Applications of the Halo Model to Large Scale Structure Measurements of the Luminous Red Galaxies

Beth Ann Reid

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Physics
Adviser: David N. Spergel

November, 2008
Abstract

The power spectrum of density fluctuations in the evolved universe provides constraints on cosmological parameters that are complementary to cosmic microwave background and other astronomical probes. The Sloan Digital Sky Survey (SDSS) Luminous Red Galaxy (LRG) sample probes a volume of \( \sim 3 \text{ (Gpc)}^3 \), and systematic errors in modeling the nonlinearities limit our ability to extract information on the shape of the linear power spectrum. There are three main effects that distort the observed power spectrum from the linear power spectrum: nonlinear gravitational evolution, redshift space distortions, and a nonlinear relation between the galaxy density field and the underlying matter density field. In this thesis we introduce a new method to mitigate the latter two distortions and rely on carefully tuned \( N \)-body simulations to model the first.

In Chapter 2 we present the technique ‘Counts-in-Cylinders’ (CiC) and use it to measure the multiplicity function of groups of LRGs in SDSS. We use the Halo Occupation Distribution description of the galaxy-matter mapping and \( N \)-body simulations to connect this observation with constraints on the distribution of LRGs in dark matter halos. In Chapter 3 we study the effects of resolution on statistics relating to both the large and small scale distributions and motions of matter and dark matter halos. We combine these results to produce a large set of high quality mock LRG catalogs that reproduce the higher order statistics in the density field probed by the CiC technique. Using these catalogs we present a detailed analysis of the method used in Tegmark et al. (2006) to estimate the LRG power spectrum, and find that the large nonlinear correction necessary for their analysis is
degenerate with changes in the linear spectrum we wish to constrain. We show that the CiC
group-finding method in Chapter 2 can be used to reconstruct the underlying halo density
field. The power spectrum of this field has only percent-level deviations from the underly-
ing matter power spectrum, and will therefore provided tighter constraints on cosmological
parameters. Techniques presented in this thesis will be useful for final analysis of the SDSS
LRGs and upcoming surveys probing much larger volumes.
First and foremost I thank David Spergel for taking on a wandering graduate student and helping her find her scientific home. It has been a pleasure and honor learning from you. I have also had a wonderful community of cosmologists to interact with in and around Princeton: Joe Fowler, Ed Groth, Jim Gunn, Jack Hughes, Raul Jimenez, David Nice, Jerry Ostriker, Lyman Page, Jim Peebles, Suzanne Staggs, and Licia Verde. From the Sloan Digital Sky Survey Large Scale Structure working group I thank Daniel Eisenstein, Will Percival, and David Weinberg. Many thanks to Paul Bode for all his time, effort, and patience with the simulations in Chapter 4. Thanks also to Annika, Chris, Fiona, Janice, Judy, Katie, Eric S., Eric D., Mihail, and David for being wonderful colleagues and friends during my time at Princeton.

I am indebted to Diane Leighty and the entire faculty of Chesterfield County Mathematics and Science High School for a challenging and delightful four years. I thank Uwe Täuber, Beate Schmittmann, Royce Zia, Lay Nam Chang, Peter Haskell, Thomas Quinn, and the faculty in the Virginia Tech Physics Department for introducing me to the wonders of modern physics and the joys of research. Thanks also to my family for their unending love and support for all my ventures. Thanks to Alma, Jen, Marc, Scott, Meghan, Danica, Madhuri, Pam, Tess, and Geek Week for keeping me sane for the last five years. Finally, many thanks and love to my husband Burke for braving life in the New Jersey suburbs. You are my unending source of love, comfort, hilarity, and gourmet desserts, and I am your ticket to Spain.
To my parents
Contents

Abstract iii

Acknowledgements v

Contents vii

List of Figures x

List of Tables xxiii

1 Introduction 1

1.1 Cosmological Model ............................................. 3
    1.1.1 The Homogeneous Universe ............................. 3
    1.1.2 Density Perturbations in Linear Theory ................... 4

1.2 Nonlinearity in the Galaxy Power Spectrum .................. 6
    1.2.1 Nonlinear Growth of Structure .......................... 6
    1.2.2 Redshift Space Distortions .............................. 7
    1.2.3 The Galaxy-Dark Matter Connection ....................... 8
    1.2.4 Solution: Reconstructing the Halo Density Field .. 10

1.3 SDSS LRG galaxy sample .................................. 10

1.4 Outline of the Thesis .................................. 11

2 Constraining the LRG Halo Occupation Distribution using Counts-in-
Cylinders

2.1 Introduction ......................................................... 13
2.2 Methodology ......................................................... 16
  2.2.1 Overview of the Method ........................................ 16
  2.2.2 Data ............................................................. 17
  2.2.3 Simulations and Halo Catalogs ................................. 26
  2.2.4 HOD model ...................................................... 26
  2.2.5 Counts-In-Cylinders Technique ............................... 29
  2.2.6 Maximum Likelihood Parameter Estimation .................. 30
2.3 Results ............................................................. 32
  2.3.1 SDSS LRG $N_{CiC}$ ............................................. 32
  2.3.2 CiC offset $\Delta g(n_{sat})$ and Group-Finding Accuracy .... 33
  2.3.3 HOD Constraints ............................................... 34
  2.3.4 Projected Correlation Function $w_p(r_p)$ .................... 42
  2.3.5 Comparison with Results using FoF Catalogs and Previous Works 44
2.4 Assessment of the CiC method ................................... 47
2.5 Conclusions ......................................................... 48
2.6 Acknowledgments .................................................... 49

3 Optimizing Dark Matter Simulations for LRG Mock Catalogs: Resolution Study

  3.1 Rationale .......................................................... 51
  3.2 Simulations ......................................................... 53
  3.3 Mass Function ...................................................... 55
  3.4 Comparing two-point statistics .................................. 59
    3.4.1 Quantifying the level of systematic bias due to mass resolution ... 61
  3.5 Two Point Statistics Results ..................................... 63
    3.5.1 Comparison of Dark Matter Power Spectra .................. 63
    3.5.2 Halo Power Spectra ........................................... 64
List of Figures

2.1  The mean $c_\parallel - z$ relation and bands including 95% of objects in each $\Delta z = 0.01$ bin. $c_\parallel$ (Eqn. 2.1) is used as a redshift indicator for objects targeted as LRGs but lacking spectra. .................................................. 20

2.2  Contours for $\Delta \ln L = \{-1.15, -3.09, -5.9\}$ in the $M_1$ vs. $M_{cut}$ plane after marginalizing over $\alpha$. For a $\chi^2$ distribution, these contours would enclose 1, 2, and 3$\sigma$ confidence regions. The cross indicates the maximum likelihood parameter values. .................................................. 36

2.3  Same as Figure 2.2, but for $\alpha$ vs. $M_{cut}$. $M_{cut}/(10^{14} M_\odot) + \alpha$ is tightly constrained, while $M_{cut}/(10^{14} M_\odot) - \alpha$ is only weakly constrained. ....... 37

2.4  Same as Figure 2.2, but for $\alpha$ vs. $M_1$. ................................. 38

2.5  The dashed curve ranging from 0 to 1 shows the $N_{cen}(M)$ term for the maximum likelihood HOD; it should vary only slightly with HOD since the satellite fraction is well constrained by our model, and we hold $\sigma_{log M}$ fixed. The solid curve shows $\langle N_{sat}(M) N_{cen}(M) \rangle$, with the error bars computed by Eqn. 2.20. 39

2.6  The probability $P(n_{sat} = 1|M)$ (long-short dashed curve), $P(n_{sat} = 2|M)$ (long dashed curve), $P(n_{sat} = 3|M)$ (short dashed curve), $P(n_{sat} = 4|M)$ (solid curve) vs. halo mass $M/M_{sun}$. Error bars are computed by Eqn. 2.20. Though $P(n_{sat}|M)$ are evaluated at identical values of $M$, points are slightly staggered for clarity. ................................. 40
2.7 The total number density of satellite galaxies occupying a halo of mass $M$ per $\log_{10} M$ (thick solid upper curve). The long-short dashed curve shows the number density of satellites occupying halos with $n_{\text{sat}} = 1$, long dashed curve for satellites in halos with $n_{\text{sat}} = 2$, short dashed curve for $n_{\text{sat}} = 3$, and solid curve for $n_{\text{sat}} = 4$. Error bars are computed by Eqn. 2.20.

2.8 The projected correlation function $w_p(r_p)$ vs. projected separation $r_p$, both in Mpc$/h$. The points with error bars taken from Masjedi et al. (2006). The solid curve is an average over 20 mock catalogs at the maximum likelihood HOD parameters presented in the text for $\sigma_{\log M} = 0.7$. The Masjedi et al. (2006) data was used in our analysis only to optimize the value of $\sigma_{\log M}$, which sets the large scale clustering. The striking agreement shows that using SO halo catalogs we can simultaneously reproduce the CiC group multiplicity function and the clustering on scales $\sim 1 - 20$ Mpc$/h$. For comparison, we produce a catalog with $\sigma_{\log M} = 0.2$ (long dashed curve) and $\sigma_{\log M} = 1.3$ (short dashed curve). Both catalogs match the observed clustering at $r_p \lesssim 0.8$ Mpc$/h$ where we have CiC constraints, but disagree with the observations on larger scales.

2.9 The projected correlation function $w_p(r_p)$ vs. projected separation $r_p$, both in Mpc$/h$. The points with error bars from taken from Masjedi et al. (2006). The solid curve is an average over 20 mock catalogs at the maximum likelihood HOD parameters presented in the text for $\sigma_{\log M} = 0.7$ and using the SO halo catalog. The short dashed curve shows $w_p(r_p)$ averaged over 5 mock catalogs generated by populating an FoF halo catalog. The FoF halo catalog has a severe suppression of two-halo pairs at separations of $\sim 1$ Mpc$/h$, equal to the half of the total number of pairs in the one-halo regime, $r_p \lesssim 0.8$ Mpc$/h$.

3.1 The expected number of halos in a simulation with $L_{\text{box}} = 558$ Mpc$/h$ in logarithmic mass bins ($\Delta \log_{10} M = 0.2$) according to the fitting formula in Tinker et al. (2008) at $z = 0.34$. 

xi
3.2 The ratio of the counts in the low (blue), medium (pink), and high (green) resolution simulation to the analytic fit presented in Tinker et al. (2008) in logarithmic mass bins ($\Delta \log_{10} M = 0.2$); we have used the initial comoving size of our box, $L_{\text{box}} = 558 \, \text{Mpc}/h$, to estimate the expected counts. This partially accounts for the underdensity of our realization. The low and medium resolution simulation results are output $z = 0.344$, while the output for the high resolution simulation is from $z = 0.339$.

3.3 The ratio of the counts in the low (blue) and medium (pink) resolution simulation to the high resolution simulation in logarithmic mass bins ($\Delta \log_{10} M = 0.2$). The low and medium resolution simulation results are for $z = 0.344$, while the output for the high resolution simulation is from $z = 0.339$.

3.4 The ratio of the counts in the low (blue) and medium (pink) resolution simulation to the high resolution simulation in logarithmic mass bins ($\Delta \log_{10} M = 0.2$), after scaling the low and medium resolution mass values by the factor 1.015. This factor more than accounts for the evolution in the mass function between $z = 0.344$ and $z = 0.339$.

3.5 The ratio of the counts in the low resolution simulation to the medium resolution simulation in logarithmic mass bins ($\Delta \log_{10} M = 0.2$). These simulations are both at $z = 0.344$.

3.6 The red curve shows the ratio of the power spectrum of the medium resolution simulation to the high resolution simulation, after scaling by the expected linear growth between the redshift of observation in the low and medium resolution simulations, $z = 0.344$, and the redshift of the high resolution simulation output, $z = 0.339$. Error bars are estimated as Poisson shot noise from the finite number of dark matter particles. The blue curve shows the ratio of the low resolution simulation to the medium resolution simulation, while the green curve is the ratio of the low resolution simulation to the scaled high resolution simulation (simply the product of the red and blue curves).
3.7 The ratio of the halo to dark matter power spectra vs. $k$ in units $(\text{Mpc}/h)^{-1}$ for the medium resolution simulation. The blue curve is for the low mass bin auto-power spectrum, the upper gold curve is for the high mass bin auto-power spectrum, and the middle red curve is the cross power spectrum of the low and high mass bins.

3.8 The ratio of the halo-dark matter cross spectrum vs. $k$ in units $(\text{Mpc}/h)^{-1}$ to the dark matter power spectra for the medium resolution simulation. The blue curve is for the low mass bin, and the red curve is for the high mass bin.

3.9 The ratio of the halo spectra for the medium to high resolution simulation vs. $k$ in units $(\text{Mpc}/h)^{-1}$. The blue curve is for the low mass bin auto-power spectrum, the gold curve is for the high mass bin auto-power spectrum, and the middle red curve is the cross power spectrum of the low and high mass bins. Linear fits to the ratio $r_0 + r_1 k$ are shown as well. The Poisson errors of Eqns. 3.8 to 3.13 are shown only for the low mass halo sample for clarity.

3.10 The ratio of the halo-dark matter cross spectra vs. $k$ in units $(\text{Mpc}/h)^{-1}$ in the medium and high resolution simulation. The blue curve is for the low mass bin, and the red curve is for the high mass bin. Linear fits to the ratio $r_0 + r_1 k$ are shown as well.

3.11 The autocorrelation of the high mass halos (gold), low mass halos (blue), their cross spectrum (red), and the linear $\xi$ averaged in the same bins of $r$ (green). $\xi(r)$ is computed from bins of size $\Delta r = 5 \text{ Mpc}/h$.

3.12 The autocorrelation of the high mass halos (gold), low mass halos (blue), and their cross spectrum (red). $\xi(r)$ is computed from bins of size $\Delta r = 0.5 \text{ Mpc}/h$.

3.13 $\xi_{\text{high}}(r) - \xi_{\text{med}}(r)$ computed from bins of size $\Delta r = 0.5 \text{ Mpc}/h$ for the low halo mass bin autocorrelation.

3.14 $\xi_{\text{high}}(r) - \xi_{\text{med}}(r)$ computed from bins of size $\Delta r = 0.5 \text{ Mpc}/h$ for the halo mass bins cross correlation.
3.15 $\xi_{\text{high}}(r) - \xi_{\text{med}}(r)$ computed from bins of size $\Delta r = 0.5$ Mpc/$h$ for the high halo mass bin autocorrelation. ................................. 70
3.16 $\xi_{\text{high}}(r) - \xi_{\text{med}}(r)$ computed from bins of size $\Delta r = 5$ Mpc/$h$ for the low halo mass bin autocorrelation. ................................. 71
3.17 $\xi_{\text{high}}(r) - \xi_{\text{med}}(r)$ computed from bins of size $\Delta r = 5$ Mpc/$h$ for the halo mass bins cross correlation. ................................. 71
3.18 $\xi_{\text{high}}(r) - \xi_{\text{med}}(r)$ computed from bins of size $\Delta r = 5$ Mpc/$h$ for the high halo mass bin autocorrelation. ................................. 72
3.19 The normalized halo-halo velocity correlation function, $\left\langle \vec{v}_i(\vec{r}) \cdot \vec{v}_j(\vec{r}') \right\rangle / \sqrt{\left\langle \vec{v}_i^2 \right\rangle \left\langle \vec{v}_j^2 \right\rangle}$, vs $|\vec{r} - \vec{r}'|$. Gold is the high mass bin, blue is the low mass bin, and red is the for low mass - high mass pairs. Results are binned in intervals of 0.5 Mpc/$h$ in $|\vec{r} - \vec{r}'|$. ................................. 72
3.20 The difference at small separations in the halo-halo velocity correlation function vs. $|\vec{r} - \vec{r}'|$ in Mpc/$h$ between high and medium resolution simulations. Gold is the high mass bin, blue is the low mass bin, and red is the for low mass - high mass pairs. Results are binned in intervals of 0.5 Mpc/$h$ in $|\vec{r} - \vec{r}'|$. ................................. 73
3.21 Same as Figure 3.20, but results are binned in intervals of 5 Mpc/$h$ and extend to $|\vec{r} - \vec{r}'| = 150$ Mpc/$h$. ................................. 73
3.22 The halo-halo relative velocity correlation function along the halo separation vector, $\left\langle (\vec{v}_i(\vec{r}) - \vec{v}_j(\vec{r}')) \cdot (\vec{r} - \vec{r}') / |\vec{r} - \vec{r}'| \right\rangle$ vs $|\vec{r} - \vec{r}'|$. Gold is the high mass bin, blue is the low mass bin, and red is the for low mass - high mass pairs. Results are binned in intervals of 0.5 Mpc/$h$ in $|\vec{r} - \vec{r}'|$. ................................. 74
3.23 The difference in the halo-halo relative velocity correlation function along the halo separation vector vs. $|\vec{r} - \vec{r}'|$ in Mpc/$h$ in the high and medium resolution simulations. Gold is the high mass bin, blue is the low mass bin, and red is the for low mass - high mass pairs. Results are binned in intervals of 5 Mpc/$h$ in $|\vec{r} - \vec{r}'|$. ................................. 75
3.24 The $\langle N_{\text{cen}}(M) \rangle$ (plateaus to 1) and $\langle N_{\text{sat}}(M) \rangle$ vs. $M$ in units of $10^{14} M_\odot$ for the HOD parameters listed in Table 3.1 for the low (blue), medium (red), and high (green) resolution mocks with $\sigma_{\log M} = 0.3$. 

3.25 The average number of central and satellite LRGs in each logarithmic mass bin with $\Delta \log_{10} M = 0.2$ vs. $M$ in units of $10^{14} M_\odot$. Low resolution (blue), medium (red), and high (green) resolution mocks are shown for $\sigma_{\log M} = 0.3$. 

3.26 The average number of satellite LRGs in each logarithmic mass bin with $\Delta \log_{10} M = 0.2$ vs. $M$ in units of $10^{14} M_\odot$. Low resolution (blue), medium (red), and high (green) resolution mocks are shown for $\sigma_{\log M} = 0.3$. 

3.27 The $\langle N_{\text{cen}}(M) \rangle$ (plateaus to 1) and $\langle N_{\text{sat}}(M) \rangle$ vs. $M$ in units of $10^{14} M_\odot$ for the HOD parameters listed in Table 3.1 for the low (blue), medium (red), and high (green) resolution mocks with $\sigma_{\log M,\text{high}} = 0.8$ (green), $\sigma_{\log M,\text{med}} = 0.92$ (red), and $\sigma_{\log M,\text{low}} = 1.8$ (blue). 

3.28 The average number of central and satellite LRGs in each logarithmic mass bin with $\Delta \log_{10} M = 0.2$ vs. $M$ in units of $10^{14} M_\odot$ for catalogs with $\sigma_{\log M,\text{high}} = 0.8$ (green), $\sigma_{\log M,\text{med}} = 0.92$ (red), and $\sigma_{\log M,\text{low}} = 1.8$ (blue). Note that the SO catalogs go to 50 particles/halo, which corresponds to a mass limit of $1.69 \times 10^{13} M_\odot$, $7.15 \times 10^{12} M_\odot$, and $3.66 \times 10^{12} M_\odot$ for the low, medium, and high resolution catalogs. 

3.29 The average number of satellite LRGs in each logarithmic mass bin with $\Delta \log_{10} M = 0.2$ vs. $M$ in units of $10^{14} M_\odot$ for catalogs with $\sigma_{\log M,\text{high}} = 0.8$ (green), $\sigma_{\log M,\text{med}} = 0.92$ (red), and $\sigma_{\log M,\text{low}} = 1.8$ (blue).
3.30 Comparison of our high resolution simulation $\sigma_{\log M} = 0.3$ (blue), $\sigma_{\log M} = 0.8$ (red), and Tegmark et al. (2006) real space $P(k)$ (green) in $(\text{Mpc}/h)^3$ vs $k$ in $(\text{Mpc}/h)^{-1}$. We estimate the real space $P(k)$ from our redshift space power spectrum by dividing by the linear theory $P_{s,0}/P_{\text{real}}$ with $\beta = f/b_{\text{LRG}}$ and $b_{\text{LRG}}$ estimated from the $k = 0$ limit of $P_{\text{LRG}}(k)/P_{\text{ic}}(k)$. For $\sigma_{\log M} = 0.3$, $b_{\text{LRG}} = 2.35$ and for $\sigma_{\log M} = 0.8$, $b_{\text{LRG}} = 2.05$. These two values of $\sigma_{\log M}$ bracket the observed large scale amplitude. The comparison is not perfect, of course, since our simulation is small compared to the SDSS LRG volume and slightly underdense. We have not applied the FOG compression in Tegmark et al. (2006); this should not matter for the amplitude of the largest modes.

3.31 $\log_{10} w_p(r_p)$ vs. $\log_{10} r_p$, both in Mpc$/h$, for $\sigma_{\log M} = 0.3$. Low (blue), medium (red), and high (green) resolution are shown with Masjedi et al. (2006) SDSS measurements. The error bars on the simulated catalogs show the error on the average $w_p(r_p)$ for the simulated box arising from the stochasticity of populating halos after averaging over 40 HOD realizations; this does not represent any cosmic variance from dark matter/halo density fields.

3.32 Same as Figure 3.31, but for $\tilde{w}_p$ as defined in Eqn. 3.17.

3.33 The ratio of the average redshift space power spectra vs. $k$ in units $(\text{Mpc}/h)^{-1}$ of the low resolution simulation to the medium resolution simulation for the $\sigma_{\log M} = 0.3$ catalogs, averaged over 40 HOD realizations, from which we derive the HOD-induced error bars representing an acceptable level of systematic resolution bias.

3.34 Same as Figure 3.33, but for the ratio of the medium to high resolution simulation for the $\sigma_{\log M} = 0.3$ catalogs.

3.35 The ratio of the average redshift space power spectra of the medium resolution simulation $\sigma_{\log M} = 0.92$ catalogs to the high resolution simulation $\sigma_{\log M} = 0.8$ catalogs vs. $k$ in units $(\text{Mpc}/h)^{-1}$. We averaged over 40 HOD realizations, from which we derive the HOD-induced error on the average.
3.36 $\log_{10} w_p(r_p)$ vs. $\log_{10} r_p$, both in Mpc/$h$, for $\sigma_{\log M,\text{high}} = 0.8$, $\sigma_{\log M,\text{med}} = 0.92$, and $\sigma_{\log M,\text{low}} = 1.8$ catalogs. Low (blue), medium (red), and high (green) resolution are shown with Masjedi et al. (2006) SDSS measurements. The error bars on the simulated catalogs show the error on the average $w_p(r_p)$ for the simulated box arising from the stochasticity of populating halos after averaging over 40 HOD realizations; this does not represent any cosmic variance from dark matter/halo density fields.

3.37 Same as Figure 3.36, but for $\tilde{w}_p$ as defined in Eqn. 3.17.

3.38 The cumulative distribution of LRGs as a function of their host halo mass in units of $10^{14} M_\odot$ for the $\sigma_{\log M,\text{high}} = 0.8$ (gold), $\sigma_{\log M,\text{med}} = 0.92$ (red), and $\sigma_{\log M,\text{low}} = 1.8$ (blue) catalogs; this is just the integral of Figure 3.28.

3.39 $P(k)$ in units of $(\text{Mpc}/h)^3$ vs $k$ in $(\text{Mpc}/h)^{-1}$ for the high resolution $\sigma_{\log M} = 0.6$ catalog where we have scaled the redshift space $P(k)$ using Eqn. 1.20 to estimate the real space power spectrum, as in Fig. 3.30.

3.40 The cumulative distribution of LRGs as a function of their host halo mass in units of $10^{14} M_\odot$ for the high resolution simulation mocks with $\sigma_{\log M} = 0.6$, $\sigma_{\log M} = 0.7$, and $\sigma_{\log M} = 0.8$ catalogs.

