217

Tinker J. L., Weinberg D. H., Zheng Z., Zehavi I., 2005, ApJ, 631, 41

Tinker J. L., Weinberg D. H., Zheng Z., 2006, MNRAS, 368, 85

Tremonti C. A. et al., 2004, ApJ, 613, 898

Vale A., Ostriker J. P., 2004, MNRAS, 353, 189

Vale A., Ostriker J. P., 2005, astro-ph/0511816

van den Bosch F. C., Yang X., Mo H. J., Norberg P., 2005a, MNRAS, 356, 1233

van den Bosch F. C., Weinmann S. M., Yang X., Mo H. J., Li C., Jing Y. P., 2005b, MNRAS, 361, 1203

Vitvitska M., Klypin A. A., Kravtsov A. V., Wechsler R. H., Primack J. R., Bullock J. S., 2002, ApJ, 581, 799

Wang L., Li C., Kauffmann G., De Lucia G., 2006, astro-ph/0603546

Wechsler R. H., Bullock J. R., Primack J. R., Kravtsov A. V., Dekel A., 2002, ApJ, 568, 52

Wechsler R. H., Zentner A. R., Bullock J. R., Kravtsov A. V., 2005, astro-ph/0512416

Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., 2006, MNRAS, 366, 2

Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., Croton D. J., Moore B., 2006, astro-ph/0606458

Wetzel A. R., Cohn J. D., White M., Holz D. E., Warren M. S., 2006, astro-ph/0606699

White M., 2001, MNRAS, 321, 1

White S. D. M., Frenk C. S., 1991, ApJ, 379, 52

Yang X., Mo H. J., van den Bosch F. C., 2003, MNRAS, 339, 1057

Yang X., Mo H. J., van den Bosch F. C., Jing Y. P., 2005a, MNRAS, 357, 608

Yang X., Mo H. J., Jing Y. P., van den Bosch F. C., 2005b, MNRAS, 358, 217

Yang X., Mo H. J., van den Bosch F. C., Weinmann S. M., Li C., Jing Y. P., 2005c, MNRAS, 362, 711

Yee H. K. C., López-Cruz O., 1999, AJ, 117, 1985

Yee H. K. C., Ellingson E., 2003, ApJ, 585, 215

York D. G. et al., 2000, AJ, 120, 1579

Yoshida N., Sheth R. K., Diaferio A., 2001, MNRAS, 328, 669

Zandivarez A., Martínez H. J., Merchán M. E., 2006, astro-ph/0602405

Zaroubi S., Hoffman Y., 1993, ApJ, 416, 410

Zehavi I. et al., 2004, ApJ, 608, 16

Zehavi I. et al., 2005, ApJ, 630, 1

Zentner A. R., Berlind A. A., Bullock J. R., Kravtsov A. V., Wechsler R. H., 2005, ApJ, 624, 505

Zhao D. H., Mo H. J., Jing Y. P., Börner G., 2003, MNRAS, 339, 12

Zheng Z. et al., 2005, ApJ, 633, 791

Zhu G., Zheng Z., Lin W. P., Jing Y. P., Kang X., Gao L., 2006, ApJ, 639, L5