3.41 $P(k)$ in $(\text{Mpc}/h)^3$ vs $k$ in $(\text{Mpc}/h)^{-1}$ for the medium resolution NEAR sample ($z = 0.233$) mock catalogs with $\sigma_{\log M} = 0.5$ (light blue), $\sigma_{\log M} = 0.6$ (blue) and $\sigma_{\log M} = 0.8$ (red). We have estimated the real space $P(k)$ from the redshift space $P(k)$ as in Figure 3.30. Tegmark et al. (2006) shown in green.
3.42 $P(k)$ in $(\text{Mpc}/h)^3$ vs $k$ in $(\text{Mpc}/h)^{-1}$ for the medium resolution FAR sample ($z = 0.42$) mock catalogs with $\sigma_{\log M} = 0.9$, $\bar{n} = 4.3 \times 10^{-5} (\text{Mpc}/h)^3$ central objects only (blue), central and satellite galaxies (red), and $P(k)$ published in Tegmark et al. (2006) (green). We have estimated the real space $P(k)$ from the redshift space $P(k)$ as in Figure 3.30. The cyan curve shows the power spectrum of a galaxy catalog with $\sigma_{\log M} = 0.9$ and $\bar{n} = 10^{-4} (\text{Mpc}/h)^3$. Because of the form of $N_{\text{cen}}(M)$ we have chosen, this catalog has a lower $P(k)$ amplitude than the lower density catalogs.

3.43 $P(k)$ in $(\text{Mpc}/h)^3$ vs $k$ in $(\text{Mpc}/h)^{-1}$ for the medium resolution FAR sample ($z = 0.42$) mock catalogs with $\sigma_{\log M} = 1.2$ central objects only (blue), central and satellite galaxies (red), and $P(k)$ published in Tegmark et al. (2006) (green). We have estimated the real space $P(k)$ from the redshift space $P(k)$ as in Figure 3.30.

3.44 $N_{\text{cen}}(M)$ vs $M$ in units of $10^{14} M_\odot$ for the three FAR ($z = 0.42$) mock catalogs: $\sigma_{\log M} = 0.9$, $\bar{n} = 4.3 \times 10^{-5} (\text{Mpc}/h)^3$ (red), $\sigma_{\log M} = 1.2$, $\bar{n} = 4.3 \times 10^{-5} (\text{Mpc}/h)^3$ (green), and $\sigma_{\log M} = 0.9$, $\bar{n} = 10^{-4} (\text{Mpc}/h)^3$ downsampled by a 0.43 (blue).

3.45 The cumulative distribution of LRGs as a function of their host halo mass $M$ in units of $10^{14} M_\odot$ for the medium resolution FAR sample ($z = 0.42$) mock catalogs: $\sigma_{\log M} = 0.9$, $\bar{n} = 4.3 \times 10^{-5} (\text{Mpc}/h)^3$ (blue), $\sigma_{\log M} = 1.2$, $\bar{n} = 4.3 \times 10^{-5} (\text{Mpc}/h)^3$ (red), and $\sigma_{\log M} = 0.9$, $\bar{n} = 10^{-4} (\text{Mpc}/h)^3$ (gold).

4.1 Black points show the error bars divided by the bandpower for the LRG power spectrum published in Tegmark et al. (2006). The black curve shows $(1 + Q_{NL} k^2)/(1 + 1.4 k) - 1$, the fractional amplitude of the nonlinear $Q_{NL}$ correction to the linear power spectrum for $Q_{NL} = 30$, the best fit value in Tegmark et al. (2006). For comparison, we show the smooth nonlinear correction fit to the dark matter power spectrum of Fig. 4.4. This is the size of the correction for our reconstructed halo density field.
4.2 The ratio of the volume-weighted SO mass function from our simulations with the spline fit given in Tinker et al. (2008) for the NEAR (solid line), MID (dashed line), and FAR (dotted line) in bins of $\Delta \log_{10}M = 0.1$. The agreement is within the stated accuracy of the Tinker et al. (2008) spline fit ($\sim 5\%$).

4.3 The ratio of the volume-averaged matter power spectra at $z_{FAR} = 0.421$ (black), $z_{MID} = 0.342$ (blue), and $z_{NEAR} = 0.235$ (green) to the input power spectrum scaled by $D^2$. The nonlinear correction to the matter power spectrum measured from our simulations is larger than expected from halofit (Smith et al., 2003), shown in red and evaluated at $z_{MID}$. We also overlay $\log k P_{IC}(k)$ to indicate the location of the baryon wiggles in the initial power spectrum. The nonlinear evolution generates power that smoothly increases with $k$ and damps the baryon oscillations. Error bars are shown only for the MID sample for clarity and are estimated from the inverse of the model covariance matrix defined in Eqn. 4.10 where the shot noise is negligible and $\beta_{beat} = 1$; the beat-coupling term in Eqn. 4.10 correlates the bandpowers.

4.4 $P_{DM}(k)/P_{IC}(k)$ for the MID sample (cyan), $P_{DM}(k)/P_{\text{smear}}(k)$ for $k_{BAO} = 0.14$ (blue), and our model fit $1.026 - 1.199k + 11.06k^2 - 8.426k^3$ (red). We also overlay $\log k P_{IC}(k)$ to indicate the location of the baryon wiggles in the initial power spectrum. Using error bars derived from the inverse of the model covariance matrix defined in Eqn. 4.10, $\chi^2 = 21$ for 35 bandpowers and 5 parameters. Comparison of the cyan and blue curves shows that $P_{\text{smear}}(k)$ adequately models the BAO features, and comparison of the blue and red curves demonstrates that our third-order polynomial in $k$ is sufficient for $k \leq 0.4 \ h/\text{Mpc}$.
4.5 The blue curve shows the measured diagonal elements of the covariance matrix. Error bars are estimated from the variance of $\langle (\Delta P_i)^2 \rangle$ across our simulations. The red curve shows the Gaussian prediction for the diagonal covariance, the inverse of the number of complex modes in each linearly-spaced $k$-bin. The green curve shows the sum of the Gaussian term and the beat-coupling term in Hamilton et al. (2006), which provides a good model for the full covariance matrix. In black we show the inverse of the diagonal elements of the inverse covariance matrix. We find that while the errors on the bandpowers are very close to the Gaussian expectation, the beat-coupling terms in the covariance matrix mean the bandpowers are correlated.

4.6 $kP_s(k)$ for the NEAR (blue), MID (green), and FAR (red) mock galaxy samples with the FOG compression algorithm of Tegmark et al. (2006) applied compared with their observed power spectrum (black points with errors). Our NEAR and MID samples are virtually identical, so we show error bars only for the MID and FAR samples. Our error bars are derived from the diagonal elements of the inverse $C_{ij,\text{red}}$ matrix. We scale the real space $P(k)$ of Tegmark et al. (2006) by $(0.8)^{-1}$ to approximate the redshift space monopole power spectrum; this is the relation between the real and redshift monopole spectra if one neglects the small contributions from the quadrupole and hexadecapole detailed in their §A3. For comparison, the jagged cyan curve shows our simulation initial conditions drawn from the convolved power spectrum according to the $ic$ algorithm, and the smooth cyan curve shows the linear power spectrum for the cosmological parameters adopted in this study.

4.7 Figure 6 of Tegmark et al. (2006) showing the power spectra of the NEAR, MID, and FAR subsamples along with the MAIN galaxies. Our mock FAR sample in Fig. 4.6 also has slightly more power at small scales compared to the MID and NEAR subsamples.
4.8 We separately fit our NEAR, MID, and FAR FOG-compressed samples to the nonlinear correction model of Tegmark et al. (2006) in Eqn. 4.6. Best fit parameters are $b_{NEAR}^2 = 5.61$, $Q_{NL,NEAR} = 17.4$; $b_{MID}^2 = 6.15$, $Q_{NL,MID} = 18.4$; and $b_{FAR}^2 = 6.67$, $Q_{NL,FAR} = 23.3$. Each subsample is scaled by the input power spectrum and the best fit $b^2$ value to accentuate the variation in the power spectrum shapes for these three samples. The best fit vanilla model in Tegmark et al. (2006) is shown by two cyan curves, where we have adjusted the normalization to agree at $k \sim 0.05$. The top curve has $A = 1.4$ in Eqn. 4.4 and the bottom curve has $A = 1.7$, and both have $Q_{NL} = 31$. The black curve shows the ratio of the best fit linear power spectrum of Tegmark et al. (2006) and the linear power spectrum adopted in this work, demonstrating that $Q_{NL}$ is degenerate with changes in the spectral shape, so that cosmological parameter information is lost to $Q_{NL}$. 

4.9 The ratio of the LRG to nonlinear dark matter spectrum for the MID sample. The lower red curve shows only central galaxies with the power spectrum computed in real space. The upper red curve is for the same galaxy sample but measured in redshift space. The lower and upper blue curves are for the sample of both central and satellite galaxies in real and redshift space. The green curve shows the redshift space power spectrum of the central and satellite galaxies after undergoing Tegmark et al. (2006) FOG compression. The black curve is the power spectrum of the CiC groups in redshift space. The nonlinear correction associated with this sample is much smaller than the FOG-compressed sample. 

4.10 Same as Figure 4.9, but for the FAR sample. 

4.11 Same as Figure 4.9, but for the NEAR sample.
4.12 We plot the normalized covariance matrix elements $C_{ij}/P(k_i)P(k_j)$ scaled by $\sqrt{N_iN_j}$, the number modes in bins $i$ and $j$. The solid lines show our measurements and the dashed lines show the model fits from Table 4.4. We plot $C_{ij,\text{tot}}$ (black), $C_{ij,\text{reduced}}$ (blue), and $C_{ij,\text{HOD}}$ (red). The top left panel shows the diagonal elements, and the other panels show cross sections of the covariance matrix with $k_i = 0.09$, 0.15, and 0.3 $h$/Mpc.

4.13 The bottom right panel shows the ratio of the redshift space monopole to the real space power spectrum for the MID sample. The straight solid lines show the prediction of Eqn. 4.28 for both samples. We fit the bias in Eqn. 4.27 to the large scale ratio $P_{LRG}/P_{DM}$. The black curve is for the sample of central galaxies only ($b_{\text{cen}} = 2.11$) and the red includes both central and satellite galaxies ($b_{\text{sat}} = 2.23$). The cyan curve shows the ratio of the redshift space monopole of the central and satellite galaxies to the real space power spectrum of the central objects (scaled by $b_{\text{cen}}^2/b_{\text{sat}}^2$). The other three panels show the quadrupole to monopole ratio in redshift space. Error bars are only shown for the central galaxy sample (black) for clarity. The central and satellite sample is shown in red, the reconstructed halo density field in blue, and the FOG-compressed sample in green. The expected values from Eqn. 4.28 are shown as the solid lines for the central (black) and central and satellite (red) samples.
List of Tables

2.1 This table lists the CiC group multiplicity function of the SDSS LRG spectroscopic + photometric sample for \( \Delta r_{\perp,max} = 0.8 \text{Mpc}/h \) and \( \beta_{max} = 0.006 \). Column 1 lists the number of satellites in the group, \( n_{sat} = n_{\text{group}} - 1 \). Column 2 contains the raw group multiplicity before any corrections have been made (including for double counting of photometric objects). Column 3 lists the final group multiplicity function after all the corrections of § 2.2.2. Column 4 shows the number of photometric objects in groups with \( n_{sat} \) satellite galaxies, Column 5 shows the net effect of the Bright Star Mask correction, and Column 6 shows the number of groups of photometric objects only with \( n_{sat} \) satellites satisfying the criteria listed in § 2.2.2. These groups are largely false detections and do not contribute to the counts in Column 2. The total area of the survey, excluding the boundary region but including the bright star mask area, is 1.549 steradians.
2.2 Column 2 shows the average offset \( \Delta g(n_{\text{sat}}) = N_{\text{CiC}}(n_{\text{sat}}) - N_{\text{true}}(n_{\text{sat}}) \) over 600 mock catalogs for HOD parameters \( \sigma_{\log M} = 0.7 \), \( M_{\text{min}} = 8.05 \times 10^{13} M_\odot \), 
\( M_{\text{cut}} = 4.66 \times 10^{13} M_\odot \), \( M_1 = 4.95 \times 10^{14} M_\odot \), and \( \alpha = 1.07 \), which are close to the final maximum likelihood parameters reported in § 2.3.3. Column 3, \( \sigma^2_g \), is our estimate of the variance of \( \Delta g \) in a subsample of the simulation box with volume equal to our SDSS subsample, 0.46 (Gpc/h)^3. Comparison with Table 2.1, Column 2 (reproduced here in Column 4) shows that this variance is comparable to the expected variance from Poisson sampling of \( \langle N_{\text{CiC}}(n_{\text{sat}}) \rangle \). Column 5 is the fraction of true one-halo groups with \( n_{\text{sat}} \) satellites that are exact matches to a CiC group. Column 6 shows the fraction of CiC groups that do not exactly match a true one-halo group.

2.3 Column 1 is the number of satellites in the group. Column 2 is a copy of Column 3 of Table 2.1, our best estimate of the SDSS CiC multiplicity function. Column 3 shows the CiC multiplicity function averaged over 20 mock catalogs produced with our maximum likelihood HOD using the SO halo catalogs and scaled by the volume ratio \( V_{SDSS}/V_{\text{sim}} \). Column 4 shows the CiC multiplicity function for mocks made with the FoF catalog. Column 5 shows our estimate of the expected \( N_{\text{true}} \) in \( V_{SDSS} \) for the Kulkarni et al. (2007) HOD.

3.1 HOD parameters for \( \sigma_{\log M} = 0.3 \) for the 3 different resolution simulations, and for the large \( \sigma_{\log M} \) catalogs, with \( \sigma_{\log M,\text{high}} = 0.8 \), \( \sigma_{\log M,\text{med}} = 0.92 \), and \( \sigma_{\log M,\text{low}} = 1.8 \). \( M_{\text{min}} \) was adjusted for each resolution to match a fixed \( N_{\text{LRG,tot}} \). Masses in units of \( 10^{14} M_\odot \).

3.2 Group biases and errors on the average after 40 HOD realizations with \( \sigma_{\log M} = 0.3 \). The eighth column shows the average number of one-halo groups with \( n_{\text{sat}} \) satellites in the high resolution simulation. The number of CiC groups on average is just column 8 multiplied by column 6.
3.3 Group biases and errors on the average after 40 HOD realizations with 
\( \sigma_{\log M,\text{low}} = 1.8, \sigma_{\log M,\text{med}} = 0.92, \sigma_{\log M,\text{high}} = 0.8 \). The eighth column 
shows the average number of one-halo groups with \( n_{\text{sat}} \) satellites in the high 
resolution simulation. The number of CiC groups on average is just column 
8 multiplied by column 6.

3.4 Group biases and errors on the average after 40 HOD realizations with 
\( \sigma_{\log M,\text{low}} = 0.8, \sigma_{\log M,\text{med}} = 0.8, \sigma_{\log M,\text{high}} = 0.8 \). The eighth column shows 
the average number of one-halo groups with \( n_{\text{sat}} \) satellites in the high resolution simulation. The number of CiC groups on average is just column 8 
multiplied by column 6.

4.1 The final \( N_{\text{CiC}}(n_{\text{sat}}) \) group multiplicity function following the method in 
Ch. 2 for our NEAR, MID, and FAR LRG subsamples.

4.2 Mock catalog parameters. Masses are for SO halos with \( \Delta = 200 \) in units of 
\( 10^{14} M_\odot \) and number densities are in units of \( 10^{-4} (\text{Mpc}/h)^{-3} \).

4.3 Fits to nonlinear power spectra. \( k_{\text{max,fit}} \) and \( k_{\text{BAO}} \) are in units of \( h/\text{Mpc} \). 
For \( P_{DM}(k)/P_{\text{IC}}(k) \) we fit to Eqn. 4.22. When \( k_{\text{max,fit}} = 0.2 \) we hold \( a_3 = 0 \). 
Some of the redshift space mock catalog fits are not well-behaved at very small 
k’s, and the corrections to a constant should be suppressed at \( k \lesssim 0.05 \). This 
also makes \( a_0 \) deviate from the large scale bias, given in the last column and 
fit using only \( k < 0.06 \). \( P_{\text{sat},z}(k) \) denotes the power spectrum of a catalog 
containing central and satellite galaxies in redshift space, \( P_{\text{CiC},z}(k) \) denotes 
the power spectrum of the reconstructed halo density field in redshift space, 
and \( P_{\text{teg},z}(k) \) denotes an FOG compressed following Tegmark et al. (2006).
4.4 Fits for the mock catalog covariance matrices using Eqns. 4.23 - 4.26 using all bandpowers between $k = 0.01$ and $k = 0.4$. ‘cen, real/redshift’ denotes a sample of only central LRGs in real/redshift space; ‘sat, real/redshift’ denotes a sample with both central and satellite galaxies in real/redshift space; ‘CiC, redshift’ denotes the reconstructed halo density field using the CiC method, and ‘Teg, redshift’ denotes the FOG-compressed sample according to Tegmark et al. (2006).

4.5 Fits for the mock catalog covariance matrices using Eqns. 4.23 - 4.26 using all bandpowers between $k = 0.05$ and $k = 0.2$. Samples are the same as in Table 4.4.
Chapter 1

Introduction

The past decade has seen a dramatic increase in the quantity of cosmological data, and with it, a new understanding of the components of the universe. Using supernovae as standard candles to study the luminosity distance - redshift relation, the High-Z Supernova Search (Riess et al., 1998) and the Supernova Cosmology Project (Perlmutter et al., 1999) discovered that the universe’s expansion rate is accelerating. In a universe obeying general relativity and containing only matter, the expansion rate must decelerate. At present, the “dark energy” causing the acceleration can be modeled with Einstein’s cosmological constant and be consistent with the host of astronomical data (Komatsu et al., 2008).

In 1992, the Cosmic Background Explorer collaboration reported the first detection of the tiny cosmic microwave background (CMB) anisotropies (Smoot et al., 1992; Wright et al., 1992), which then quickly became a prominent source of cosmological data. The power spectrum of the CMB anisotropies measured by the Wilkinson Microwave Anisotropy Probe (WMAP) is cosmic variance limited out to $\ell \sim 500$, encompassing $\sim 500,000$ independent spherical harmonics (Nolta et al., 2008) and providing the tightest constraints on cosmological parameters for a single data set (Dunkley et al., 2008).

Galaxy redshift surveys have also dramatically grown in scope: the first CfA Survey
was completed in 1982 and compiled redshifts for 2401 galaxies (Huchra et al., 1983); the PSCz (Saunders et al., 2000) contains \( \sim 15000 \) IRAS galaxies out to \( z = 0.1 \); the final data release of the Sloan Digital Sky Survey (SDSS) Luminous Red Galaxy (LRG) sample alone will have \( \sim 8 \) times more galaxies and cover \( \sim 2 \ (\text{Gpc}/h)^3 \), or \( \sim 20 \) times more volume than PSCz. Eisenstein et al. (2005) report the detection of the baryon acoustic oscillation (BAO) peak, a relic from the epoch of recombination, in the clustering of the LRGs at \( z \sim 0.3 \). Komatsu et al. (2008) combine the most recent BAO and supernovae results with the WMAP five year measurements to constrain cosmological models. They detect no significant deviation from the minimal \( \Lambda \)CDM model, and the parameters of the model are constrained within a few percent.

The shape and amplitude of galaxy power spectra can provide additional cosmological constraints. The number of independent modes available for a survey volume \( V \) and maximum wavenumber \( k_{\text{max}} \) is \( N_{\text{modes}} = k_{\text{max}}^3 V/6\pi^2 = 130,000 \) for \( k_{\text{max}} = 0.2 \ h/\text{Mpc} \) and the SDSS LRG volume. \( k_{\text{max}} \) is limited by the onset of nonlinearities; current measurements usually constrain their cosmological analyses to \( k_{\text{max}} \sim 0.1 \) (Tegmark et al., 2006). Since galaxy clustering measurements probe smaller scales than CMB measurements, connecting the two would allow tight constraints on the shape of the initial spectrum of fluctuations. However, the galaxy power spectrum has been transformed from the linear power spectrum through nonlinear growth of structure, redshift space distortions, and the uncertain relation between the galaxy and dark matter density fields. Sanchez & Cole (2007) find significant systematic differences in the galaxy power spectrum results from the Two-Degree Field Galaxy Redshift Survey (2dFGRS) and SDSS. Moreover, Dunkley et al. (2008) find that the best fit \( \Omega_m \) varies with \( k_{\text{max}} \) when combining the LRG power spectrum with WMAP, and so they choose not to include the galaxy power spectrum in their cosmological parameter analysis that combines several independent data sets. Verde & Peiris (2008) find that the LRG sample has less statistical power than the MAIN SDSS or 2dFGRS galaxies because of the strong nonlinear corrections necessary. Improved modeling of the nonlinearities in
galaxy power spectra is therefore essential to harnessing their power as cosmological probes, and is the underlying goal of this thesis. We first review the basic cosmological model and results of the theory of structure formation.

1.1 Cosmological Model

1.1.1 The Homogeneous Universe

In this thesis we will assume the minimal ΛCDM model, which requires 5 parameters to specify the linear matter power spectrum $P_{\text{lin}}(k)$ at low redshift. Komatsu et al. (2008) place stringent limits on the possible deviations from this model. We focus on the epoch of the universe well after recombination, where the energy density in radiation is negligible, and so only consider the energy density in matter and the cosmological constant. We now summarize the basic cosmological model, drawing from Dodelson (2003), Peebles (1980), and Verde (2000) throughout this section.

The cosmological parameters of this model are

- $\Omega_m$: the density of matter today ($a = 1$), both baryonic and cold dark matter, in units of the critical density $\rho_{c,0}$
- $\Omega_b$: the baryonic matter density today in units of the critical density $\rho_{c,0}$
- $h$: The expansion rate $\dot{a}/a$ in units of 100 km/s/Mpc
- $\sigma_8$: The linear rms variance of matter on 8 Mpc/$h$ scales
- $n_s$: The power-law index of the primordial perturbations

The critical density is given by

$$\rho_c(a) = \frac{8\pi G}{3H^2(a)}. \quad (1.1)$$
\( \Omega_\Lambda = 1 - \Omega_m \) accounts for the dark energy density in the form of the cosmological constant. This minimal model assumes that the curvature of the universe is 0, so that the Friedmann-Walker-Robertson (FRW) metric for a homogeneous universe is
\[
d\vec{s} \cdot d\vec{s} = dt^2 - a^2(t)d\vec{x} \cdot d\vec{x} \tag{1.2}
\]
where \( d\vec{x} \) is the usual three-dimensional Euclidean line element and \( a(t) \) is the scale factor of the universe at time \( t \). \( a(t) \) obeys
\[
\left( \frac{\dot{a}}{a} \right)^2 = H^2(a) = \frac{8\pi G}{3} \rho_{c,0} \left[ \Omega_m a^{-3} + \Omega_\Lambda \right]. \tag{1.3}
\]
In a homogeneous universe, light emitted when the universe had scale factor \( a \) will be redshifted according to
\[
a = \frac{1}{1 + z}. \tag{1.4}
\]
Throughout this thesis we use comoving distances. \( \chi(a) \) is the distance light has traveled that was emitted when the universe had scale factor \( a \) and is observed by us today:
\[
\chi(a) = \int_{t(a)}^{t_0} \frac{dt'}{a(t')} = \int_a^1 \frac{da'}{a'^2 H(a')}. \tag{1.5}
\]

### 1.1.2 Density Perturbations in Linear Theory

Fluctuations in the matter density field are defined by
\[
\delta(\vec{x}, a) = \rho(\vec{x}) / \rho_m(a) - 1 \tag{1.6}
\]
where \( \rho_m(a) = \rho_{c,0} \Omega_m a^{-3} \) is the mean matter density in the universe. Consistent with the WMAP results (Komatsu et al., 2003), primordial matter fluctuations are assumed to be isotropic, adiabatic, and Gaussian. Moreover, we parametrize the primordial power spectrum as a power law:
\[
P_{\text{prim}}(k) \propto k^{n_s}. \tag{1.7}
\]
Fluctuations are damped on small scales by the photon-baryon fluid that dominated the early universe, so that the power spectrum of fluctuations that seeded the large scale structure is given by
\[
P_{\text{lin}}(k) = T^2(k; \Omega_m, \Omega_b, h) P_{\text{prim}}(k). \tag{1.8}
\]
This is the linear power spectrum we refer to throughout this thesis; on large scales, the
galaxy power spectrum is proportional to $P_{lin}(k)$. $T$ is the transfer function, and can be ef-
ciently computed as a function of the cosmological parameters using codes like CMBFAST
(Seljak & Zaldarriaga, 1996) or CAMB Lewis et al. (2000), or approximated analytically
(Bardeen et al., 1986; Eisenstein & Hu, 1999). We parametrize the amplitude of $P(k)$
through $\sigma_8$, the linear rms fluctuations on 8 Mpc/$h$ scales today:

$$\sigma_R^2 = \frac{1}{2\pi^2} \int P_{lin}(k)W_R^2(k)k^2dk$$

(1.9)

where $W_R(k)$ is the Fourier transform of a spherical top hat of radius $R = 8 \, h^{-1} \, \text{Mpc}$.

In the Newtonian approximation, the cosmological fluid obeys

$$\frac{\partial \delta}{\partial t} + \frac{1}{a} \nabla \cdot [(1 + \delta) \vec{v}_p] = 0$$

(1.10)

$$\frac{\partial \vec{v}_p}{\partial t} + \frac{\dot{a}}{a} \vec{v}_p + \frac{1}{a} (\vec{v} \cdot \nabla) \vec{v}_p = -\frac{1}{a} \nabla \phi$$

(1.11)

$$\nabla^2 \phi = 4\pi G \rho_m a^2 \delta.$$  

(1.12)

We have adopted comoving coordinates, but $\vec{v}_p$ is the proper velocity relative to the Hubble
flow: $\vec{v}_p = a \dot{x}$. To first order in $\delta$ or $\vec{v}_p$, $\delta$ obeys

$$\frac{\partial^2 \delta}{\partial t^2} + 2\frac{\dot{a}}{a} \frac{\partial \delta}{\partial t} = 4\pi G \rho_m \delta.$$  

(1.13)

Retaining only the growing mode, we find

$$\delta(\vec{x}, a) = D(a) \delta_0(\vec{x}) / D(1)$$

(1.14)

$$D(a) = \frac{5\Omega_m H_0^2 \dot{a}}{2} \int_0^a \frac{da'}{a'^3}.$$  

(1.15)

That is, the initial density fluctuation field maintains the same shape and grows in amplitude
according to $D(a)$. Taking the curl of both sides of Eqn. 1.11, one finds that any vorticity
in the primordial velocity field will decay as $\nabla \times \vec{v}_p \propto 1/a$. Therefore, one can assume that
$\vec{v}$ is irrotational, and Eqn. 1.10 becomes

$$\vec{v}(\vec{k}) = \frac{dD(a)}{da} \frac{i a^2 H(a) \delta(\vec{k}, a)}{D(a)k} \hat{k}$$

(1.16)
1.2 Nonlinearity in the Galaxy Power Spectrum

There are three principal sources of nonlinearity between the observed galaxy power spectrum and the linear matter power spectrum (Eqn. 1.8). First, nonlinearities develop in the matter density field and Eqn. 1.13 is no longer valid. Second, the galaxy power spectrum can only be observed in redshift space, where the line-of-sight (LOS) coordinate is a combination of the real space galaxy position and its peculiar velocity. Finally, the galaxy density field is not a simple linear function of the underlying mass distribution.

1.2.1 Nonlinear Growth of Structure

In the linearized solution to Eqns. 1.10 - 1.12, the initial fluctuation field changes only in amplitude. If, as we assume throughout this thesis, the primordial fluctuations at each $k$ were independent and drawn from a Gaussian field, the linear density field will also consist of independent Gaussian modes. However, when $\delta \sim 1$, the linearized solution to Eqns. 1.10 - 1.12 breaks down. Modes at different $k$ are coupled, higher order correlations are generated, and the probability distribution function of $\delta(k)$ becomes non-Gaussian. Small scale modes go nonlinear first, while the large scale modes can still be treated in the linear regime.

There have been three main avenues for understanding and quantifying the nonlinear regime of gravitational instability: higher order perturbation theory of Eqns. 1.10 - 1.12 (Bernardeau et al., 2002), analytic techniques in the highly nonlinear regime (Press & Schechter, 1974; Bardeen et al., 1986; Lacey & Cole, 1993), and N-body simulations (Bertschinger, 1998).

As the density field evolves, small regions collapse into virialized objects called halos. Press & Schechter (1974) pioneered an analytic description of the highly nonlinear regime by computing the halo mass function $n(M)$, the number density of halos with mass between $M$ and $M + dM$, from the initial density field. Bardeen et al. (1986) showed that peaks
in a Gaussian random field larger than some value will be biased relative to the density field. Navarro et al. (1997) fit halo density profiles from N-body simulations to a universal function. These three concepts form the basis of the halo model description of large scale structure, reviewed in Cooray & Sheth (2002). In its simplest form, all matter resides in halos, and the nonlinearity of the matter density field is encompassed by the conditional mass function \( n(M|\delta) \), the number density of halos of mass \( M \) in regions where the large scale overdensity is \( \delta \). To compute an \( n \)-point correlation function, one considers all permutations of \( n \) objects from \( m \leq n \) different halos. The halo overdensity field \( \delta_h \) is related to the linear matter density field \( \delta_{DM, \text{lin}} \) by the expansion (Eqn. 67 in Cooray & Sheth (2002))

\[
\delta_h = \sum_{k>0} b_k(M) \delta_{DM, \text{lin}}^k.
\]

\( k > 1 \) terms can be neglected on sufficiently large scales, where halos are approximately linearly biased.

### 1.2.2 Redshift Space Distortions

The observed redshift of an object is a combination of its cosmological redshift \( z_d \) and its peculiar velocity along the LOS, \( v_{p,\text{LOS}} \):

\[
z_{\text{obs}} = z_d + \left(1 + z_d\right) v_{p,\text{LOS}}
\]

where its real space comoving distance from the observer is \( \chi(z_d) \). The relation between the comoving redshift space coordinate and the real space comoving distance to the object is

\[
s = \chi(z_{\text{obs}}) \approx \chi(z_d) + \frac{(1 + z_d)v_{p,\text{LOS}}}{H(z_d)} \approx \chi(z_d) + \frac{(1 + z_{\text{obs}})v_{p,\text{LOS}}}{H(z_{\text{obs}})}.
\]

Kaiser (1987) worked out the relation between the redshift and real space galaxy density field in linear theory to be

\[
\delta_s(\vec{k}) = \left[1 + \beta \mu_k^2\right] \delta(\vec{k})
\]
where $\mu_k = \hat{k} \cdot \hat{z}$ and

$$\beta = \frac{1}{b_{gal}} \frac{a}{D(a)} \frac{dD}{da}.$$  \hfill (1.21)

$b_{gal}$ is the linear bias between the galaxy and matter density fields. In linear theory modes along the LOS are enhanced, so that overdense regions appear squashed in that direction. On scales comparable to the size of a virialized halo, the Finger-of-God (FOG) effect dominates. Objects orbiting in virialized halos will have LOS velocities with dispersions set by the halo virial velocity:

$$\sigma_{LOS}^2 \sim \frac{GM}{2R}.$$ \hfill (1.22)

which can be several times the rms linear velocity. For objects in the same dark matter halo, the velocity contribution to Eqn. 1.19 will be much larger than the intergalaxy separation, and the group will be stretched along the LOS.

1.2.3 The Galaxy-Dark Matter Connection

The simplest model for the clustering of a particular type of galaxy $g$ is to assume that they are biased Poisson tracers of the underlying matter density field, which predicts a simple relation between the galaxy and matter power spectra:

$$\delta_g(\bar{x}) = b_g \delta_{DM}(\bar{x}) + n_{shot}(\bar{x})$$ \hfill (1.23)

$$\langle n_{shot}(\bar{x}) n_{shot}(\bar{x}') \rangle = \frac{1}{n_g} \delta^D(\bar{x} - \bar{x}')$$ \hfill (1.24)

$$P_g(k) = b_g^2 P_{DM}(k) + \frac{1}{n_g}.$$ \hfill (1.25)

While the ‘gastrophysics’ involved in galaxy formation is too complex to make detailed predictions from first principles, it is commonly believed that galaxies form in the potential
well of dark matter halos. Therefore a natural improvement on the Poisson model is to associate galaxies with dark matter halos. The halo model can be simply extended to describe galaxies with the assumption that average galaxy properties depend only on the host halo mass. In principle, the halo model could be extended to include additional dependencies on local environment or halo merger history. However, the halo model has had considerable success without invoking dependencies other than on halo mass (e.g., Tinker et al. (2007)). For our purposes of investigating nonlinearities in large scale structure statistics, this approximation is probably sufficient. This model provides a simple framework to compute galaxy statistics, since \( N \)-body simulations can be used to find the mass function of halos and their clustering properties.

In a halo model for galaxies, one must specify the Halo Occupation Distribution (HOD) \( P(N|M) \), the probability that \( N \) galaxies reside in a halo of mass \( M \), along with the spatial distribution of galaxies within a halo. Then the computation of \( n \)-point statistics proceeds as for dark matter halos, described in §1.2.1. In § 2.2.4 we provide the details of our implementation of the halo model.

Sánchez & Cole (2008) compare the galaxy power spectra from the 2dFGRS galaxies, the SDSS main sample, and the SDSS LRG sample and find that the 2dFGRS spectrum has relatively more large scale power; equivalently, the SDSS has relatively more small scale power. Sánchez & Cole (2008) conclude that the \( r \)-band selected SDSS galaxies have more scale-dependent biasing. As we will demonstrate in this thesis, such an effect will arise whenever two populations have differing halo occupation distributions. The discrepancy found by Sánchez & Cole (2008) causes changes in the inferred cosmological parameters that are larger than the statistical errors. Therefore, the modeling of galaxy samples must be improved before we can harness the statistical power of current and more ambitious redshift surveys of the future. The goal of this thesis is to address this very issue for the SDSS LRGs.
1.2.4 Solution: Reconstructing the Halo Density Field

The one-halo contribution to the galaxy power spectrum produces a scale-dependent bias that depends on the galaxy sample HOD. To eliminate this term for the SDSS LRG sample, we use FOG features in the LRG density field to reconstruct the halo density field. The power spectrum of the halo density field will have only a residual contribution from the one-halo term. Moreover, this will partially eliminate the FOG effect in the resulting $P_s(k)$, making estimates of $\beta$ and the real space $P(k)$ more robust. Finally, the nonlinear mapping between the halo density field and the linear matter field can easily be studied using $N$-body simulations.

1.3 SDSS LRG galaxy sample

The SDSS LRG sample selection was designed to optimize the effective survey volume in the range $k \sim 0.05 - 0.1 \, h$/Mpc (Tegmark et al., 2006):

$$V_{\text{eff}}(k) = \int \left[ \frac{\bar{n}(\vec{r}) P_g(k)}{1 + \bar{n}(\vec{r}) P_g(k)} \right]^2 d^3 r$$

(1.26)

where the error on $\Delta P_g(k) \propto V_{\text{eff}}(k)^{-1/2}$. Moreover, the SDSS LRGs have bias $b_{\text{LRG}} \approx 2$, which reduces the relative shot noise contribution per galaxy. Because the effective volume in the linear regime of the power spectrum is large compared to other currently available samples (SDSS MAIN galaxies, 2dFGRS), we focus on this sample in the development of our large scale structure techniques.

The targeting algorithm (Eisenstein et al., 2001) produces a sample that is approximately volume limited with $\bar{n} \approx 10^{-4} \, (\text{Mpc}/h)^{-3}$ out to $z \sim 0.36$, and then becomes flux limited to the maximum redshift $z = 0.55$ (Zehavi et al., 2005). LRGs often sit at the centers of groups and clusters (Ho et al., 2007). Our HOD modeling in Ch. 2 finds that only 6% of LRGs are satellite galaxies (i.e., not the central LRG in their halos).
1.4 Outline of the Thesis

The low number density of the LRGs suggests that LRG pairs occupying the same dark matter halo can be separated from pairs occupying distinct dark matter halos with high fidelity. We explore this possibility in Ch. 2 as we compare groups found by our Counts-in-Cylinders (CiC) method with groups occupying the same dark matter halos in our $N$ body simulation mock catalogs. We measure the multiplicity function of CiC groups in the SDSS LRG sample to constrain the parameters of their HOD.

In Ch. 3 our aim is to determine the optimal parameters for a set of $N$ body simulations from which we can generate high fidelity mock LRG catalogs. We compare a host of large and small scale clustering and velocity correlation statistics between three simulation boxes with the same initial conditions but different mass resolutions.

Despite the much larger effective volume (Eqn. 1.26) of the LRGs compared with the SDSS MAIN or 2dFGRS sample galaxies, when Verde & Peiris (2008) reconstruct the primordial power spectrum with minimal assumptions about its shape, they find that the LRG sample has less statistical power than the SDSS MAIN or 2dFGRS. Since the authors restrict themselves to $k \leq 0.1 \, h$/Mpc, they are unable to constrain both the nonlinear correction and the primordial power spectrum. Dunkley et al. (2008) find that the best fit $\Omega_m$ varies with $k_{\text{max}}$ when combining the LRG power spectrum with WMAP, and so they choose not to include the galaxy power spectrum in their cosmological parameter analysis that combines several independent data sets. Both difficulties stem from the large nonlinear correction to the LRG power spectrum, which is above the statistical error in each $k$-bin for $k \gtrsim 0.07 \, h$/Mpc. There are many potential improvements to the analysis in Tegmark et al. (2006) that could reduce the size of the nonlinear correction. First, the effect of the FOG compression algorithm employed in their analysis was not examined; we find that this procedure can add a significant amount of power, thus increasing the amplitude of the necessary nonlinear correction factor. Second, the nonlinear correction
fitting function, the ‘$Q_{nl}$’ model, was tested and calibrated on LRG mock catalogs in real space, neglecting all the complications of redshift space. Moreover, the halo occupation model used to generate their mock catalogs did not allow for any satellite galaxies. Armed with a set of high fidelity mock catalogs calibrated using the SDSS CiC group multiplicity function and observed power spectrum amplitude, in Ch. 4 we present measurements of the effects of the FOG algorithm of Tegmark et al. (2006) and show that our reconstructed halo density field using CiC groups has a much smaller nonlinear correction. We present a new nonlinear correction model calibrated on our high fidelity mock catalogs in redshift space which can be used in the analysis of the final power spectrum analysis of the SDSS LRGs.
Chapter 2

Constraining the LRG Halo
Occupation Distribution using
Counts-in-Cylinders

This chapter is based on work done in collaboration with David Spergel and has been submitted for publication to *The Astrophysical Journal*.

2.1 Introduction

The Sloan Digital Sky Survey (SDSS; York et al. (2000)) has recorded the largest sample of Luminous Red Galaxies (LRGs), probing a volume of $\sim 1 \ (h^{-1} \text{ Gpc})^3$ out to $z \sim 0.5$ (Eisenstein et al., 2001) and making it ideal for studying large scale structure. Understanding the small-scale relationship between the galaxy and dark matter density fields is essential to extracting the linear matter power spectrum from the galaxy power spectrum, even on very large scales (Schulz & White, 2006; Sánchez & Cole, 2008).

The Halo Occupation Distribution (HOD) is a popular and useful description of this relationship (Seljak, 2000; Peacock & Smith, 2000; Cooray & Sheth, 2002). Conroy et al.
(2007a) and White et al. (2007) have used this framework to constrain the rate at which LRGs merge or are disrupted in clusters. Zheng et al. (2007) constrain stellar mass growth between DEEP2 \((z \sim 1)\) and SDSS \((z \sim 0)\) galaxies using the HOD description, and Conroy et al. (2007b) employ HOD modeling to illuminate the fate of \(z \sim 2\) star-forming galaxies. Others researchers, e.g., Chen (2007) and Ho et al. (2007), use the HOD description to study the spatial distribution of satellite galaxies.

Several groups have used two and three point statistics (Blake et al., 2007; Kulkarni et al., 2007; White et al., 2007) as well as galaxy-galaxy lensing (Mandelbaum et al., 2006) to constrain the HOD of LRGs. Ho et al. (2007) have taken a more direct approach and used X-ray determined cluster masses to measure \(N_{\text{LRG}}(M)\). Though these analyses were performed on samples with different luminosity and redshift ranges, they offer seemingly conflicting results on both the slope \(\alpha\) at the high mass limit of the satellite term \(N_{\text{sat}} \sim M^\alpha\) and the fraction of LRGs that are satellite galaxies. Ho et al. (2007) find \(\alpha \approx 0.6\) when fitting the total LRG number \(N_{\text{tot}}(M) \propto M^\alpha\), Kulkarni et al. (2007) find \(\alpha = 1.4\), and Blake et al. (2007) find \(\alpha \sim 2.1 - 2.6\); Kulkarni et al. (2007) report a satellite fraction of \(\sim 17\%\), while Blake et al. (2007)'s redshift slices span 3-8%. The most luminous elliptical galaxies in Mandelbaum et al. (2006) have a satellite fraction of \(\lesssim 10\%\).

The low number density of the SDSS Luminous Red Galaxy (LRG) sample suggests that LRG pairs occupying the same dark matter halo can be separated from pairs occupying distinct dark matter halos with high fidelity. In this Chapter we explore that intuition, and show that one-halo pairs can be identified with \(\sim 75\%\) completeness and \(\lesssim 27\%\) contamination by simple cuts in the transverse separation \(\Delta r_\perp\) and LOS separation \(\Delta r_\parallel\). Furthermore, these pairs can be grouped together using a Friends-of-Friends (FoF) algorithm to estimate the LRG group multiplicity function. We apply this technique to a sample of LRGs from SDSS to constrain their HOD. We find that both the high values of \(\alpha \sim 2\) and high satellite fractions reported in previous papers are inconsistent with the \(0.16 < z < 0.36\) SDSS LRG
group multiplicity function measured here. In contrast to previous methods which rely on 2 and 3 point statistics to constrain the HOD (as in Kulkarni et al. (2007)), our method probes the HOD more directly by estimating the group multiplicity function from the higher order statistics in the LRG density field in the one-halo dominant regime.

We present an overview of the Counts-in-Cylinders (CiC) method in § 2.2.1 and apply it to an approximately volume limited subsample of SDSS LRGs in § 2.2.2, addressing the complications of fiber collisions, incompleteness, and complex angular masks. The CiC technique developed here requires calibration on mock galaxy catalogs. We summarize our N-body simulation parameters in § 2.2.3. § 2.2.4 presents the HOD model we employ throughout this analysis and details how we populate our simulations with galaxies. § 2.2.5 describes the Counts-In-Cylinders (CiC) technique to measure the LRG group multiplicity function and its calibration with simulations. The HOD parameters are fit using a maximum likelihood analysis explicated in § 2.2.6. In § 2.3 we present the CiC multiplicity function of our SDSS LRG subsample and describe the relation between the CiC and true group multiplicity functions. We present the constraints on the HOD parameters and their implications for the fraction of LRGs that are satellites, as well as the mass distribution of halos hosting LRG groups with $n_{\text{sat}}$ satellites. Mock catalogs produced using the CiC maximum likelihood HOD and a spherical overdensity (SO) halo catalog agree with the Masjedi et al. (2006) measurement of $w_p(r_p)$ for this sample when the large scale bias is adjusted with a single parameter for the central galaxy HOD. In § 2.3.5 we compare these results with a mock LRG catalog based on a FoF halo catalog and with other HOD measurements in the literature. We show that while the FoF and SO catalogs can both match the observed CiC multiplicity function, the FoF catalog produces mock catalogs with a deficit of halos at 1 Mpc/$h$ that is evident in the projected correlation function. In § 2.4 we comment on the strengths and weaknesses of the CiC method and summarize our conclusions in § 2.5.

Throughout this Chapter we adopt the Spergel et al. (2007) cosmological parameters
used in our simulations to convert redshifts to distances: \((\Omega_m, \Omega_b, \Omega_\Lambda, n_s, \sigma_8, h) = (0.26, 0.044, 0.74, 0.95, 0.77, 0.72)\). All distances and separations are in comoving coordinates.

### 2.2 Methodology

#### 2.2.1 Overview of the Method

The goal of this section is to measure the group multiplicity function of a subsample of the SDSS LRGs. For a complete spectroscopic sample covering the full sky (or in a periodic simulation box), our method is as follows:

- Identify ‘one-halo’ pairs of galaxies. In what follows, our criteria for two galaxies to be a one-halo pair is \(\Delta r_\perp \leq 0.8 \text{ Mpc}/h\) and \(\Delta r_\parallel = 20 \text{ Mpc}/h\) (equivalently \(\Delta z/(1+z) \leq \beta_{\text{max}} = 0.006\)), where both are comoving separations. These choices were motivated by results on mock LRG catalogs and will be discussed in more detail in later sections.

- Group pairs of galaxies into groups using a FoF algorithm. The number of groups with \(n_{\text{sat}}\) satellites, \(N_{\text{CiC}}(n_{\text{sat}})\) is the CiC multiplicity function.

In § 2.2.2 we present the technical details of accounting for the facts that the SDSS has boundaries and holes, that the spectroscopic sample of LRGs is incomplete, and that the SDSS cannot simultaneously take spectra of two objects separated by \(< 55''\), so that regions of the sky observed only once spectroscopically may have missing close pairs of LRGs.

We use the LRGs from the SDSS imaging sample to supplement the spectroscopic sample (Blanton et al., 2005; Adelman-McCarthy et al., 2007). We identify potential pairs from the imaging sample and calibrate this step using pairs of objects from the spectroscopic sample. Since most nonisolated LRGs are in groups of 2 and candidate pairs from the imaging sample neighboring more than one LRG are highly likely to be group members, we apply the small correction for false LRG pair detections to \(N_{\text{CiC}}(n_{\text{sat}} = 1)\). A more complex scheme involving corrections at each \(n_{\text{sat}}\) would not have enough statistics to calibrate on the spectroscopic sample. We eliminate from our sample LRGs close to the survey
boundary, though they are allowed to be grouped with LRGs away from the boundary. This ensures that our multiplicity function is not biased due to unobserved LRGs outside the boundary. However, since the bright star masks are numerous and individually very small, this approach is not practical for dealing with objects near bright star masks. Instead, we adjust $N(n_{sat})$ by estimating the probability that there is an LRG covered by each bright star mask, and then computing the change in $N(n_{sat})$ if there were one. Table 2.1 shows that all of these corrections are small.

### 2.2.2 Data

The SDSS (Stoughton et al., 2002; Abazajian et al., 2003) has imaged $10^4$ deg$^2$ in $u$, $g$, $r$, $i$, and $z$. From this sample, spectroscopic LRG targets are efficiently selected using two color/magnitude cuts (Eisenstein et al., 2001). The tiling algorithm ensures nearly complete samples (Blanton et al., 2003). However, spectroscopic fiber collisions prohibit simultaneous spectroscopy for objects separated by $< 55''$, leaving $\sim 7\%$ of targeted objects without redshifts (Masjedi et al., 2006). Overlapping plates on $\sim 1/3$ of the survey area mitigate this problem and permit us to calibrate this effect in our analysis, as detailed below. The 'photometric sample' as referred to below consists of objects from the imaging sample that were targeted as LRGs according to the color/magnitude cuts laid out in Eisenstein et al. (2001) but lack spectra. The 'spectroscopic sample' consists of objects from the imaging sample that were targeted as LRGs and subsequently observed.

The goal of this analysis is to measure the group multiplicity function for the spectroscopic LRG sample with $-23.2 < M_g < -21.2$ and $0.16 < z < 0.36$. This sample is approximately volume-limited with $\bar{n} \approx 9.7 \times 10^{-5} (h^{-1} \text{Mpc})^{-3}$, and Zehavi et al. (2005) and Masjedi et al. (2006) have measured the projected correlation function for this sample on small and intermediate scales. We begin with the entire set of LRG target galaxies for the DR4+ sample from the NYU Value-Added Galaxy Catalog (VAGC) (Blanton et al.,
2005; Adelman-McCarthy et al., 2007) so that our sample includes the complete sample of galaxies satisfying our $M_g$ and $z$ cuts. We correct our multiplicity function statistically for the inclusion of a small number of interlopers.

Our analysis requires the identification of all close pairs of LRGs in the sample, where close pairs satisfy $\Delta \theta \leq \theta_{max}(z)$ and $\Delta r_\parallel \leq \Delta r_{\parallel, max}$. The dominant source of LRG close pair incompleteness is from LRG-LRG fiber collisions; this results in a deficit of pairs separated by $<55''$ in the spectroscopic sample. LRG-main galaxy fiber collisions or spectroscopic incompleteness result in further pair incompleteness. Finally, the bright star mask occupies 1.88% of the survey area. We statistically correct for the unobserved LRGs in these regions.

Following Eisenstein et al. (2005) we cut sectors with $<60\%$ spectroscopic completeness. These sectors lie primarily along the boundary of the survey, and remove only 276 objects from our sample. The remaining survey area is 5564 deg$^2$, of which 104 deg$^2$ is covered by a bright star mask. Using inverse random catalogs from the VAGC to trace the survey geometry, we remove all objects within $\theta_{max}(z_{min} = 0.16)$ of the survey boundary, which amounts to 469 deg$^2$, or 8.6% of the total area. However, we keep track of all LRGs in the boundary, because they are allowed to form pairs with the objects remaining in our sample and therefore contribute to the group multiplicity function. Only 53 objects from the boundary were included in CiC groups. Note that while there are many corrections for incompleteness, they are all small and well-constrained.

Sample Model and Color/Magnitude cuts

There are 41721 objects in our spectroscopic sample passing the redshift, $M_g$, sector completeness, and boundary cuts, and 8167 objects from the photometric sample passing the same sector completeness and boundary cuts. To reduce the level of contamination in our LRG multiplicity function, we apply color and magnitude cuts to objects from the photo-
metric sample falling within $\Delta \theta_{\text{max}}$ of a spectroscopic LRG to select a high-fidelity sample of close pair candidates. Since we know the redshift of the proposed group from the spectroscopic neighbor, we use $c_\parallel$ (Eisenstein et al., 2001) as a redshift indicator and $r_{\text{pet}}$ of the photometric sample object as an absolute magnitude indicator at the spectroscopic redshift.

$$c_\parallel = 0.7(g - r) + 1.2(r - i - 0.18)$$ (2.1)

Using the colors of objects in our spectroscopic sample, we find the upper and lower limits of $c_\parallel$ that encompass 95% of the spectroscopic LRGs as function of redshift in bins of $\Delta z = 0.01$. This relation is shown in Fig. 2.1. For a candidate close pair match, we discard photometric sample objects with $c_\parallel$ falling outside this region, given the redshift of the spectroscopic group member. Since the k+e corrections described in Eisenstein et al. (2001) have already been applied to the spectroscopic sample, we estimate the expected $M_g$ of a photometric object at redshift $z$ of the spectroscopic group member by

$$M_{g,\text{photo}} = M_{g,\text{spec}} + (r_{\text{pet,photo}} - r_{\text{pet,spec}})$$ (2.2)

and disregard objects with $M_{g,\text{photo}} < -23.2$ or $M_{g,\text{photo}} > -21.2$.

For a random sample of LRG targets, 36.6% of them will pass the Eisenstein et al. (2001) LRG cuts as well as our sample’s magnitude and redshift cuts, based on the ratio in DR4+ for targeted LRGs that have spectra. The fraction of objects in our photometric sample that would pass our spectroscopic magnitude and redshift cuts will be slightly higher than for the random sample of LRG targets, since it includes fiber collision pairs very likely to be LRGs. We compute this fraction for the photometric sample at the end of the group-finding algorithm to circumvent this issue. We model the photometric sample as composed of $N_{\text{pass}}$ objects which would pass the absolute magnitude and redshift cuts of our sample, if they had spectra; the remaining objects we label as $N_{\text{fail}}$. A large fraction of objects in the $N_{\text{fail}}$ group are below our minimum redshift cut $z_{\text{min}} = 0.16$. Tests on targeted LRGs with spectra that fail our LRG subsample cuts show that the $N_{\text{fail}}$ group is well-approximated as uncorrelated with the spectroscopic LRG sample.
Figure 2.1: The mean $c_\| - z$ relation and bands including 95% of objects in each $\Delta z = 0.01$ bin. $c_\|$ (Eqn. 2.1) is used as a redshift indicator for objects targeted as LRGs but lacking spectra.
Candidate close pairs of spectroscopic LRGs from the photometric sample are naturally divided into two groups. The first, denoted ‘FB’ for ‘fiber’, includes objects \( \theta \leq 55'' \) from a spectroscopic object. These objects do not have spectra primarily due to a fiber collision with the neighboring spectroscopic LRG. The bulk of candidate close pairs fall into this category, and the contamination for such pairs by objects at different redshifts is low. The remaining close pairs have separations \( 55'' < \theta < \theta_{\text{max}}(z_{\text{spec}}) \), arise from the small overall incompleteness of the survey or fiber collisions with MAIN galaxies, and are denoted ‘INC’. The probability of contamination is larger for this type, but still manageable. We use all targeted LRGs with spectra but failing our \( M_g \) or \( z \) cuts to compute an average rate at which an \( N_{\text{fail}} \) object will pass our color/magnitude cut on the photometric sample by comparison with the spectroscopic LRG pairs in each type of collision: \( p_{\text{pass,FB}} = 0.078 \) and \( p_{\text{pass,INC}} = 0.070 \). The rate is slightly lower for INC collisions because they are more likely to be at a lower redshift where \( c_{||} \) is more discriminating, since \( \theta_{\text{max}}(z) \) decreases with \( z \). The number of close pair contaminants we expect from this sample for our two types of collisions are

\[
N_{FB,\text{fail}} = \frac{A_{FB}}{A_{\text{survey}}} \times N_{\text{fail}} \times p_{\text{pass}} \approx 2.1
\]

\[
N_{INC,\text{fail}} = \frac{A_{LRG} - A_{FB}}{A_{\text{survey}}} \times N_{\text{fail}} \times p_{\text{pass}} \approx 26.8
\]

where \( A_{FB} \) and \( A_{LRG} \) are the total areas enclosed by annuli of \( 55'' \) and \( \theta_{\text{max}}(z_{\text{spec}}) \) around each LRG, and \( A_{\text{survey}} \) is the total survey area after removal of the boundary. \( N_{\text{fail}} \) was estimated at the end of the group-finding algorithm; see below.

To estimate the completeness and contamination for the \( N_{\text{pass}} \) sample, we rely on observed collisions from the spectroscopic sample. For FB collisions, we found 295.5 collisions between objects in the spectroscopic sample that passed the color/magnitude cuts. 276 of these passed the \( \Delta z \) cut, so 93.4% of objects satisfying \( \Delta \theta \leq 55'' \) and passing the color/magnitude cuts would be considered one-halo pairs if the redshift were known. Fur-
thermore, 19 pairs failed the color/magnitude cuts but passed the $\Delta z$ cut. Therefore, the sample is 93.5% complete after the color/magnitude cuts and 6.6% contaminated. For INC collisions, we found 2494 collisions passing the photometric color/magnitude cuts; 2020 of these passed the $\Delta z$ cuts, so would be assigned as pairs. 110.5 pairs passing the $\Delta z$ cuts failed the color magnitude cuts. Therefore, this collision sample should be 94.8% complete and 19.0% contaminated.

Our approach is to estimate the number of interlopers and apply the correction to the number of groups of 2 LRGs. 568 objects from the photometric sample are $\leq 55''$ from a spectroscopic LRG object and pass the color/magnitude cuts at the spectroscopic redshift; 250 objects from the photometric sample are $55'' < \theta < \theta_{\text{max}}(z)$ and pass the color/magnitude cuts at the spectroscopic redshift. Based on the rates measured from spectroscopic collisions and after correcting for the expected number of interlopers $N_{FB, fail} + N_{\text{INC, fail}}$, we expect 709.3 of the remaining 789.1 collisions to be ‘true’ collisions (i.e., would pass the $\Delta z$ criterion as well, if the redshift were measured), and we expect to have missed 46.3 true collisions due to our color/magnitude cuts. Therefore, we overestimate the total number of pairs by 62.4 (a 2.8% correction).

Finally, we must estimate the total number of isolated LRGs (i.e., those with no neighboring LRGs passing our CiC cuts) missing from our sample due to the incompleteness of the spectroscopic sample. Again making use of the assumption that objects in the photometric sample that would fail the color/magnitude cuts of our LRG sample are uncorrelated with the spectroscopic sample and that the number of expected interlopers is negligible (Eqns. 2.3 and 2.4), we expect the total number of isolated galaxies in the photometric sample to be

$$N_{\text{photo, iso}} = N_{\text{pass}}(1 - p_{\text{group}}) + N_{\text{fail}},$$ (2.5)

where $p_{\text{group}}$ is the probability that an LRG is in a CiC group with $n_{\text{sat}} \geq 1$. Using the observable $N_{\text{photo, iso}}$, the number of objects from the photometric catalog not grouped with spectroscopic objects, along with the pass rate for a random set of LRG targets of 36.6%,
allows us to solve for the ratio $N_{pass}/N_{fail}$ in the photometric sample. Given the final estimate of the group multiplicity function (see below), we find $p_{\text{group}} = 0.132$ and the isolated LRG contribution from the photometric sample, $N_{pass} * (1 - p_{\text{group}})$, is 2900. This produces an increase in total objects in our sample of 8%, in agreement with the incompleteness rate reported by Masjedi et al. (2006).

**Group-finding Algorithm and Bright Star Mask Corrections**

One-halo pairs of galaxies are assigned by satisfying the criteria $r_{\perp} \leq r_{\perp,\text{max}}$ and $\Delta z/(1 + z) \leq \beta_{\text{max}}$. Galaxies are then grouped by a FoF algorithm. The challenges in this section are to self-consistently incorporate the photometric sample into the calculation of the FoF group multiplicity function, and to marginalize over tiny Bright Star Mask holes in the survey geometry.

The group finding algorithm first assigns candidate pairs for each galaxy. As described in § 2.2.2, the redshift space criterion for a spectroscopic-photometric pair becomes $c_{\parallel,\text{min}}(z_{\text{spec}}) \leq c_{\parallel,\text{photo}} \leq c_{\parallel,\text{max}}(z_{\text{spec}})$ and $-23.2 < M_{g,\text{photo}} < -21.2$. Using these pair assignments, candidate groups are formed with the FoF algorithm. Spectroscopic neighbors of photometric objects in the group are kept if they satisfy the $\Delta r_{\parallel}$ requirement with at least one spectroscopic object in the group, and photometric neighbors of photometric objects in the group are kept if they satisfy the redshift space criterion with at least one spectroscopic object in the group. We correct the group assignments for the 23 photometric objects in two distinct groups. We first give preference to groups in which the photometric object is a fiber-collision pair; otherwise, we assign the photometric object to the larger group. We must then recompute the membership of the group losing the photometric object.

The last step of the group-finding algorithm accounts for Bright Star Mask and counts the number of isolated photometric objects, $N_{\text{photo,iso}}$ used in Eqn. 2.5. The Bright Star
Mask catalog, consisting of the \textbf{ra, dec}, and radius for each circular masked region, was generated from the Tycho2 catalog as described by the NYU VAGC (Blanton et al., 2005). The Bright Star Mask covers 1.88\% of the total survey region. Using the observed pair counts as a function of radius, we estimate the probability per unit area of finding a 1-halo pair satisfying $r_{\perp} \leq r_{\perp,max} = 0.8$ Mpc/$h$ and $\Delta z/(1 + z) \leq \beta_{max}$ as a function of $r_{\perp}$. The result is well-fit with a power law: $dP/dA = 0.027r_{\perp}^{-1.08} (\text{Mpc}/h)^{-2}$. $dP/dA$ is similar to the projected correlation function, except that it does not subtract the uncorrelated contribution, and it only includes pairs meeting our $\Delta r_{\parallel}$ criterion. For each bright star mask within $r_{\perp,max}$ of an LRG, we evaluate the probability that an unobserved LRG resides under the Bright Star Mask by integrating our power law fit to $dP/dA$ over the area of overlap between the Bright Star Mask and the LRG annulus satisfying $r_{\perp} \leq r_{\perp,max}$. We assign the probability of a given Bright Star Mask to cover an LRG as the maximum probability assigned from all the observed LRG neighbors of the Bright Star Mask elements. We modify the final group multiplicity function to account for each possible unobserved LRG, including its ability to bridge two previously existing groups. Therefore a single group of LRGs can contribute to many group multiplicity bins with weights according to the probability of a covered LRG beneath each Bright Star Mask intersecting a group member. This procedure adds 79.8 net nonisolated LRGs to the group multiplicity function; the rest of the expected $0.0188 \times N_{\text{LRG,observed}}$ are added to the isolated LRG count. See Table 2.1 for the final multiplicity function along with the contributions of each of these corrections.

Finally, we place an upper limit on the contribution from photometric-object only groups. As listed in the final column in Table 2.1, we find 64 photometric object pairs and a single group of 3 photometric objects for which the $c_{\parallel}$ and $r_{\text{petro}}$ for each galaxy in the group fall into the 95\% range of the same redshift bin of Fig. 2.1. 30\% of such pairs fall into the $z = 0.35 - 0.36$ bin, for which $c_{\parallel}$ is the least discriminating, while only 7.8\% of the spectroscopic LRGs lie in this redshift range. Therefore, we conclude that we miss a negligible number of groups by ignoring photometric object only groups.
Table 2.1: This table lists the CiC group multiplicity function of the SDSS LRG spectroscopic + photometric sample for $\Delta r_{\perp, \text{max}} = 0.8\, \text{Mpc}/h$ and $\beta_{\text{max}} = 0.006$. Column 1 lists the number of satellites in the group, $n_{\text{sat}} = n_{\text{group}} - 1$. Column 2 contains the raw group multiplicity before any corrections have been made (including for double counting of photometric objects). Column 3 lists the final group multiplicity function after all the corrections of § 2.2.2. Column 4 shows the number of photometric objects in groups with $n_{\text{sat}}$ satellite galaxies, Column 5 shows the net effect of the Bright Star Mask correction, and Column 6 shows the number of groups of photometric objects only with $n_{\text{sat}}$ satellites satisfying the criteria listed in § 2.2.2. These groups are largely false detections and do not contribute to the counts in Column 2. The total area of the survey, excluding the boundary region but including the bright star mask area, is 1.549 steradians.

<table>
<thead>
<tr>
<th>$n_{\text{sat}}$</th>
<th>$N_{\text{raw}}(n)$</th>
<th>$N_{\text{final}}(n)$</th>
<th>$n_{\text{photo}}$</th>
<th>$\Delta N_{\text{BSMASK}}$</th>
<th>$N_{\text{photo,only}}$</th>
<th>$N_{\text{CiC,mocks}}(n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>36778</td>
<td>40407.56</td>
<td>2907.71</td>
<td>702.49</td>
<td>0</td>
<td>40546.1</td>
</tr>
<tr>
<td>1</td>
<td>2314</td>
<td>2301.12</td>
<td>612</td>
<td>63.54</td>
<td>64</td>
<td>2190.4</td>
</tr>
<tr>
<td>2</td>
<td>281</td>
<td>285.86</td>
<td>146</td>
<td>6.86</td>
<td>1</td>
<td>323.2</td>
</tr>
<tr>
<td>3</td>
<td>56</td>
<td>54.29</td>
<td>38</td>
<td>1.29</td>
<td>0</td>
<td>65.8</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>20.20</td>
<td>14</td>
<td>0.20</td>
<td>0</td>
<td>15.2</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>4.13</td>
<td>5</td>
<td>0.13</td>
<td>0</td>
<td>4.7</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1.05</td>
<td>2</td>
<td>0.05</td>
<td>0</td>
<td>1.5</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1.02</td>
<td>1</td>
<td>0.02</td>
<td>0</td>
<td>0.2</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0.02</td>
<td>0</td>
<td>0.02</td>
<td>0</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Note on average $\bar{n}_{\text{LRG}}$

Zehavi et al. (2005) construct a model of the spectroscopic LRG sample and estimate $\bar{n}_{\text{LRG}} = 9.7 \times 10^{-5} \, (\text{Mpc}/h)^{-3}$ for the $0.16 < z < 0.36$, $-23.2 < M_g < -21.2$ subsample. This is in good agreement with our estimate $N_{\text{LRG}}/V = 1.0 \times 10^{-4} \, (\text{Mpc}/h)^{-3}$ where V is the product of the fraction of the sky covered by our sample and the comoving volume between $z = 0.16$ and $z = 0.36$. As discussed in the previous sections, the number of LRGs in our sample is increased by 8% when including objects from the imaging sample and covered by Bright Star Masks. For our choice of cosmological parameters, the volume of the $0.16 < z < 0.36$ shell is 5% larger than in the Zehavi et al. (2005) cosmology, so a difference of 3% in $\bar{n}_{\text{LRG}}$ is expected.
2.2.3 Simulations and Halo Catalogs

In order to constrain the LRG HOD, we require mock LRG catalogs to calibrate the relationship between the CiC group multiplicity function and the true HOD multiplicity function. The mock LRG catalogs are derived from halo catalogs from a $z = 0.2$ N-body simulation snapshot using the HOD formalism detailed in § 2.2.4. We use a $1024^3$ particle TPM (Bode & Ostriker, 2003) $1 (h^{-1} \text{Gpc})^3$ simulation with particle mass $M_p = 6.72 \times 10^{10} M_\odot$ and $\epsilon = 16.28 h^{-1} \text{kpc}$ described in more detail in Sehgal et al. (2007). The cosmological parameters are set to $(\Omega_m, \Omega_b, \Omega_\Lambda, n_s, \sigma_8, h) = (0.26, 0.044, 0.74, 0.95, 0.77, 0.72)$. For our FoF catalogs (Davis et al., 1985), all halos containing $\geq 10^7$ particles were identified using a linking parameter $b = 0.2$. The virial mass and radius for each halo are measured in spheres with overdensity determined by the spherical top hat collapse model (Bryan & Norman, 1998). We also find the radius $r_{v_{\text{max}}}$ at which the circular velocity is maximum in order to estimate the halo concentration $c_{\text{vir}}$. Eqn. 11 in Lokas & Mamon (2001) implies

$$c_{\text{vir}} = 2.16 r_{\text{vir}} / r_{v_{\text{max}}}.$$

We produce catalogs using the spherical overdensity (SO) halo finder with $\Delta = 200 \rho_b$ described in Tinker et al. (2008) and code kindly provided by J. Tinker. We record halos down to 50 particles. Ch. 3 shows that the clustering and velocity statistics of halos containing 50-80 particles are not altered by the mass resolution. The mass function in this mass range is overestimated by $\sim 5\%$ and unaffected in higher mass bins. As we demonstrate in later sections, only a small fraction of LRGs occupy halos in the lowest mass bin for our best fit HOD, so the effect of simulation resolution on the HOD parameters should be minimal.

2.2.4 HOD model

The Halo Occupation Distribution (HOD) model assumes that the probability $P(N_{\text{LRG}}|M)$ of $N_{\text{LRG}}$ LRGs occupying a dark matter halo of mass $M$ at redshift $z$ depends only on the halo mass (for a review, see Cooray & Sheth (2002)). However, the application of the HOD
formalism requires that we make several further assumptions about $P(N_{\text{LRG}}|M)$. Detailed studies of dark matter halos and subhalos suggest a division of galaxies into central and satellite galaxies (Kravtsov et al., 2004). The central galaxies are assumed to sit at the halo center, consistent with the observation that most ($\sim 80\%$) of the brightest cluster LRGs are found within $0.2r_{\text{vir}}$ of the center of the cluster potential well as traced by X-rays. Satellite galaxies occur in the more massive halos already containing a central galaxy. In high resolution simulations they can be directly associated with dark matter subhalos (Vale & Ostriker, 2006); here we will assume they have the same distribution as the halo dark matter. We will use these functional forms with the five free parameters $M_{\text{min}}, \sigma_{\log M}, M_1, M_{\text{cut}},$ and $\alpha$ to describe the mass dependence of the average halo occupation as a function of halo mass $M$:

$$\langle N(M) \rangle = \langle N_{\text{cen}} \rangle (1 + \langle N_{\text{sat}} \rangle)$$

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

$$\langle N_{\text{sat}} \rangle = \left( \frac{M - M_{\text{cut}}}{M_1} \right)^\alpha$$

This form for $N_{\text{cen}}(M)$ was suggested in Zheng et al. (2005) and adopted in the analysis of Blake et al. (2007). Both Blake et al. (2007) and Kulkarni et al. (2007) use Eqn. 2.9 with $M_{\text{cut}} = 0$. As shown below, we find evidence for $M_{\text{cut}} > 0$. Half of halos with mass $M_{\text{min}}$ host central LRGs, and $\sigma_{\log M}$ quantifies the width of the transition from $N_{\text{cen}}(M) = 0$ to $N_{\text{cen}}(M) = 1$. $M_1$ sets the mass scale at which satellite galaxies become probable, and $M_{\text{cut}}$ sets a cut-off below which halos do not host satellites. In the case $M \gg M_{\text{cut}}$ and $M_1 \gg M_{\text{min}}$, we expect $\alpha = 1$, or the number of satellite galaxies to be proportional to the halo mass.

$P(N_{\text{LRG, sat}}|N(M))$ is assumed to be Poisson distributed; this assumption is supported by the distribution of subhalo counts in simulations (Kravtsov et al., 2004) as well as observations (Lin et al., 2004; Ho et al., 2007). Often two-point statistics are used to fit the HOD; they depend only on the first and second moments of $P(N_{\text{LRG}}|M)$. The technique
we present here to constrain the HOD, Counts-In-Cylinders, makes use of higher-order statistics in the galaxy distribution. This technique has excellent constraining power for the parameters of $N_{sat}$, but is dependent upon the accuracy of the Poisson assumption for deriving accurate HOD parameters. Using the Poisson assumption, the expected number of halos with $N_{sat} = n$ satellites is given by

$$\langle N(N_{sat} = n) \rangle = \int dM \ n_{halo}(M) \ exp(-N_{sat}(M)N_{cen}(M)) \ \frac{(N_{sat}(M)N_{cen}(M))^n}{n!}. \ (2.10)$$

Eqn. 2.10 is central to our maximum likelihood analysis described in § 2.2.6.

The CiC method constrains only the HOD parameters in $N_{sat}(M)$. In this work we fix the $N_{cen}$ parameters $M_{min}$ and $\sigma_{logM}$ by matching the observed $\bar{n}$ and amplitude of the projected correlation function in the 2-halo regime.

**Populating the Simulations**

Halos are populated with a central galaxy with probability $N_{cen}(M)$. Central galaxies are placed at the center of their host halos and assigned the peculiar velocity of their halos. Halos with a central galaxy are populated with $N_{sat}$ galaxies, where $P(N_{sat}|N(M))$ is drawn from a Poisson distribution. Our parameter constraints are derived using SO halo catalogs. For those, the position and velocity of the satellite galaxies are taken to be that of a randomly selected dark matter particle halo member. For mock catalogs based on the FoF halos, the satellite galaxies are independently distributed following an NFW profile with concentration of the dark matter halo determined by Eqn. 2.6. The peculiar velocity of a satellite galaxy is the sum of the halo peculiar velocity and a random velocity drawn from a Gaussian distribution determined by the virial velocity of the halo (Lokas & Mamon, 2001):

$$\sigma^2_{vir,1D} = \frac{GM_{vir}}{2R_{vir}}. \quad (2.11)$$
We assign comoving redshift space position \( s \) to an object in our mock catalogs using the conversion at \( z_{\text{box}} = 0.2 \):

\[
s = x_{\text{LOS}} + (1 + z_{\text{box}})v_p/H(z_{\text{box}})
\]

(2.12)

where \( x_{\text{LOS}} \) is the comoving distance along the line of sight in real space.


\section*{2.2.5 Counts-In-Cylinders Technique}

The hypothesis underlying the Counts-In-Cylinders technique to constrain the LRG HOD is that 1-halo and 2-halo LRG pairs are separable based on their relative angular and redshift space positions. In the regime of small separations where the 1-halo term dominates, a cylinder should be a good approximation to the density contours surrounding central galaxies, as long as the satellite velocity is uncorrelated with its distance from the halo center, and the relative velocity dominates the separation of central and satellite objects in the redshift direction. Based on our initial analysis of completeness and contamination of mock catalogs derived from FoF halos, we set \( r_{\perp,\text{max}} = 0.8 \) Mpc/\( h \) and \( z_{\text{max}} = 20 \) Mpc/\( h \) for our \( z = 0 \) catalogs. \( r_{\perp,\text{max}} \) is set by the typical comoving size of halos hosting satellite galaxies, and \( z_{\text{max}} \) is set by the amplitude of the velocity dispersion in halos massive enough to host satellite galaxies. In later work we plan to improve the fidelity of our CiC group identification. However, the choice made here is sufficient since we calibrate the relatively small bias in the method using mock catalogs.

Finally, we must select a redshift dependence for the parameters \( \Delta r_{\perp,\text{max}} \) and \( \Delta r_{||,\text{max}} \). The virial radius of a halo of fixed mass in \emph{comoving} coordinates decreases by \( < 6\% \) over our SDSS sample redshift range; this decrease will be offset as massive halos hosting LRGs accrete and grow. We approximate the net result by fixing the transverse separation in comoving coordinates. The virial velocity of a halo of fixed mass varies by a similar fraction, but again we expect it to grow with halo mass. We keep the effective maximum relative velocity fixed as we translate the cylinder parameters to different redshifts. For an object
at $\chi_o = \chi(z_d)$, the observed redshift will be

$$z_{\text{obs}} = z_d + (1 + z_d)v_p/c$$  \hspace{1cm} (2.13)$$

so for pairs in massive halos where $\Delta z_d$ is relatively small, $\Delta z_{\text{obs}}/(1 + z_{\text{obs}}) \approx \Delta v_p/c$.

### 2.2.6 Maximum Likelihood Parameter Estimation

While the main purpose of this technique is to produce mock catalogs with higher-order statistics in agreement with the observed sample (specifically the CiC multiplicity function), in the process we derive constraints on HOD parameters. These results are dependent on the input mass function, which we take directly from our simulation. The HOD model we are constraining predicts the expected number of groups containing $N_{\text{sat}}$ satellite LRGs for each positive integer $n = N_{\text{sat}}$ via Eqn. 2.10. The CiC technique presented here produces a group multiplicity function which may have both an offset and scatter around the true number of groups with $n_{\text{sat}}$ satellites,

$$\Delta g(n_{\text{sat}}) = N_{\text{CiC}}(n_{\text{sat}}) - N_{\text{true}}(n_{\text{sat}}),$$  \hspace{1cm} (2.14)$$

where by ‘true’ groups we mean groups of LRGs occupying the same halo. The overall offset reflects both the incompleteness of CiC method for finding pairs of LRGs in the same halo, as well as the contamination from LRGs in other halos. The scatter $\Delta g(n_{\text{sat}})$ arises both from the stochasticity of the $N_{\text{LRG}} - M_{\text{halo}}$ relation and cosmic variance of halo alignments that cause pairs of galaxies in separate halos to be grouped together by CiC. Using our mock catalogs we have estimated $P(\Delta g(n_{\text{sat}}))$ for each $n_{\text{sat}}$; see § 2.3.2 for more details.

While in principle $\Delta g(n_{\text{sat}})$ should be measured at each HOD point, this is not feasible since one must make many mock catalogs in order to estimate the distribution of $\Delta g(n_{\text{sat}})$. As we show below, the tightest parameter constraints from our CiC group multiplicity measurement are on $M_1$ and $\alpha + M_{\text{cut}}/(10^{14}M_\odot)$. Examination of $\Delta g(n_{\text{sat}})$ over a subset of points around the maximum likelihood HOD indicates that the variance of $\Delta g(n_{\text{sat}})$ is
much larger than any systematic variation of $\Delta g(n_{\text{sat}})$ with HOD parameters for $n_{\text{sat}} = 2, 3$. There may be slight trends at larger $n_{\text{sat}}$. This may be because the CiC parameters adopted in this thesis miss a significant fraction of large groups, so that a better model for the offset between $N_{\text{CiC}}$ and $N_{\text{true}}$ would be $N_{\text{CiC}}(n_{\text{sat}}) = N_{\text{true}}(n_{\text{sat}}) * b(n_{\text{sat}})$. We chose not to adopt this parametrization because at large $n_{\text{sat}}$, either multiplicity function can be 0 and lead to ambiguities in $b(n_{\text{sat}})$. We hope to improve the accuracy of the CiC technique at $n_{\text{sat}} > 2$ in future work. However, we expect this effect to have only a small impact on the final parameter constraints.

We assume that for each value of $n_{\text{sat}}$, the number of groups with $n_{\text{sat}}$ satellites, $j = N_{\text{CiC}}(n_{\text{sat}}) - \Delta g(n_{\text{sat}})$, is Poisson-distributed about the expectation value $\mu_n$ in Eqn. 2.10:

$$P(j = N_{\text{CiC}}(n_{\text{sat}}) - \Delta g(n_{\text{sat}}) | \mu) = e^{-\mu_n} \frac{\mu_n^j}{j!}$$

(2.15)

and independent of the other observed and expected multiplicities. Next we marginalize over the distribution of $\Delta g(n_{\text{sat}})$ which we measured from our mock catalogs:

$$P(N_{\text{CiC}}(n_{\text{sat}})|\mu_n) = \int P(j = N_{\text{CiC}}(n_{\text{sat}}) - \Delta g(n_{\text{sat}}) | \mu) P(\Delta g) d\Delta g.$$

(2.16)

Then the probability of observing the set of CiC multiplicities $\tilde{N}_{\text{CiC}} = \{N_1, N_2, ..., N_{k_{\text{max}}}\}$ given the expectation values $\bar{\mu} = \{\mu_1, \mu_2, ..., \mu_{k_{\text{max}}}\}$ (with the subscript denoting the number of satellites in the group) is

$$P(\tilde{N}_{\text{CiC}} | \bar{\mu}) = \Pi_{n=1}^{k_{\text{max}}} P(N_n | \mu_n)$$

(2.17)

In order to make use of Bayes’ Theorem to find the parameters $\bar{p}$ that maximize the likelihood of the observations $\tilde{N}_{\text{CiC}}$,

$$P(\bar{p} | \tilde{N}_{\text{CiC}}) = \frac{P(\tilde{N}_{\text{CiC}} | \bar{p})P(\bar{p})}{P(\tilde{N}_{\text{CiC}})}$$

(2.18)

we must assume a prior $P(\bar{p})$. We choose a flat prior on $\alpha, M_1, \text{ and } M_{\text{cut}}$, except that they are required to be positive.

Finally, we note that the $N_{\text{CiC,SDSS}}(n_{\text{sat}}) = 0$ for $n_{\text{sat}} > 9$. In Eqn. 2.17 we set $k_{\text{max}} = 18$. 

The exact cutoff is unimportant since \( n_{\text{sat}} \leq 9 \) multiplicities already constrain the HOD parameter space to one in which \( N(n_{\text{sat}}) \) is sharply falling.

In Figures 2.5 through 2.7 we compute the expectation of several quantities as

\[
\langle y \rangle = \int y(\bar{p}) P(\bar{p}|N_{CiC}) d\bar{p}
\]

(2.19)

\[
\Delta x = \sqrt{\langle (x - \langle x \rangle)^2 \rangle}.
\]

(2.20)

Eqn. 2.20 is used to compute the error bars in those figures. Confidence intervals for some function \( y(\bar{p}) \) of the HOD parameters given in § 2.3.3 are computed directly from the marginalized distribution:

\[
P(y') dy' = \int_{y(\bar{p}) \in [y', y'+dy']} P(\bar{p}|N_{CiC}) d\bar{p}.
\]

(2.21)

### 2.3 Results

We use the SDSS LRG \( N_{CiC} \) to evaluate the likelihood of a given HOD model using Eqn. 2.18. However, \( M_{\text{min}} \) and \( \sigma_{\log M} \) are still unconstrained. We choose \( \sigma_{\log M} = 0.7 \) for the analysis that follows, and set \( M_{\text{min}} \) so that the number density of LRGs in the mock catalog matches that of our SDSS sample, \( \bar{n}_{\text{LRG}} = 1.0 \times 10^{-4} \text{ (Mpc/}h^3 \text{). This choice of } \sigma_{\log M} = 0.7 \text{ provides excellent agreement with the observed projected correlation function; see § 2.3.4 and Figure 2.8.}

#### 2.3.1 SDSS LRG \( N_{CiC} \)

Column 3 of Table 2.1 presents our final estimation of \( N_{CiC}(n_{\text{sat}}) \) for \( n_{\text{sat}} = 0 - 8 \), the group multiplicity function for our \( 0.16 < z < 0.36, -23.2 < M_g < -21.2 \) subsample of SDSS LRGs.
2.3.2 CiC offset $\Delta g(n_{\text{sat}})$ and Group-Finding Accuracy

To characterize the offsets $\Delta g(n_{\text{sat}})$ (Eqn. 2.14) between the CiC group multiplicity function and the true multiplicity function, we produce 600 mock catalogs using the HOD parameters $\sigma_{\log M} = 0.7$, $M_{\text{min}} = 8.05 \times 10^{13} M_{\odot}$, $M_{\text{cut}} = 4.66 \times 10^{13} M_{\odot}$, $M_1 = 4.95 \times 10^{14} M_{\odot}$, and $\alpha = 1.07$, which are close to the final maximum likelihood parameters reported in § 2.3.3; the two produce very similar CiC group multiplicity functions. We measured $\Delta g(n_{\text{sat}})$ in a randomly selected cubic subsample of each mock catalog, selected to have a volume equal to that of our SDSS LRG subsample, $0.46 (\text{Gpc}/h)^3$. The width of $P(\Delta g)$ may be underestimated, since each subsample is drawn from the same dark matter simulation. However, the broadness of $N_{\text{cen}}(M)$ means that our simulation box contains many more halos than LRGs, and so the variance resulting from cosmic structures is somewhat sampled. The average and variance of $\Delta g(n_{\text{sat}})$ are reported in Table 2.2. Comparison to the observed $N_{\text{CiC}}(n_{\text{sat}})$ in Table 2.1 shows that the width of the $\Delta g$ distribution produces comparable uncertainty in $\langle N_{\text{true}}(n_{\text{sat}}) \rangle$ as the Poisson sampling. Therefore, neglecting the integral in Eqn. 2.16 would cause an underestimation of the errors on the HOD parameters. In our likelihood calculation we use the histogram of $\Delta g(n_{\text{sat}})$ values to estimate $P(\Delta g(n_{\text{sat}}))$ in Eqn. 2.16.

Table 2.2 also reports a measure of the completeness and contamination with which we find groups of size $n_{\text{sat}}$. Column 5 is the fraction of true one-halo groups with $n_{\text{sat}}$ satellites that are exact matches to a CiC group. This is a rather stringent definition, since it excludes groups with only one missing satellite of many, or with only one galaxy from a different halo. Column 6 shows the fraction of CiC groups that do not exactly match a true one-halo group. Our optimization mainly focused on groups with only one satellite, since they are most numerous. Moreover, the CiC parameters were established while still using FoF halo catalogs; there are non-negligible differences in the two-halo term between the two catalogs (see §2.3.5). Therefore, we are still optimistic that we can improve the accuracy of our method at $n_{\text{sat}} > 1$, which we will address in a later paper. The issue is of less concern in the current work, since we have calibrated the observed statistics on our mock catalogs.
Table 2.2: Column 2 shows the average offset $\Delta g(n_{\text{sat}}) = N_{\text{CiC}}(n_{\text{sat}}) - N_{\text{true}}(n_{\text{sat}})$ over 600 mock catalogs for HOD parameters $\sigma_{\log M} = 0.7$, $M_{\text{min}} = 8.05 \times 10^{13} M_\odot$, $M_{\text{cut}} = 4.66 \times 10^{13} M_\odot$, $M_1 = 4.95 \times 10^{14} M_\odot$, and $\alpha = 1.07$, which are close to the final maximum likelihood parameters reported in § 2.3.3. Column 3, $\sigma_g^2$, is our estimate of the variance of $\Delta g$ in a subsample of the simulation box with volume equal to our SDSS subsample, 0.46 (Gpc$/h)^3$. Comparison with Table 2.1, Column 2 (reproduced here in Column 4) shows that this variance is comparable to the expected variance from Poisson sampling of $(N_{\text{CiC}}(n_{\text{sat}}))$. Column 5 is the fraction of true one-halo groups with $n_{\text{sat}}$ satellites that are exact matches to a CiC group. Column 6 shows the fraction of CiC groups that do not exactly match a true one-halo group.

Large biases are unlikely as long as the mock catalogs are reproducing the salient features of the small-scale clustering and distribution of satellites within dark matter halos.

2.3.3 HOD Constraints

With the width of the $N_{\text{cen}}(M)$ distribution fixed at $\sigma_{\log M} = 0.7$, we compute the likelihood at each point in the three dimensional space of $N_{\text{sat}}$ parameters $\alpha$, $M_{\text{cut}}$, and $M_1$ according to Eqns. 2.15 - 2.18. We evaluate $M_{\text{min}}$ at each HOD point so that $\bar{n}_{\text{LRG}}$ remains fixed. The maximum likelihood HOD parameters and marginalized one-dimensional 68% and 95% confidence intervals are $M_{\text{cut}} = 5.0^{+1.5}_{-1.3}(^{+2.9}_{-2.6}) \times 10^{13} M_\odot$, $M_1 = 4.95^{+0.37}_{-0.26}(^{+0.70}_{-0.53}) \times 10^{14} M_\odot$, and $\alpha = 1.035^{+0.10}_{-0.17}(^{+0.24}_{-0.31})$, with $M_{\text{min}} = 8.05 \times 10^{13} M_\odot$ at the maximum likelihood point. In Figures 2.2, 2.3, and 2.4 we show two dimensional likelihood contours for $\Delta L = \{-1.15, -3.09, -5.9\}$ after marginalizing over the remaining parameter.
For a $\chi^2$ distribution, these contour values correspond to 1, 2, and 3$\sigma$ confidence regions. We find a strong degeneracy in the $M_{\text{cut}} - \alpha$ plane. We tightly constrain the fraction of galaxies that are satellites, $f_{\text{sat}} = 0.0636^{+0.0019}_{-0.0020}(^{+0.0038}_{-0.0039})$ and the parameter combination $\alpha + M_{\text{cut}}/(10^{14} M_\odot) = 1.53^{+0.039}_{-0.047}(^{+0.080}_{-0.090})$. In the orthogonal direction, we find only a weak constraint: $M_{\text{cut}}/(10^{14} M_\odot) - \alpha = -0.54^{+0.32}_{-0.22}(^{+0.60}_{-0.49})$. In Column 3 of Table 2.3 we list the number of groups with $n_{\text{sat}}$ satellites averaged over 20 mock catalogs evaluated at our maximum likelihood HOD, which show good agreement with the SDSS CiC multiplicity function; recall that the broad distribution in $N_{\text{CiC}} - N_{\text{sat}}$ introduces noise in addition to the Poisson term.

In Figures 2.5 through 2.7, we show the implications of these results for the distribution of both satellite galaxies as a whole and groups containing $n_{\text{sat}}$ satellites as a function of halo mass. Figure 2.5 shows the mean and rms (Eqns. 2.19 and 2.20) of the expected number of satellites, $\langle N_{\text{cen}}(M) N_{\text{sat}}(M) \rangle$ for several halo masses. $N_{\text{sat}}(M)$ is less well constrained at the high mass end, where the number density of halos is low. Furthermore, we have not included cosmic variance of halo counts in this analysis; we have only used the mass function from our simulation. By comparison with Figure 2.7, we see that $N_{\text{cen}}(M) < 1$ even when $M$ is large enough to host satellite galaxies. Thus the tight constraints on the low mass end of $N_{\text{sat}}(M)$ are dependent on the accuracy of our parametrization of $N_{\text{cen}}(M)$. Figure 2.6 shows the probability that a halo of mass $M$ hosts $n_{\text{sat}}$ satellites. Again, the constraints are weaker as $M$ increases and $n(M)$ decreases. Figure 2.7 shows the distribution of all satellite galaxies as a function of halo mass, as well as the distribution of satellites in groups with $n_{\text{sat}} = 1, 2, 3$ and 4. The width of these distributions is comparable to the difference in mean halo mass as function of $n_{\text{sat}}$; this is qualitatively in line with the large scatter seen in $N_{\text{LRG}}$ vs. $M_{200}$ in Ho et al. (2007). Attempts to measure the mass of dark matter halos using LRG groups as a tracer should expect a broad distribution.
Figure 2.2: Contours for $\Delta \ln L = \{-1.15, -3.09, -5.9\}$ in the $M_1$ vs. $M_{\text{cut}}$ plane after marginalizing over $\alpha$. For a $\chi^2$ distribution, these contours would enclose 1, 2, and $3\sigma$ confidence regions. The cross indicates the maximum likelihood parameter values.
Figure 2.3: Same as Figure 2.2, but for $\alpha$ vs. $M_{\text{cut}}$. $M_{\text{cut}}/(10^{14} M_{\odot}) + \alpha$ is tightly constrained, while $M_{\text{cut}}/(10^{14} M_{\odot}) - \alpha$ is only weakly constrained.
Figure 2.4: Same as Figure 2.2, but for $\alpha$ vs. $M_1$. 

$M_1 / (10^{14} \text{ M}_{\odot})$
Figure 2.5: The dashed curve ranging from 0 to 1 shows the $N_{\text{cen}}(M)$ term for the maximum likelihood HOD; it should vary only slightly with HOD since the satellite fraction is well constrained by our model, and we hold $\sigma_{\log M}$ fixed. The solid curve shows $\langle N_{\text{sat}}(M)N_{\text{cen}}(M) \rangle$, with the error bars computed by Eqn. 2.20.
Figure 2.6: The probability $P(n_{\text{sat}} = 1|M)$ (long-short dashed curve), $P(n_{\text{sat}} = 2|M)$ (long dashed curve), $P(n_{\text{sat}} = 3|M)$ (short dashed curve), $P(n_{\text{sat}} = 4|M)$ (solid curve) vs. halo mass $M/M_{\text{sun}}$. Error bars are computed by Eqn. 2.20. Though $P(n_{\text{sat}}|M)$ are evaluated at identical values of $M$, points are slightly staggered for clarity.
Figure 2.7: The total number density of satellite galaxies occupying a halo of mass $M$ per $\log_{10} M$ (thick solid upper curve). The long-short dashed curve shows the number density of satellites occupying halos with $n_{\text{sat}} = 1$, long dashed curve for satellites in halos with $n_{\text{sat}} = 2$, short dashed curve for $n_{\text{sat}} = 3$, and solid curve for $n_{\text{sat}} = 4$. Error bars are computed by Eqn. 2.20.
2.3.4 Projected Correlation Function \( w_p(r_p) \)

Zehavi et al. (2005) and Masjedi et al. (2006) measure the projected correlation function \( w_p(r_p) \) for this sample on small and intermediate scales:

\[
 w_p(r_p) = 2 \int_0^{\pi_{max}} d\pi \xi(r_p, \pi). \tag{2.22}
\]

We follow Zehavi et al. (2005) and set \( \pi_{max} = 80 \) Mpc/\( h \), which is large enough to include most correlated pairs and produce stable estimates of \( w_p(r_p) \). Masjedi et al. (2006) recover missing fiber collision pairs by computing \( w_p(r_p) \) by cross correlation between the SDSS spectroscopic and imaging samples. They also correct for photometric biases of close galaxy pairs, which can introduce incompleteness of pairs with separation \( r_p \lesssim 0.1 \) Mpc/\( h \).

We present the projected correlation function \( w_p(r_p) \) averaged over 20 mock catalogs produced with our SO halo catalog using our maximum likelihood HOD in Figure 2.8. We find excellent agreement with the measurements of Masjedi et al. (2006). Using the diagonal error bars reported in Masjedi et al. (2006), we find \( \chi^2 = 7.5 \) for the outer 15 points. There is substantial discrepancy with the inner 3 points at \( r_p = 0.01, 0.016, 0.026 \) (not shown in Fig. 2.8); \( \chi^2 = 29 \) for all 18 points. The discrepancy is not surprising since these small distances are comparable to the force resolution of our simulation. Though the CiC method relies primarily on pairs with \( r_p \leq 0.8 \) Mpc/\( h \), our mock catalogs reproduce the features of the observed \( w_p(r_p) \) by adjusting a single parameter \( \sigma_{\log M} \) to match the large scale (\( \sim 20 \) Mpc/\( h \)) bias probed by \( w_p(r_p) \). Note that a sharp transition from 0 to 1 for \( N_{cen}(M) \) can be ruled out with confidence. Figure 2.8 shows \( w_p(r_p) \) for catalogs with \( \sigma_{\log M} = 0.2, 0.7, \) and 1.3 for comparison. All three catalogs match the observed clustering at \( r_p \lesssim 0.8 \) Mpc/\( h \) where we have CiC constraints, but only catalogs with \( \sigma_{\log M} \sim 0.7 \) match the observed clustering on \( \sim 2 - 20 \) Mpc/\( h \) scales, the regime where two-halo pairs dominate.
Figure 2.8: The projected correlation function $w_p(r_p)$ vs. projected separation $r_p$, both in Mpc/$h$. The points with error bars taken from Masjedi et al. (2006). The solid curve is an average over 20 mock catalogs at the maximum likelihood HOD parameters presented in the text for $\sigma_{\log M} = 0.7$. The Masjedi et al. (2006) data was used in our analysis only to optimize the value of $\sigma_{\log M}$, which sets the large scale clustering. The striking agreement shows that using SO halo catalogs we can simultaneously reproduce the CiC group multiplicity function and the clustering on scales $\sim 1 - 20$ Mpc/$h$. For comparison, we produce a catalog with $\sigma_{\log M} = 0.2$ (long dashed curve) and $\sigma_{\log M} = 1.3$ (short dashed curve). Both catalogs match the observed clustering at $r_p \lesssim 0.8$ Mpc/$h$ where we have CiC constraints, but disagree with the observations on larger scales.
2.3.5 Comparison with Results using FoF Catalogs and Previous Works

We also generate mock LRG catalogs based on a FoF halo catalog produced with linking length $b = 0.2$. For this catalog we use virial masses as defined in Bryan & Norman (1998). Setting $\sigma_{\text{log} M} = 0.5$, we are able to match both the SDSS multiplicity function (Column 4 of Table 2.3) and the large scale projected correlation. However, Figure 2.9 shows that the FoF catalog does not reproduce the observed $w_p(r_p)$ at $r_p \sim 1$ Mpc/$h$. Using the diagonal error bars only presented in Masjedi et al. (2006), this catalog has $\chi^2 = 104$ for the last 15 points (the mock catalog based on SO halos has $\chi^2 = 7.5$). The large discrepancy between the FoF and SO-based mock catalogs arises from the FoF algorithm’s tendency to link nearby halos. The SO algorithm explicated in Tinker et al. (2008) allows halo overlap in order to correctly separate distinct density peaks likely to host LRGs. This problem should also be circumvented by algorithms which identify subhalos.

If one sums the difference between the observed and FoF catalog $w_p(r_p)$ for the three points near 1 Mpc/$h$, the number of missing pairs in that regime is equal to half the number in the one-halo pair regime, $r_p \lesssim 0.8$ Mpc/$h$. To match only $w_p(r_p)$ using an FoF catalog, one would therefore need to enhance the number of pairs at separation $\sim 1$ Mpc/$h$ without greatly altering the number of pairs at smaller radii. In the halo model, this can be accomplished by putting many satellite galaxies in a large halos whose virial radii extend to $\sim 1$ Mpc/$h$, so that satellite-satellite pairs also contribute substantially to $w_p(r_p)$. In a given halo, satellite-satellite pairs will have larger separations than central-satellite pairs. This drives $\alpha$ to large values so that the one-halo term can accommodate the missing two-halo pairs in FoF catalogs.

Column 5 of Table 2.3 shows the expected true multiplicity function from the best fit HOD presented in Kulkarni et al. (2007), scaled to an SDSS volume. Their HOD predicts many more groups with $n_{\text{sat}} = 1, 2,$ and 3 than are observed; at larger $n_{\text{sat}}$ our current CiC method may not recover such large groups. This HOD also predicts 15 objects in an
Table 2.3: Column 1 is the number of satellites in the group. Column 2 is a copy of Column 3 of Table 2.1, our best estimate of the SDSS CiC multiplicity function. Column 3 shows the CiC multiplicity function averaged over 20 mock catalogs produced with our maximum likelihood HOD using the SO halo catalogs and scaled by the volume ratio $V_{SDSS}/V_{sim}$. Column 4 shows the CiC multiplicity function for mocks made with the FoF catalog. Column 5 shows our estimate of the expected $N_{true}$ in $V_{SDSS}$ for the Kulkarni et al. (2007) HOD.

<table>
<thead>
<tr>
<th>$n_{sat}$</th>
<th>$N_{SDSSCIC,final}$</th>
<th>$N_{SOMLHODCIC}$</th>
<th>$N_{FOF}$</th>
<th>$N_{Kulkarni}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>40407.56</td>
<td>40546.1</td>
<td>40542.0</td>
<td>37855</td>
</tr>
<tr>
<td>1</td>
<td>2301.12</td>
<td>2190.4</td>
<td>2265.1</td>
<td>5202</td>
</tr>
<tr>
<td>2</td>
<td>285.86</td>
<td>323.2</td>
<td>301.3</td>
<td>998</td>
</tr>
<tr>
<td>3</td>
<td>54.29</td>
<td>65.8</td>
<td>53.7</td>
<td>343</td>
</tr>
<tr>
<td>4</td>
<td>20.20</td>
<td>15.2</td>
<td>13.4</td>
<td>155</td>
</tr>
<tr>
<td>5</td>
<td>4.13</td>
<td>4.7</td>
<td>4.0</td>
<td>82</td>
</tr>
<tr>
<td>6</td>
<td>1.05</td>
<td>1.5</td>
<td>1.6</td>
<td>48</td>
</tr>
<tr>
<td>7</td>
<td>1.02</td>
<td>0.2</td>
<td>0.92</td>
<td>30</td>
</tr>
<tr>
<td>8</td>
<td>0.02</td>
<td>0.16</td>
<td>0.28</td>
<td>21</td>
</tr>
</tbody>
</table>

SDSS volume with $n_{sat} > 15$. Interestingly, if we fix $M_{cut} = 0$, we also find $\alpha \sim 1.4$ (see Figure 2.3). The cause of the discrepancy here seems to be mainly in $M_1$, the overall normalization of their $N_{sat}(M)$. While we find $f_{sat} = 0.064$, they report $f_{sat} = 0.17$. Perhaps their large FoF linking length $b = 0.6 \text{ Mpc}/h$ severely reduces the two-halo clustering in small scales, forcing the one-halo term to compensate.

The results of Ho et al. (2007) support our findings for a relatively low cutoff for the number of LRGs per halo. Figure 12 of Ho et al. (2007) shows a single cluster with 16 LRGs; and all other clusters in their sample have $\leq 11$ LRGs. We expect this to provide a strict upper bound on our sample, since their photometric sample has a number density $\sim 4$ times larger than the spectroscopic sample we study here.
Figure 2.9: The projected correlation function $w_p(r_p)$ vs. projected separation $r_p$, both in Mpc/$h$. The points with error bars are taken from Masjedi et al. (2006). The solid curve is an average over 20 mock catalogs at the maximum likelihood HOD parameters presented in the text for $\sigma_{\log M} = 0.7$ and using the SO halo catalog. The short dashed curve shows $w_p(r_p)$ averaged over 5 mock catalogs generated by populating an FoF halo catalog. The FoF halo catalog has a severe suppression of two-halo pairs at separations of $\sim 1$ Mpc/$h$, equal to the half of the total number of pairs in the one-halo regime, $r_p \lesssim 0.8$ Mpc/$h$. 
2.4 Assessment of the CiC method

The HOD parameter constraints reported in § 2.3.3 are for $\sigma_{\log M}$ held fixed. Since the CiC constraints always involve the product $\langle N_{\text{cen}}(M)N_{\text{sat}}(M) \rangle$ and Figure 2.5 shows that $N_{\text{cen}}(M) < 1$ in regions where the satellite contribution is non-negligible, the CiC maximum likelihood HOD parameters will vary with $\sigma_{\log M}$ to hold $\langle N_{\text{cen}}(M)N_{\text{sat}}(M) \rangle$ relatively fixed. Moreover, varying $\sigma_{\log M}$ changes the average bias of the LRG-occupied halos, which may in turn affect the distribution of $\Delta g(n_{\text{sat}})$ in Eqn. 2.14. We expect this to be a smaller effect. A complete HOD analysis would include $N_{\text{cen}}(M)$ parameters when evaluating the likelihood. We find that using $w_p(r_p)$ on $\sim 20$ Mpc/$h$ to set $\sigma_{\log M}$ is sufficient for the purposes of generating mock catalogs. We hope to explore variations in the functional form of $N_{\text{cen}}(M)$ in later work. We have not explored the dependence of our method on the assumed distribution of satellite galaxies. Seo et al. (2007) find that passively evolving galaxies are more concentrated than the dark matter at low redshift; this should make the CiC group finder more accurate than we found for our mock catalogs. For the parameters chosen here, the CiC method does not extract exact matches to one-halo groups with high fidelity. We have circumvented this issue by careful calibration on our simulated catalogs, and we expect to refine the method in the future to improve the accuracy of finding groups with more than two LRGs.

Because the CiC method can simultaneously reproduce the large scale clustering of LRGs and match the higher order statistics in the density field probed by the CiC group multiplicity function, we expect it to be a useful tool both for HOD constraints and the production of good mock catalogs. Since the CiC method only uses information on scales where the one-halo term dominates, our method is less sensitive to knowing the two-halo correlation function on $\sim$ Mpc/$h$ scales. We have demonstrated that the deficit of two-halo pairs at $\sim 1$ Mpc/$h$ is problematic for HOD fitting methods using 2 and 3-pt statistics and FoF catalogs because the HOD must accommodate the missing two halo pairs at this separation, and thus predict groups containing many more LRGs than are observed.
Finally, the low number density of the LRGs makes them particularly well-suited for this method.

### 2.5 Conclusions

The low number density of SDSS LRGs allows us to partially separate LRG pairs occupying the same dark matter halo from pairs occupying distinct dark matter halos. Candidate one-halo pairs are identified using simple cuts in the transverse and LOS separations. We group these pairs using the FoF algorithm to compute the CiC group multiplicity function. We measure the CiC group multiplicity function for the subsample of SDSS LRGs satisfying $-23.2 < M_g < -21.2$ and $0.16 < z < 0.36$, carefully accounting for the effects of fiber collisions and survey boundaries, holes, and incompleteness.

In order to derive HOD constraints from our measurement, we calibrated the relation between the CiC and true one-halo group multiplicity functions using mock LRG catalogs. The variance about the mean relation is comparable to the Poisson sampling variance of the CiC multiplicity function and must be properly accounted for in the maximum likelihood parameter estimation.

The CiC group multiplicity function places strong constraints on the satellite LRG HOD, $N_{\text{sat}}(M)$. When we fix $\sigma_{\log M} = 0.7$ and $\bar{n}_{\text{LRG}} = 10^{-4} \, (\text{Mpc}/h)^{-3}$, the maximum likelihood HOD parameters and marginalized one-dimensional 68% and 95% confidence intervals are $M_{\text{cut}} = 5.0_{-1.3}^{+1.5}(-2.6) \times 10^{13} M_{\odot}$, $M_1 = 4.95_{-0.20}^{+0.37}(-0.79) \times 10^{14} M_{\odot}$, and $\alpha = 1.035_{-0.17}^{+0.10}(-0.24)$, with $M_{\text{min}} = 8.05 \times 10^{13} M_{\odot}$ at the maximum likelihood point. We tightly constrain the satellite fraction to $f_{\text{sat}} = 0.0636_{-0.0020}^{+0.0019}(-0.0038)$. The projected correlation function $w_p(r_p)$ of mock catalogs derived from an SO halo catalog is an excellent agreement with the measurements of Masjedi et al. (2006) and Zehavi et al. (2005) when the large scale clustering
is used to fix $\sigma_{\log M}$. Fig. 2.9 shows that FoF halo catalogs have a severe deficit of pairs at $\sim 1 \text{ Mpc}/h$. In § 2.3.5 we point out that methods using $w_p(r_p)$ with an analytic estimate of the two-halo term calibrated using FoF halos will severely overestimate the number of satellites and the maximum expected one-halo group size. Our measured CiC group multiplicity function rules out the best fit HOD from Kulkarni et al. (2007).

Despite the increased complexity of our approach and necessary calibration using simulations, we have produced high quality mock catalogs which reproduce both higher order statistics in the density field and the features of the projected correlation function. These mock catalogs will be used in Ch. 4 to study the large scale structure statistics of our CiC groups.

### 2.6 Acknowledgments

Daniel Eisenstein provided our final LRG sample along with excellent advice regarding the SDSS analysis and Michael Blanton provided the inverse random catalogs and Tycho2 catalog. We thank Jeremy Tinker for excellent discussions regarding FoF vs. SO halo finding algorithms as well as kindly shared his SO halo-finding code, which was used to produce the halo catalogs in this paper. Rachel Mandelbaum and Charlie Conroy provided excellent guidance and suggestions in the initial stages of this project. Paul Bode kindly provided the simulation results and assistance.

Funding for the Sloan Digital Sky Survey (SDSS) has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the U.S. Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS Web site is http://www.sdss.org/. The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The Participating Institutions are The University of Chicago, Fermilab,
the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, University of Pittsburgh, Princeton University, the United States Naval Observatory, and the University of Washington.

BAR gratefully acknowledges support from the National Science Foundation Graduate Research Fellowship. This project was supported by NSF Grant 0707731. DNS thanks the APC (Universite de Paris VII) for its hospitality while this work was completed.
Chapter 3

Optimizing Dark Matter Simulations for LRG Mock Catalogs: Resolution Study

This chapter is based on work done in collaboration with David Spergel and Paul Bode.

3.1 Rationale

The underlying goal of this thesis is to increase $k_{\text{max}}$, the maximum wavenumber in $P_{\text{LRG}}(k)$ which can be robustly used to estimate the underlying linear power spectrum. To do so we must understand in detail the onset of nonlinearities, redshift space effects, and possible scale dependent biasing of the galaxies with respect to the dark matter. The goal of this chapter is to select the parameters for a set of simulations that will allow us to create the largest total volume of mock catalogs and have enough realizations to estimate the covariance matrix of the power spectrum, all without introducing systematic errors due to insufficient mass resolution.
How much mass resolution do we need? This depends on what particular statistics of the nonlinear density field we need to capture. The goal of this work is to measure the clustering properties of halos and determine the relationship between the LRG, dark matter, and CiC group power spectra. To know the precise relation between the CiC groups defined in Ch. 2 and the underlying halo density field, our simulations would need accurate halo-halo $n$-point correlation functions in redshift space for $n \leq 9$, the maximum CiC group size measured in Ch. 2 from the SDSS LRG sample. However, as long as the dominant source of discrepancy between the CiC group density field and the halo density field is either the failure of the CiC method to group together galaxies from the same halo or contamination of CiC groups due to a single close halo along the line of sight, then accuracy in two-point halo statistics will be sufficient for our scientific goals. The results of Ch. 2 that the satellite fraction of LRGs is low ($\sim 6\%$) and that 60% of one-halo pairs are found in groups with $n_{\text{sat}} = 1$ support this assumption. Therefore, as long as the two-point statistics in redshift space for the halos of interest are independent of resolution, we will consider the simulation mass resolution sufficient for our purposes. The detailed assumptions of how to populate halos with central and satellite galaxies may produce the largest source of uncertainty in the relation between the CiC and underlying matter density field. We explore this issue in later work.

Based on the results of Ch. 2 we made an estimate of the required simulation parameters, and set these to the parameters of our medium resolution simulation. In the sections below we compare this simulation with a lower and higher resolution simulation seeded with the same initial modes. We examine the halo mass function, 2-point clustering statistics in real and Fourier space, 2-point velocity statistics, and the CiC bias function $b(n_{\text{sat}}) = N_{\text{CiC}}(n_{\text{sat}})/N_{\text{true}}(n_{\text{sat}})$ for mock catalogs. Together these statistics allow us to put constraints on the level of error on the power spectrum of the CiC groups induced by the simulation mass resolution.
3.2 Simulations

Most simulations use a periodic box with exactly the mean universal matter density. While this approach is simple, the lack of large scale power leads to errors in the estimates of both the two-point and four-point functions. Sirko (2005) emphasized the importance of the DC mode in real space clustering. Hamilton et al. (2006) showed that the covariance on small scales is dominated by beat-coupling with larger scales; this beat-coupling term will not be present when measuring the power spectrum at discrete wavenumbers as in a periodic box. We capture this effect by allowing the DC mode in the box to be nonzero.

Initial conditions for the simulations were generated with the publicly available ic code (Sirko, 2005). We updated the code to use the mt19937 “Mersenne Twister” generator (Matsumoto & Nishimura, 1998) to select the initial density mode realizations. ic was designed to generate initial conditions for an ensemble of periodic box simulations that would match real-space statistical properties such as the mass variance in spheres and $\xi(r)$. For a periodic box simulation with comoving side length $L_{uni}$ at $z = \infty$, the initial modes are drawn from a convolved power spectrum $P_{L_{uni}}(k)$ for which $P_{L_{uni}}(0)/L_{uni}^3$ is the variance of the DC mode in a volume equal to the initial periodic box volume. The DC mode is assumed to evolve linearly, and so its effects can be mimicked to first order in Lagrangian perturbation theory by a slight change in the cosmological parameters of the box as a function of the value of the DC mode in the realization (see Eqns. 19 - 22 of Sirko (2005)). The relation between the scale factor in the box cosmology, $a_{box}$, and the scale factor in the universe cosmology, $a_{uni}$ is

$$\frac{a_{box}}{a_{uni}} = 1 - \frac{1}{3} \frac{D_{uni}(a_{uni})}{D_{uni}(1)} \Delta_o$$

(3.1)

where $\Delta_o$ is the amplitude of the DC mode at $a_{uni} = 1$, and the growth function $D_{uni}$ is evaluated using the universe cosmology. In this scenario, each simulation in the ensemble represents an equal initial comoving volume, $L_{uni}^3$. Once the universe has evolved to scale factor $a_{uni}$, the comoving size of the box is $L_{box} = L_{uni}(a_{box}/a_{uni})$. Because overdense regions expand more slowly than underdense regions, we must weight the simulations by
We use the publicly-available Tree-Particle-Mesh (TPM) code (Bode & Ostriker, 2003) to run the periodic box $N$-body simulations. We fix the cosmological parameters at the values recommended from the latest WMAP5 analysis: $(\Omega_m, \Omega_b, \Omega_{\Lambda}, n_s, \sigma_8, h) = (0.2792, 0.0462, 0.7208, 0.960, 0.817, 0.701)$ (Komatsu et al., 2008). We set $L_{\text{box}} = 558 \, \text{Mpc}/h$. We ran 3 simulations of varying mass resolution: $N_p = (128 \times 3)^3$, $(128 \times 4)^3$, and $(128 \times 5)^3$, resulting in particle masses $M_p = 3.39 \times 10^{11}$, $1.43 \times 10^{11}$, and $7.33 \times 10^{10} \, M_\odot$. Each simulation had the same realizations of $\delta(\vec{k})$, apart from the cut-off at $k_{\text{Nyquist}}$ of the initial particle grid which has $N_{\text{grid},1d} = N_p^{1/3}$. In this particular realization, $\Delta_o = -0.01175$, so these simulations represent a slightly underdense region of the universe. For absolute comparisons to our simulation results, we must account for the nonzero DC mode. To estimate the expected number density of halos in a specified mass range, we assume a linear bias $b_h$ of the halos with respect to the dark matter and use Eqn. 66 of Cooray & Sheth (2002):

$$\delta_h = b_h \delta_{DM}. \quad (3.2)$$

Then the universal average number density of these halos can be estimated by

$$n_{\text{uni},i}(a_{\text{uni}}) = \frac{n_{\text{sim},i}(a_{\text{box}})}{1 + b_h \Delta_o D(a_{\text{uni}})/D(1)} \quad (3.3)$$

$$b_h = \frac{P_{hm}(k)}{P_{nm}(k)}, k \ll k_{nl} \quad (3.4)$$

where we estimate the halo bias $b_h$ from the ratio of the halo-matter cross spectrum to the matter power spectrum in the limit of small $k$. Since we know the number of halos in our simulation box, the best estimate for the Poisson contribution of the power spectrum is

$$P_{\text{poiss},hh}(k) = n_{\text{sim},h}(a_{\text{box}})/n_{\text{uni},h}^2(a_{\text{uni}}). \quad (3.5)$$

Because $P_{\text{poiss},hh} \gg P_{hh}(k)$ at large $k$ for the LRG number density, the difference between Eqn. 3.5 and $1/n_{\text{uni},h}(a_{\text{uni}})$ is noticeable.
In Ch. 4 we produce catalogs at the mean redshift of the NEAR, MID, and FAR samples in Tegmark et al. (2006). The detailed clustering and velocity statistics comparisons in this chapter will be made on the simulation outputs for the MID sample, though we generate catalogs for the NEAR and FAR sample as well to demonstrate that the halo mass resolution is sufficient for all three samples. For the MID sample output the low and medium resolution simulations were output at $z_{uni} = 0.344$ while the high resolution simulation was output slightly later at $z_{uni} = 0.339$ due to different TPM time step sizes in the simulations.

We use the spherical overdensity code described in Tinker et al. (2008) to find all halos defined by $\rho = 200\rho_m$ containing at least 50 particles. This is below the limit typically used for careful measurements of the halo mass function, but a large improvement over the 10 particle FoF halos used in the previous SDSS LRG mock catalogs, described in Seo & Eisenstein (2005). In this Chapter, we test for selection effects on halos with $\leq 100$ particles.

### 3.3 Mass Function

In this section, we explore how the mass resolution of the simulation affects the halo mass function. Tinker et al. (2008) determine a fitting function to the mass function of halos found with a spherical overdensity (SO) halo finder in a large set of $N$-body simulations. Their fit is a function of both redshift and the spherical overdensity defining the halos. Figure 3.1 shows the expected number of halos in our simulation box in logarithmic mass bins with $\Delta \log_{10} M = 0.2$ using their spline fitting function. If halos are Poisson tracers of the matter field, then Figure 3.1 is also an estimate of the expected variance in each mass bin. Figure 3.2 compares our simulation halo counts in the same mass bins with Tinker et al. (2008). We have used the initial comoving size of our box, $L_{box} = 558$ Mpc/$h$, to estimate the expected counts from Tinker et al. (2008) rather than the box size at the observation redshift, $L_{box} = 559.84$ Mpc/$h$. While this partially accounts for the underdensity of our simulation in the comparison, one expects a greater suppression in the number density of
halos in the mass range here, where \( b(M) > 1 \) (Eqn. 3.3). There is good agreement with the Tinker et al. (2008) mass function at the level of accuracy claimed for their analytic fits (\( \sim 5\% \)), especially considering that their fits use many simulations at \( \sigma_8 = 0.9 \).

The ratio of the counts in the low (blue) and medium (pink) resolution simulation to the high resolution simulation counts in logarithmic mass bins (\( \Delta \log_{10} M = 0.2 \)) are shown in Figure 3.3. The low mass boundary is set to \( M = 50 M_{\text{p,med}} = 7.15 \times 10^{12} M_\odot \). The first few mass bins are 50 - 79; 80 - 125; and 126 - 199 particles in the medium resolution simulation (53 - 83; 84 - 133; and 134 - 210 particles for the first 3 non-zero points for the low resolution simulation). At moderate and high masses, the low and medium resolution simulations have \( \sim 1 - 2\% \) too few halos. The observed discrepancy can be attributed to the discrepancy in observation redshift of the simulations: \( \ln(t_{\text{high}}/t_{\text{low,med}}) = 0.008 \). We can approximately correct for the bias by scaling all masses by the factor 1.015 (see Figure 3.4). That is, \( d \ln M/d \ln t \approx 1.8 \). This is in line with the value of 1.75 for a \( 10^{14} h^{-1} M_\odot \) halo found in Voit et al. (2003) when considering halo merger trees. The 1.5% correction to all halo masses is more than enough to explain the discrepancy in Figure 3.3.

We also compare the mass functions of the low and medium resolution simulation, which
Figure 3.2: The ratio of the counts in the low (blue), medium (pink), and high (green) resolution simulation to the analytic fit presented in Tinker et al. (2008) in logarithmic mass bins ($\Delta \log_{10} M = 0.2$); we have used the initial co-moving size of our box, $L_{\text{box}} = 558$ Mpc/$h$, to estimate the expected counts. This partially accounts for the underdensity of our realization. The low and medium resolution simulation results are output $z = 0.344$, while the output for the high resolution simulation is from $z = 0.339$.

are luckily at the same redshift; see Figure 3.5. The first few mass bins are 50 - 79; 80 - 125; and 126 - 199 particles in the low resolution simulation (119 - 187; 188 - 297; and 298 - 471 in the medium resolution simulation). The agreement is good to $\sim 1\%$ in the intermediate regime where there are many halos per bin.

Notice that in Figures 3.3 and 3.5, as one reaches the lowest mass bin (50-79 particles), there is a $\sim 5-6\%$ increase in halo counts in the lower resolution simulation. This is probably due to the increased noise in the SO halo mass estimates coupled with the scattering of halos into the lowest mass bin due to the steepness of the mass function. We must check whether the slight increase in halo counts in this bin modifies any two-point statistics relevant to making mock galaxy catalogs. We argue that any smooth and small correction to halo masses will be degenerate with small variations in the HOD; therefore, this small systematic is not important for our purposes. However, a selection effect on pairs of halos rather than a correction to the halo mass will distort the statistics important to our measurements.
Figure 3.3: The ratio of the counts in the low (blue) and medium (pink) resolution simulation to the high resolution simulation in logarithmic mass bins ($\Delta \log_{10} M = 0.2$). The low and medium resolution simulation results are for $z = 0.344$, while the output for the high resolution simulation is from $z = 0.339$.

Figure 3.4: The ratio of the counts in the low (blue) and medium (pink) resolution simulation to the high resolution simulation in logarithmic mass bins ($\Delta \log_{10} M = 0.2$), after scaling the low and medium resolution mass values by the factor 1.015. This factor more than accounts for the evolution in the mass function between $z = 0.344$ and $z = 0.339$. 
Comparing two-point statistics

Our aim in this section is to show that systematic bias in the two-point statistics in the dark matter and halo density field are small. The TPM algorithm used to run our simulations increases the resolution in regions of high density contrast. There are several scenarios under which the spatial dependence of the resolution could induce a selection effect on halos. A small subhalo or halo merging into a large object will get full resolution because its particles are considered part of the larger object, while a similar object farther away from the large object or in a filament may not be selected for high resolution because it is in a dense region, and so its density contrast with its surroundings is lower. Meanwhile, the same object in a void will get full resolution since it has high density contrast with its surroundings. These potential selection effects would show up in the comparison of cross-correlation between low and high mass halo bins as a function of simulation resolution.

Tinker et al. (2008) conservatively set a minimum of 400 particles per halo in their mass function measurements; below this value they claim that simulations with different mass resolutions begin to diverge. In § 3.3 we showed that the number density of halos increases above the expected value in the lowest mass bin with 50-80 particles; for our purposes we can neglect this effect since these halos will not host satellite galax-
ies, and slight changes in the mass function are degenerate with the parametrization of $N_{cen}(M)$. However, we want to make sure the two-point statistics of these halos are robust. We split the medium and high resolution simulation halos into two mass bins. Halos in the mass range $50M_{p,med} \leq M < 100M_{p,med} = 7.15 \times 10^{12}M_\odot \leq M < 1.43 \times 10^{13}M_\odot$ ($98M_{p,high} \leq M < 196M_{p,high} = 7.18 \times 10^{12}M_\odot \leq M < 1.44 \times 10^{13}M_\odot$) are in the low mass bin, and everything above that mass is in the high mass bin. For the medium resolution simulation, the number of halos in the low mass bin is 74466, and 64285 in the high mass bin; in the high resolution simulation, 72529 halos are in the low mass bin and 64445 are in the high mass bin. Using Eqn. 3.4 to estimate the large scale bias as $k \to 0$, the low mass halo bin has bias $b_{low} = 1.34$ and the high mass halo bin has bias $b_{high} = 1.78$.

In the following sections we check the clustering statistics on both large and small scales. We measure the dark matter, low and high mass halo bin real space power spectra along with their cross spectra, and the real space auto and cross halo correlation functions. Discrepancies between the simulations will indicate the possibility of halo selection effects.

To assure our simulations are accurate for two-point statistics in redshift space, we compare the rms halo velocities, and the velocity correlations with two statistics: 
\[
\frac{\langle \vec{v}_i(\vec{r}) \cdot \vec{v}_j(\vec{r}) \rangle}{\sqrt{\langle \vec{v}_i^2 \rangle\langle \vec{v}_j^2 \rangle}} \text{ as a function of } |\vec{r} - \vec{r}'| \text{ and } \langle (\vec{v}_i(\vec{r}) - \vec{v}_j(\vec{r}')) \cdot (\vec{r} - \vec{r}') / |\vec{r} - \vec{r}'| \rangle \text{ vs } |\vec{r} - \vec{r}'|.
\]
The first statistic encompasses the coherence of the velocity field, while the second is sensitive to the FOG effect.
3.4.1 Quantifying the level of systematic bias due to mass resolution

Halos as Poisson tracers of the matter density field

In the simplest approximation, halos can be modelled as linearly biased Poisson tracers of the matter density field:

\[
\delta_h(\vec{x}) = b_h \delta_{DM}(\vec{x}) + n_{shot}(\vec{x})
\]

\[
\langle n_{shot}(\vec{x})n_{shot}(\vec{x}') \rangle = \frac{1}{n_h} \delta^D(\vec{x} - \vec{x}').
\]

Smith et al. (2007) argue that both the assumption of linear bias and Poisson sampling break down on moderate scales. Since halos may not overlap, the effective shot noise contribution of halos cannot be white. They find evidence for scale-dependent halo biasing by examining the halo-matter cross power spectrum. However, for the purposes of estimating errors and characterizing the large-scale clustering, we will ignore these effects and quantify the clustering properties as a function of halo mass in terms of a linear bias \(b_h\) (denoted as \(b_1\) and \(b_2\) to denote the linear bias for halo mass bins 1 and 2).

If we take the underlying dark matter density field \(\delta_{DM}(\vec{x})\) as fixed (i.e., not a random variable over which we take expectations) and assume that the shot noises associated with different non-overlapping halo mass bins \(h1\) and \(h2\) are uncorrelated, then we expect

\[
\langle \delta_{h}(\vec{k})\delta_{DM}^*(\vec{k}) \rangle = b_h |\delta_{DM}(\vec{k})|^2
\]

\[
\langle \delta_{h}(\vec{k})\delta_{DM}^*(\vec{k})\delta_{h}^*(\vec{k})\delta_{DM}(\vec{k}) \rangle - \langle \delta_{h}(\vec{k})\delta_{BM}^*(\vec{k}) \rangle^2 = |\delta_{DM}(\vec{k})|^2 \frac{1}{n_h}
\]

\[
\langle \delta_{h}(\vec{k})\delta_{BM}^*(\vec{k}) \rangle = b_h^2 |\delta_{DM}(\vec{k})|^2 + \frac{1}{n_h}
\]

\[
\langle \delta_{h1}(\vec{k})\delta_{h2}^*(\vec{k})\delta_{h1}^*(\vec{k})\delta_{h2}(\vec{k}) \rangle - \langle \delta_{h1}(\vec{k})\delta_{h2}^*(\vec{k}) \rangle^2 = 2b_h^2 |\delta_{DM}(\vec{k})|^2 \frac{1}{n_h} + \left( \frac{1}{n_h} \right)^2
\]

\[
\langle \delta_{h1}(\vec{k})\delta_{h2}^*(\vec{k}) \rangle = b_1 b_2 |\delta_{DM}(\vec{k})|^2
\]

\[
\langle \delta_{h1}(\vec{k})\delta_{h2}^*(\vec{k})\delta_{h1}^*(\vec{k})\delta_{h2}(\vec{k}) \rangle - \langle \delta_{h1}(\vec{k})\delta_{h2}^*(\vec{k}) \rangle^2 = b_1^2 |\delta_{DM}(\vec{k})|^2 \frac{1}{n_{h1}}
\]

\[
+ b_2^2 |\delta_{DM}(\vec{k})|^2 \frac{1}{n_{h2}} + \frac{1}{n_{h1}n_{h2}}.
\]

The correlation function between halo bins \(h1\) and \(h2\) in the simulation box is estimated
by
\[
\xi_{h_1,h_2}(r) = \frac{N_{h_1,h_2}(r)(1 + \delta_{h_1,h_2})L_{\text{box}}^3}{N_{h_1}N_{h_2}V(r)} - 1
\] (3.14)

Here \(\xi_{h_1,h_2}(r)\) is the cross correlation (auto if \(h_1 = h_2\)) between two species of objects, \(N_{h_1,h_2}(r)\) is the number of pairs with separation \(r\) to \(r + dr\), \(V(r)\) is the volume of the bin between \(r\) and \(r + dr\), and \(\bar{N}_{h_1,h_2}\) are the expected number of objects of type \(h_1/h_2\) (estimated using Eqn. 3.3). The model of halos as Poisson tracers implies

\[
\frac{\langle \Delta \xi(r) \Delta \xi(r) \rangle}{\xi^2(r)} = \frac{1}{\bar{N}(r)}
\] (3.15)

when neglecting the error on \(\bar{N}_{h_1,h_2}\), which is appropriate for our comparisons of different resolution simulations with the same initial conditions.

**Errors on the comparison between different resolution simulations**

The errors on measuring the simulation two-point statistics arise only from rounding or from the Fast Fourier Transforms we use to measure \(P(k)\), and are negligible. All discrepancy between dark matter statistics measured from different resolution simulations can therefore be attributed to resolution effects or to the missing power at large \(k\) in the lower resolution simulations. However, it is well known that slight perturbations in initial conditions for an \(N\)-body problem result in exponentially diverging solutions in phase space (Miller, 1964), so the precise location of dark matter particles will diverge between the simulations.

Even for identical dark matter density fields sampled with different numbers of particles, the Poisson sampling noise will introduce noise in the halo finding algorithm, particularly at low halo masses. We have already seen that the effect of this noise is to increase the number of halos above the expected value at low masses. Despite this noise, we expect there to be a strong correlation between the locations of halos in different resolution simulations, so that the noise in the difference between two-point halo statistics measured in different resolution simulations should be sub-Poissonian, but at an unknown level. We
therefore use Eqns. 3.8 through 3.15 to show errors in § 3.5. This is both an upper bound on the level of noise introduced by the halo finder, and a good point of comparison for the level of systematics. We consider systematic trends acceptable as long as they are not significant compared to the Poisson variance, since the Poisson variance quantifies the noise inherent in extracting the underlying matter power spectrum from the halo power spectrum.

3.5 Two Point Statistics Results

3.5.1 Comparison of Dark Matter Power Spectra

We first compare the dark matter power spectra from the low, medium and high resolution simulation. Since we are focused on large scale statistics, we only compute $P(k)$ for $k < 0.4$ Mpc/$h$. In contrast to the discussion of halos above, the Poisson noise associated with the dark matter particles is likely to be a good estimate of the error on the measurement of the ‘underlying’ continuous field, since the Poisson noise is unlikely to be correlated between resolutions after the simulations have evolved. Since the high resolution simulation was output at $z_{\text{high}} = 0.339$ while the low and medium resolution simulations were output at $z_{\text{med,low}} = 0.344$, we scale the high resolution $P(k)$ down by the expected linear growth, $(D(z_{\text{med}})/D(z_{\text{high}}))^2$, for the comparison. In Figure 3.6 we see a slight but systematic decrease in the power at high $k$ as the resolution decreases; this is a $\sim 0.2\%$ effect at $k = 0.2$ Mpc/$h$ between the medium and high resolution simulation. Since the comparison of the low and medium resolution simulation at identical output redshifts shows a similar effect, the discrepancy is not the result of nonlinear growth between $z_{\text{med}}$ and $z_{\text{high}}$. The validity of Poisson subtraction at low $k$ has been questioned (see, e.g., Sirko (2005) and references therein). While we have subtracted the standard Poissonian shot noise here, not subtracting it would restore only half of the discrepancy shown in Figure 3.6. Therefore, we conclude that there is a small systematic suppression of power at high $k$ as the simulation resolution decreases. Figure 11 in Cooray and Sheth indicates that the one-halo term is
Figure 3.6: The red curve shows the ratio of the power spectrum of the medium resolution simulation to the high resolution simulation, after scaling by the expected linear growth between the redshift of observation in the low and medium resolution simulations, $z = 0.344$, and the redshift of the high resolution simulation output, $z = 0.339$. Error bars are estimated as Poisson shot noise from the finite number of dark matter particles. The blue curve shows the ratio of the low resolution simulation to the medium resolution simulation, while the green curve is the ratio of the low resolution simulation to the scaled high resolution simulation (simply the product of the red and blue curves).

$\gtrsim 10\%$ of the total power at $k = 0.2$ and equal in magnitude to the two-halo term at $k = 0.4$. Therefore, substantial changes in the halo profiles with resolution may cause this discrepancy; if so, this discrepancy is of little concern for us since we focus here on how the halos trace the underlying matter field. Regardless, the effect is small compared with the expected Poisson sampling errors on $P_{\text{LRG}}(k)$.

### 3.5.2 Halo Power Spectra

In this section, we explore the effects of simulation resolution on halo power spectra by comparing spectra of the low and high halo mass bins specified in § 3.4 from our medium and high resolution simulations. The ratio of halo to dark matter spectra are shown in Figure 3.7 and halo-dark cross spectra are shown in Figure 3.8 for the medium resolution simulation; the high resolution simulation results are very similar. We fit this ratio as a linear function of wavenumber, $r_0 + r_1 k$. The values extrapolated to $k = 0$ provide an estimate of the large scale bias; the fit values derived from the different types of spectra are
all consistent with each other at the level of expected error ($\lesssim 1\%$). Halos may be more biased in the medium resolution simulation at the $\lesssim 0.5\%$ level in both the low and high mass bins; this is comparable with the level of error on the measured amplitudes of the spectra, and consistent with the small discrepancies observed in the mass functions.

We compute a $\chi^2$ using the errors in Eqn. 3.8 through 3.13 and fitting a linear relation $r_0 + r_1 k$ to the first 10 points (up to $k = 0.118$). There are $10 - 2 = 8$ degrees of freedom for each fit. We find $\chi^2 = 12.2, 5.5, 5.9, 16.3, 8.6$ for $\delta h_l - \delta h_l, \delta h_l - \delta h_h, \delta h_l - \delta_{dm}, \delta h_h - \delta h_h, \delta h_l - \delta_{DM}$. Subscript $h$ denotes the high mass halo bin, $l$ the low mass halo bin, and $DM$ the dark matter. These values are consistent with the expected values for a linearly biased, Poisson model for the halos expressed in Eqn. 3.8 through 3.13, though $r_1 = 0$ in a linear biasing model.

The high mass halo bin auto-power spectrum to dark matter spectrum ratio decreases with $k$, but increases with $k$ in the halo-dark cross spectrum. Smith et al. (2007) argue that the standard Poisson correction is incorrect for halos due to halo exclusion, which would explain the suppression of our halo spectra with respect to the dark matter. Another possibility is that the cross spectrum is more sensitive to nonlinear halo biasing, since $\delta_h(\vec{k})\delta_{DM}(\vec{k})$ is computed at each $\vec{k}$, so phase information is retained. This trend is also seen in the high resolution simulation with consistent amplitudes, so it is not a resolution effect. We examine this issue for the LRGs in Ch. 4.

Comparison of Medium and High Resolution Spectra

When we directly take the ratio of the spectra from the medium and high resolution simulation, we do not detect any significant deviation from a constant ratio (see Figure 3.9 and Figure 3.10). The best fit value of the ratio deviates from 1 only in that the halos in the medium resolution are $\sim 0.5\%$ more highly biased than those in the high resolution
Figure 3.7: The ratio of the halo to dark matter power spectra vs. $k$ in units $(\text{Mpc}/h)^{-1}$ for the medium resolution simulation. The blue curve is for the low mass bin auto-power spectrum, the upper gold curve is for the high mass bin auto-power spectrum, and the middle red curve is the cross power spectrum of the low and high mass bins.

Figure 3.8: The ratio of the halo-dark matter cross spectrum vs. $k$ in units $(\text{Mpc}/h)^{-1}$ to the dark matter power spectra for the medium resolution simulation. The blue curve is for the low mass bin, and the red curve is for the high mass bin.
Figure 3.9: The ratio of the halo spectra for the medium to high resolution simulation vs. $k$ in units ($\text{Mpc}/h)^{-1}$. The blue curve is for the low mass bin auto-power spectrum, the gold curve is for the high mass bin auto-power spectrum, and the middle red curve is the cross power spectrum of the low and high mass bins. Linear fits to the ratio $r_0 + r_1 k$ are shown as well. The Poisson errors of Eqns. 3.8 to 3.13 are shown only for the low mass halo sample for clarity.

simulation. For clarity we show the Poisson error bars for the low mass bin; as discussed in § 3.4, these errors overestimate the variance between the medium and high resolution simulations. For the low mass halo sample, there appears to be a small scale dependence in the ratio between the medium and high resolution spectra ratio. However, the trend is not clear for $k < 0.2 h/\text{Mpc}$ where we focus our $P(k)$ analysis in Ch. 4. We next search directly in the correlation function for evidence of halo selection effects on small scales, where they would be more evident.

3.5.3 Halo Correlation functions

We use the large scale halo bias to estimate the expected number density of halos for the low and high mass halo bins by Eqn. 3.3. Figure 3.11 shows the autocorrelation of the low and high mass halos in the high resolution simulation as well as their cross-spectrum. All three halo correlation functions asymptote to negative values, reflecting the underdensity of our simulation. In the linear approximation we would expect $\xi_{ij}(r) = b_i b_j \xi_{\text{linear}}$; Figure 3.11 shows similar qualitative behavior. The small scale behavior is shown in Figure 3.12. At the smallest separations ($r = 0 - 5 \text{ Mpc}/h$), no halo pairs are found due to halo exclusion.
Figure 3.10: The ratio of the halo-dark matter cross spectra vs. \( k \) in units \((Mpc/h)^{-1}\) in the medium and high resolution simulation. The blue curve is for the low mass bin, and the red curve is for the high mass bin. Linear fits to the ratio \( r_0 + r_1 k \) are shown as well.

Figure 3.11: The autocorrelation of the high mass halos (gold), low mass halos (blue), their cross spectrum (red), and the linear \( \xi \) averaged in the same bins of \( r \) (green). \( \xi(r) \) is computed from bins of size \( \Delta r = 5 \text{ Mpc}/h \).

The high mass bin autocorrelation peaks at larger \( r \) than the low mass bin autocorrelation since the exclusion of high mass halos extends to larger radii. The high mass halos are more highly correlated at small \( r \).

Figures 3.13 though 3.15 show the \( \xi_{high}(r) - \xi_{med}(r) \) for the high and low mass halo bins and \( r < 15 \text{ Mpc}/h \), with error bars indicating the Poisson noise (Eqn. 3.15). There is a small indication in Figure 3.13 that close pairs in the low mass bin are over-abundant in the medium resolution simulation. Inside 4 Mpc/h we find an excess of pairs of 0.8% in the medium resolution simulation with respect to the high resolution simulation. When
we make halo catalogs we will be able to examine the effect of this small bias; it should be sufficiently diluted since only a small fraction of LRGs will reside in halos in the low mass bin. In contrast, the cross spectra and high mass halo spectra in Figures 3.14 and 3.15 seem to be consistent: in both the number of pairs within 4 Mpc/h agree to within 0.1%.

We show the behavior of $\Delta \xi$ on scales $r = 15 - 150$ Mpc/h in Figures 3.16 - 3.18. All three $\Delta \xi$ measurements are consistent with 0 at the $\sim 0.002$ level throughout this $r$ range. For the low mass halo bin and the low-high cross correlation, $\Delta \xi$ goes from slightly negative to slightly positive. On comparison with Figure 3.11, such a trend would be consistent with the medium resolution simulation halos being slightly more biased, in agreement with
Figure 3.14: $\xi_{\text{high}}(r) - \xi_{\text{med}}(r)$ computed from bins of size $\Delta r = 0.5$ Mpc/$h$ for the halo mass bins cross correlation.

Figure 3.15: $\xi_{\text{high}}(r) - \xi_{\text{med}}(r)$ computed from bins of size $\Delta r = 0.5$ Mpc/$h$ for the high halo mass bin autocorrelation.
the large scale bias values deduced from the power spectra; moreover, evolution between $z_{med} = 0.334$ and $z_{high} = 0.339$ will slightly lower the bias with fixed mass boundaries. In summary, we find no statistically significant evidence for resolution-dependent effects in $\xi(r)$ between the medium and high resolution simulations.

3.6 Velocity Correlations

The rms velocity for the low mass halo bin is 520 km/s and 519 km/s for the high mass halo bin. Since the velocity field is well-correlated on large scales, $\langle \mathbf{v}_i(\mathbf{r}) \cdot \mathbf{v}_j(\mathbf{r}') \rangle / \sqrt{\langle v_i^2 \rangle \langle v_j^2 \rangle}$ drops slowly with $|\mathbf{r} - \mathbf{r}'|$ (see Fig. 3.19). The correlation is slightly lower for the high mass

Figure 3.16: $\xi_{high}(r) - \xi_{med}(r)$ computed from bins of size $\Delta r = 5$ Mpc/h for the low halo mass bin autocorrelation.

Figure 3.17: $\xi_{high}(r) - \xi_{med}(r)$ computed from bins of size $\Delta r = 5$ Mpc/h for the halo mass bins cross correlation.
Figure 3.18: $\xi_{\text{high}}(r) - \xi_{\text{med}}(r)$ computed from bins of size $\Delta r = 5$ Mpc$/h$ for the high halo mass bin autocorrelation.

Figure 3.19: The normalized halo-halo velocity correlation function, $\langle \vec{v}_i(\vec{r}) \cdot \vec{v}_j(\vec{r'}) \rangle / \sqrt{\langle v_i^2 \rangle \langle v_j^2 \rangle}$, vs $|\vec{r} - \vec{r'}|$. Gold is the high mass bin, blue is the low mass bin, and red is the for low mass - high mass pairs. Results are binned in intervals of 0.5 Mpc$/h$ in $|\vec{r} - \vec{r'}|$.

halos ($\sim 5\%$) at $r \lesssim 15$ Mpc$/h$. Figs. 3.20 and 3.21 compare this statistic for the medium and high resolution simulation. Figure 3.21 indicates a slight suppression of this statistic for the low mass bin at the $\sim 0.001$ level, or few percent level as one reaches $r \sim 150$ Mpc$/h$ where $\langle \vec{v}_i(\vec{r}) \cdot \vec{v}_j(\vec{r'}) \rangle / \sqrt{\langle v_i^2 \rangle \langle v_j^2 \rangle} = 0.03$.

A statistic perhaps more relevant to checking the accuracy of our simulations’ nonlinear redshift space distortions is the relative halo velocities along the vector separating the halos: $\langle (\vec{v}_i(\vec{r}) - \vec{v}_j(\vec{r'})) \cdot (\vec{r} - \vec{r'})/|\vec{r} - \vec{r'}| \rangle$, shown in Figure 3.22. The larger mass halos have larger infall velocities, reaching 400 km/s around 2 Mpc$/h$. Figure 3.23 shows that the medium
Figure 3.20: The difference at small separations in the halo-halo velocity correlation function vs. $|\vec{r} - \vec{r}'|$ in Mpc/$h$ between high and medium resolution simulations. Gold is the high mass bin, blue is the low mass bin, and red is the for low mass - high mass pairs. Results are binned in intervals of 0.5 Mpc/$h$ in $|\vec{r} - \vec{r}'|$.

Figure 3.21: Same as Figure 3.20, but results are binned in intervals of 5 Mpc/$h$ and extend to $|\vec{r} - \vec{r}'| = 150$ Mpc/$h$. 
Figure 3.22: The halo-halo relative velocity correlation function along the halo separation vector, \( \langle (\vec{v}_i(r) - \vec{v}_j(r')) \cdot (\vec{r} - \vec{r'}) / |\vec{r} - \vec{r'}| \rangle \) vs \( |\vec{r} - \vec{r'}| \). Gold is the high mass bin, blue is the low mass bin, and red is the for low mass - high mass pairs. Results are binned in intervals of 0.5 Mpc/h in \( |\vec{r} - \vec{r'}| \).

and high resolution simulations agree in this statistic to within \( \sim 1 \) km/s. The slight discrepancy, if significant, may be explained by a slight difference in the bias of the halo mass bins between the medium and high resolution simulations.

### 3.7 Mock Catalog Results

In this section we compare several statistics of mock catalogs generated using the algorithm in Ch. 2. To summarize, the CiC method is used to constrain the parameters of \( N_{sat} \) while \( \bar{n} \) and the large scale clustering amplitude are used to fix \( \sigma_{logM} \). We have adopted different cosmological parameters and different output redshifts than in Ch. 2, so we must repeat the process of finding the HOD parameters.

There are two main differences between the catalogs produced in this chapter and those in Ch. 2. Here we use a previous incarnation of our maximum likelihood evaluation that neglects the width of the distribution in \( \Delta g(n_{sat}) = N_{CiC}(n_{sat}) - N_{true}(n_{sat}) \). Instead, we assume a deterministic relation \( N_{CiC}(n_{sat}) = b(n_{sat}) \times N_{CiC}(n_{sat}) \). We refer to \( b(n_{sat}) \) as the multiplicity bias and examine it as a function of simulation resolution. These mock
Figure 3.23: The difference in the halo-halo relative velocity correlation function along the halo separation vector vs. |\vec{r} - \vec{r}'| in Mpc/h in the high and medium resolution simulations. Gold is the high mass bin, blue is the low mass bin, and red is the for low mass - high mass pairs. Results are binned in intervals of 5 Mpc/h in |\vec{r} - \vec{r}'|.

catalogs will be sufficiently close to the maximum likelihood HOD for the purpose here of comparing mock catalogs from different resolution simulations. Also note that we used the $N_{CiC}(n_{sat})$ multiplicity function measured in Ch. 2 for this analysis, while in Ch. 4 we recompute $N_{CiC}(n_{sat})$ for the NEAR, MID, and FAR redshift LRG subsamples.

Second, using the large scale bias to fix $\sigma_{\log M}$ is more challenging because we are working with an underdense simulation box, and the volume is small enough for the DC mode to be important. The discrepancy in the LRG number density in the box, $n_{box}$, and the universe average, $n_{uni}$, causes $w_p$ to drop at large $r_p$. We therefore compute $w_p(r_p)$ using both $n_{uni}$, appropriate when this is a small region of a much larger survey over which to estimate $n_{uni}$, and $\tilde{w}(r_p)$ using $n_{box}$, appropriate if this box were the entire survey, and one estimated

$n = N_{LRG,box}/L^3$:

$$w_p(r_p) = \frac{N(r_p)}{n_{true}L^3_{box}A(r_p)} - 2\Delta z_{max}$$  \hspace{1cm} (3.16)

$$\tilde{w}_p(r_p) = \frac{w_p(r_p) + 2\Delta z_{max}}{(1 + b\Delta aD(a_{uni})/D(1))^2} - 2\Delta z_{max}$$  \hspace{1cm} (3.17)

We have used Eqn. 3.3 to relate $\tilde{w}_p(r_p)$ to $w_p(r_p)$.

Since the comparison of $w(r_p)$ or $\tilde{w}(r_p)$ to the results for the SDSS LRGs (Masjedi et al.,
2006) is still imprecise due to the small volume of our simulation, we consider two values of \( \sigma_{\log M} \) here, \( \sigma_{\log M} = 0.3 \) and \( \sigma_{\log M} = 0.8 \). For the latter catalog, the halo mass limit of the halo catalogs of mass \( 50M_\odot \) becomes important (\( 1.69 \times 10^{13}M_\odot, \ 7.15 \times 10^{12}M_\odot, \) and \( 3.66 \times 10^{12}M_\odot \) for the low, medium, and high resolution catalogs respectively), since halos at the mass limit of the high resolution simulation host mock LRGs. For the high \( \sigma_{\log M} \) catalogs, we therefore set \( \sigma_{\log M, high} = 0.8, \sigma_{\log M, med} = 0.92, \) and \( \sigma_{\log M, low} = 1.8 \) to match the large scale bias of all three catalogs (Figs 3.36 and 3.37). Table 3.1 shows the HOD parameters of both mocks for each resolution simulation. Figures 3.24 and 3.27 compare \( N_{cen}(M) \) and \( N_{sat}(M) \) for the 3 resolutions for both catalogs. They are in excellent agreement, except for the low resolution catalog at \( \sigma_{\log M} = 1.8 \). In Figures 3.25 and 3.28, we show the product of \( \langle N_{cen}(M) (1 + N_{sat}(M)) \rangle \) with the number of halos per logarithmic mass bin (\( \Delta \log_{10} M = 0.2 \)) in the halo catalog. This is just the average number of LRGs residing in a halo of mass \( M_i - 1.58M_i \). For the \( \sigma_{\log M} = 0.3 \) catalog, the number of LRGs in halos near the halo catalog mass limit is negligible; this is not the case for the high \( \sigma_{\log M} \) catalog. This is the source of discrepancy between HOD parameters derived from the different resolution simulations. Because increasing \( \sigma_{\log M} \) decreases the probability of hosting a galaxy for a range of higher masses, the satellite parameters must compensate for the vacancy of some high mass halos. Therefore \( M_1 \) has decreased as we increased \( \sigma_{\log M} \) in the lower resolution simulations to match the large scale clustering. Figures 3.26 and 3.29 show that the resulting distribution of satellite galaxies as a function of halo mass is in good agreement between the three resolution mocks. From Figure 3.28 we conclude that the low resolution simulation has insufficient mass resolution to reach halos to attain this level of large scale bias. In § 3.8 we argue that the medium resolution simulation will be sufficient.

In Figure 3.30 we compare the high resolution mock LRG power spectra for our \( \sigma_{\log M} = 0.3 \) and \( \sigma_{\log M} = 0.8 \) catalogs to the real space \( P(k) \) reported in Tegmark et al. (2006). For these catalogs we estimate \( b_{\sigma=0.3} = 2.35 \) and \( b_{\sigma=0.8} = 2.05 \) in the large scale limit, which in linear theory (Eqn. 1.20) implies that the ratio of the redshift space to real space
Figure 3.24: The \( \langle N_{\text{cen}}(M) \rangle \) (plateaus to 1) and \( \langle N_{\text{sat}}(M) \rangle \) vs. \( M \) in units of \( 10^{14} M_\odot \) for the HOD parameters listed in Table 3.1 for the low (blue), medium (red), and high (green) resolution mocks with \( \sigma_{\log M} = 0.3 \).

Power spectrum is \( \approx 1.2 - 1.24 \). We divide our redshift space power spectra by this factor to compare with the real space \( P(k) \) reported in Tegmark et al. (2006). Note that in normalizing the power spectra from our mocks, we used \( \bar{n} = N_{\text{objects}} / L_{\text{uni}}^3 \); to match the expected \( \bar{n}_{\text{uni,LRG}} \), these spectra should be further multiplied by \( (1 + b \Delta_{\text{box}})^2 (1 - \Delta_{\text{box}} / 3)^6 = 0.978 \). Furthermore, the underdensity of this realization means that linear modes have grown less by the factor \( D(z, \Delta) / D(z, \Delta = 0) = 0.992 \). In combination, we expect the results shown in Figure 3.30 to be diminished by 0.6% in a \( \Delta_{\text{o}} = 0 \) box, at least for modes still in the linear regime. The \( \sigma_{\log M} = 0.8 \) catalog \( P(k) \) is smaller than the Tegmark et al. (2006) \( P(k) \) by 10-15%, ensuring that our chosen values of \( \sigma_{\log M} \) safely bracket the observed amplitude.

Figure 3.31 and 3.32 compare the projected correlation function in the 3 resolution mock catalogs with \( \sigma_{\log M} = 0.3 \). The agreement is very good for separations \( r_p > 0.026 \) Mpc/\( h \), sufficiently beyond the spatial resolution of the simulations. We compute a \( \chi^2 \) on the last 15 points to assess the level of consistency between the mock catalogs drawn from the different resolution simulations. As we discussed in § 3.4.1, quantifying the expected variation between the simulations is not possible. For the \( \chi^2 \) discussed here, we adopt error bars that are the estimated error on the average \( w_p(r_p) \) computed with 40 independent mock catalogs produced from each resolution simulation’s SO catalog; consistency at this level assures that any discrepancies due to the mass resolution are \( \sim 1/40 \) of the variance.
Figure 3.25: The average number of central and satellite LRGs in each logarithmic mass bin with $\Delta \log_{10} M = 0.2$ vs. M in units of $10^{14} M_\odot$. Low resolution (blue), medium (red), and high (green) resolution mocks are shown for $\sigma_{\log M} = 0.3$ catalogs.

Figure 3.26: The average number of satellite LRGs in each logarithmic mass bin with $\Delta \log_{10} M = 0.2$ vs. M in units of $10^{14} M_\odot$. Low resolution (blue), medium (red), and high (green) resolution mocks are shown for $\sigma_{\log M} = 0.3$.

Figure 3.27: The $\langle N_{\text{cen}}(M) \rangle$ (plateaus to 1) and $\langle N_{\text{sat}}(M) \rangle$ vs. M in units of $10^{14} M_\odot$ for the HOD parameters listed in Table 3.1 for the low (blue), medium (red), and high (green) resolution mocks with $\sigma_{\log M, \text{high}} = 0.8$ (green), $\sigma_{\log M, \text{med}} = 0.92$ (red), and $\sigma_{\log M, \text{low}} = 1.8$ (blue).
Figure 3.28: The average number of central and satellite LRGs in each logarithmic mass bin with $\Delta \log \log M = 0.2$ vs. $M$ in units of $10^{14} M_\odot$ for catalogs with $\sigma_{\text{log}M,\text{high}} = 0.8$ (green), $\sigma_{\text{log}M,\text{med}} = 0.92$ (red), and $\sigma_{\text{log}M,\text{low}} = 1.8$ (blue). Note that the SO catalogs go to 50 particles/halo, which corresponds to a mass limit of $1.69 \times 10^{13} M_\odot$, $7.15 \times 10^{12} M_\odot$, and $3.66 \times 10^{12} M_\odot$ for the low, medium, and high resolution catalogs.

Figure 3.29: The average number of satellite LRGs in each logarithmic mass bin with $\Delta \log_{10} M = 0.2$ vs. $M$ in units of $10^{14} M_\odot$ for catalogs with $\sigma_{\text{log}M,\text{high}} = 0.8$ (green), $\sigma_{\text{log}M,\text{med}} = 0.92$ (red), and $\sigma_{\text{log}M,\text{low}} = 1.8$ (blue).
Figure 3.30: Comparison of our high resolution simulation $\sigma_{\log M} = 0.3$ (blue), $\sigma_{\log M} = 0.8$ (red), and Tegmark et al. (2006) real space $P(k)$ (green) in $(\text{Mpc}/h)^3$ vs $k$ in $(\text{Mpc}/h)^{-1}$. We estimate the real space $P(k)$ from our redshift space power spectrum by dividing by the linear theory $P_s/k$ with $\beta = f/b_{\text{LRG}}$ and $b_{\text{LRG}}$ estimated from the $k = 0$ limit of $P_{\text{LRG}}(k)/P_{\text{IC}}(k)$. For $\sigma_{\log M} = 0.3$, $b_{\text{LRG}} = 2.35$ and for $\sigma_{\log M} = 0.8$, $b_{\text{LRG}} = 2.05$. These two values of $\sigma_{\log M}$ bracket the observed large scale amplitude. The comparison is not perfect, of course, since our simulation is small compared to the SDSS LRG volume and slightly underdense. We have not applied the FOG compression in Tegmark et al. (2006); this should not matter for the amplitude of the largest modes.

introduced by the HOD stochasticity. Comparing low to high, we find $\chi^2 = 40$; medium to high, $\chi^2 = 28$; and low to medium, $\chi^2 = 15$. We expect these to be $\sim N_{\text{DOF}} = 15$ only if there are no discrepancies in the underlying DM/halo density field from simulation to simulation. Therefore, $\chi^2 = 28$ between the medium and high resolution simulation indicates that the level of variation induced by systematic resolution effects and stochastic variation from the halo finder and $N$-body simulation phase space divergences is $\sim 1/40$ of the variance induced by the HOD stochasticity. We consider this an acceptable level of systematic bias in our mock catalogs. The most significant discrepancy between low and high and medium and high resolution simulations comes at $r_p = 1.05$ and 1.66 Mpc/h, where $\Delta w_p(1.05) = 11.7 \pm 2.7; \Delta w_p(1.66) = 4.2 \pm 1.3$ for low compared with high, and $\Delta w_p(1.05) = 7.0 \pm 2.7; \Delta w_p(1.66) = 4.7 \pm 1.3$ for medium to high. Assuming the errors on the SDSS measured $w_p$ scale by the square root of the survey area, the $1\sigma$ errors will be 12 and 8 at these values of $r_p$. Therefore, even if the discrepancy is entirely due to resolution effects of our simulation, they are at or below the level measurable from the entire survey.
Table 3.1: HOD parameters for $\sigma_{\log M} = 0.3$ for the 3 different resolution simulations, and for the large $\sigma_{\log M}$ catalogs, with $\sigma_{\log M, high} = 0.8$, $\sigma_{\log M, med} = 0.92$, and $\sigma_{\log M, low} = 1.8$. $M_{\text{min}}$ was adjusted for each resolution to match a fixed $N_{\text{LRG,tot}}$. Masses in units of $10^{14} M_\odot$.

<table>
<thead>
<tr>
<th>res</th>
<th>$M_{\text{min}}$</th>
<th>$\sigma_{\log_{10} M}$</th>
<th>$M_{\text{cut}}$</th>
<th>$M_1$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>0.490</td>
<td>0.3</td>
<td>0.386</td>
<td>5.72</td>
<td>1.36</td>
</tr>
<tr>
<td>med</td>
<td>0.495</td>
<td>0.3</td>
<td>0.387</td>
<td>5.74</td>
<td>1.36</td>
</tr>
<tr>
<td>high</td>
<td>0.498</td>
<td>0.3</td>
<td>0.484</td>
<td>5.84</td>
<td>1.26</td>
</tr>
<tr>
<td>low</td>
<td>1.605</td>
<td>1.8</td>
<td>0.811</td>
<td>3.14</td>
<td>0.96</td>
</tr>
<tr>
<td>med</td>
<td>1.218</td>
<td>0.92</td>
<td>0.625</td>
<td>4.26</td>
<td>0.98</td>
</tr>
<tr>
<td>high</td>
<td>1.082</td>
<td>0.8</td>
<td>0.533</td>
<td>4.80</td>
<td>1.03</td>
</tr>
<tr>
<td>low</td>
<td>0.727</td>
<td>0.8</td>
<td>0.540</td>
<td>5.10</td>
<td>1.09</td>
</tr>
<tr>
<td>med</td>
<td>0.995</td>
<td>0.8</td>
<td>0.512</td>
<td>4.82</td>
<td>1.07</td>
</tr>
</tbody>
</table>

$w_p(r_p)$ shows good agreement with the Masjedi et al. (2006) measurement up to the last few points. Since $\Delta_0 < 0$ in this realization, we would expect the clustering in the simulation to be slightly weaker than the data. Figure 3.32 shows that if we adjust $\bar{n}$ to the box value, the agreement with the data is fairly good. However, $\sigma_{\log M}$ is likely lower than the best fit value for a larger volume (with $\Delta_0$ closer to 0), since it falls slightly above the data in Fig 3.32. Moreover, the amplitude of the redshift space power LRG power spectrum is large compared with the Tegmark et al. (2006) measured values, supporting the conclusion that the best fit $\sigma_{\log M}$ for this cosmology will be larger so that the overall bias of the LRG sample will decrease. However, for $\sigma_{\log M} = 0.3$, the mocks from the different resolution boxes are sufficiently consistent in the projected correlation function.

Another statistic that contains information about moments of the higher order density field relevant to our mocks is the group multiplicity bias $b(n_{\text{sat}}) = N_{\text{CiC}}(n_{\text{sat}})/N_{\text{HOD}}(n_{\text{sat}})$, the ratio of the group multiplicity function measured using the CiC algorithm to the true one-halo group multiplicity function. As in our $w_p(r_p)$ analysis, the error bars are estimated errors on each $\bar{b}(n_{\text{sat}})$ averaged over 40 independent HOD realizations. Table 3.2 shows that the group multiplicity biases $b(n_{\text{sat}})$ between different simulation resolutions are consistent with one another at this level, and therefore the systematic bias introduced by the simulation mass resolution is sufficiently small. Above $n_{\text{sat}} = 3$, the number of groups in a particular
Figure 3.31: $\log_{10} w_p(r_p)$ vs. $\log_{10} r_p$, both in Mpc/$h$, for $\sigma_{\log M} = 0.3$. Low (blue), medium (red), and high (green) resolution are shown with Masjedi et al. (2006) SDSS measurements. The error bars on the simulated catalogs show the error on the average $w_p(r_p)$ for the simulated box arising from the stochasticity of populating halos after averaging over 40 HOD realizations; this does not represent any cosmic variance from dark matter/halo density fields.

Figure 3.32: Same as Figure 3.31, but for $\bar{w}_p$ as defined in Eqn. 3.17.
<table>
<thead>
<tr>
<th>$n_{sat}$</th>
<th>$b_{low}$</th>
<th>$\sigma_{b,low}$</th>
<th>$b_{med}$</th>
<th>$\sigma_{b,med}$</th>
<th>$b_{high}$</th>
<th>$\sigma_{b,high}$</th>
<th>$\langle N_{HOD,high} \rangle$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.179</td>
<td>0.004</td>
<td>1.190</td>
<td>0.004</td>
<td>1.184</td>
<td>0.005</td>
<td>728.825</td>
</tr>
<tr>
<td>2</td>
<td>1.079</td>
<td>0.013</td>
<td>1.088</td>
<td>0.016</td>
<td>1.065</td>
<td>0.015</td>
<td>100.500</td>
</tr>
<tr>
<td>3</td>
<td>0.882</td>
<td>0.032</td>
<td>0.923</td>
<td>0.037</td>
<td>0.905</td>
<td>0.042</td>
<td>22.400</td>
</tr>
<tr>
<td>4</td>
<td>0.736</td>
<td>0.068</td>
<td>0.790</td>
<td>0.074</td>
<td>0.751</td>
<td>0.053</td>
<td>7.250</td>
</tr>
<tr>
<td>5</td>
<td>0.665</td>
<td>0.085</td>
<td>0.650</td>
<td>0.115</td>
<td>0.562</td>
<td>0.081</td>
<td>2.150</td>
</tr>
<tr>
<td>6</td>
<td>0.375</td>
<td>0.090</td>
<td>0.352</td>
<td>0.067</td>
<td>0.567</td>
<td>0.116</td>
<td>0.625</td>
</tr>
<tr>
<td>7</td>
<td>0.077</td>
<td>0.042</td>
<td>0.400</td>
<td>0.077</td>
<td>0.143</td>
<td>0.055</td>
<td>0.175</td>
</tr>
</tbody>
</table>

Table 3.2: Group biases and errors on the average after 40 HOD realizations with $\sigma_{\log M} = 0.3$. The eighth column shows the average number of one-halo groups with $n_{sat}$ satellites in the high resolution simulation. The number of CiC groups on average is just column 8 multiplied by column 6.

Each realization is too low for accurate comparisons in our small simulation volume.

Finally we compare the mock catalog $P(k)$ for the 3 resolution simulations and $\sigma_{\log M} = 0.3$. We consider $P(k)$ for the central galaxies only, for central and satellite galaxies, and for the reconstructed halo density field (see Ch. 4 for the details of reconstructing this field). Figure 3.33 and Figure 3.34 show the ratio of the low to medium and medium to high resolution simulation $P(k)$ for the LRG central+satellite mock catalogs; the results are very similar for the central galaxy only and reconstructed halo density field $P(k)$’s. Again the error bars are the error on the average over 40 HOD realizations and represent an acceptable level of systematic resolution bias. At this level of error, the ratios are consistent with 1; for both LRG power spectra and LRG-dark matter cross spectra, the ratios of low to medium, medium to high, and low to high resolution spectra are all consistent with $1 \pm 0.01$ and show no significant dependence on $k$. Therefore in the redshift space power spectrum of the $\sigma_{\log M} = 0.3$ catalogs, we have not detected any significant resolution dependence.

Next we consider our higher $\sigma_{\log M}$, lower biased catalogs with $\sigma_{\log M,high} = 0.8$, $\sigma_{\log M,med} = 0.92$, and $\sigma_{\log M,low} = 1.8$. These values were chosen to get a good match to the large scale bias. Figure 3.35 shows the ratio of $P(k)$ from the medium to the high resolution
Figure 3.33: The ratio of the average redshift space power spectra vs. $k$ in units $(\text{Mpc}/h)^{-1}$ of the low resolution simulation to the medium resolution simulation for the $\sigma_{\log M} = 0.3$ catalogs, averaged over 40 HOD realizations, from which we derive the HOD-induced error bars representing an acceptable level of systematic resolution bias.

Figure 3.34: Same as Figure 3.33, but for the ratio of the medium to high resolution simulation for the $\sigma_{\log M} = 0.3$ catalogs.
Figure 3.35: The ratio of the average redshift space power spectra of the medium resolution simulation $\sigma_{\log M} = 0.92$ catalogs to the high resolution simulation $\sigma_{\log M} = 0.8$ catalogs vs. $k$ in units $\left(Mpc/h\right)^{-1}$. We averaged over 40 HOD realizations, from which we derive the HOD-induced error on the average.

mock catalog. The amplitude is $\sim 1\%$ too small without substantial dependence on $k$; this offset could be fixed with by slightly varying the $\sigma_{\log M}$ of the medium resolution catalog. The $P(k)$ amplitude of the low resolution catalog is $\sim 4\%$ too large. It is difficult to lower the bias any more because the mass limit on the low resolution catalog is so high (see Figure 3.28).

If the discrepancy between our high $\sigma_{\log M}$ catalogs of different resolution is only due to mass function differences or halo catalog mass limits and not any two-point selection effects, then adjusting $\sigma_{\log M}$ to match the large scale bias will hopefully leave the two-point statistics intact. Since the precise form of Eqn. 2.8 is not physically motivated, it seems harmless to modify its form through multiplication with $n_{\text{low}}(M)/n_{\text{true}}(M)$ for the lower resolution simulations, if we can still match all of the desired correlation statistics. However, the hard mass limit of the halo catalogs does introduce a sharp cutoff in $N_{\text{cen}}(M)$ if $N_{\text{cen}}(M)$ is non-negligible at the catalog mass limit. Figure 3.36 and Figure 3.37 compare the high resolution $\sigma_{\log M} = 0.8$ mocks with a medium resolution $\sigma_{\log M} = 0.92$ and low resolution $\sigma_{\log M} = 1.8$. There is excellent agreement between the medium and high resolution catalog $w_p(r_p)$. We compute a $\chi^2$ as for the $\sigma_{\log M} = 0.3$ catalogs and find $\chi^2 = 20$ for the last 15 data points, indicating no statistically significant systematic resolution bias.
Figure 3.36: $\log_{10} w_p(r_p)$ vs. $\log_{10} r_p$, both in Mpc/$h$, for $\sigma_{\log M,\text{high}} = 0.8$, $\sigma_{\log M,\text{med}} = 0.92$, and $\sigma_{\log M,\text{low}} = 1.8$ catalogs. Low (blue), medium (red), and high (green) resolution are shown with Masjedi et al. (2006) SDSS measurements. The error bars on the simulated catalogs show the error on the average $w_p(r_p)$ for the simulated box arising from the stochasticity of populating halos after averaging over 40 HOD realizations; this does not represent any cosmic variance from dark matter/halo density fields.

Figure 3.37: Same as Figure 3.36, but for $\bar{w}_p$ as defined in Eqn. 3.17.
In Table 3.3 we compare the bias factors \( b(n_{sat}) \) for the group multiplicity functions in each resolution’s catalog (\( \sigma_{\log M, low} = 1.8, \sigma_{\log M, med} = 0.92, \sigma_{\log M, high} = 0.8 \)). We detect a discrepancy in the bias factors at \( n_{sat} = 1 \) and \( n_{sat} = 2 \) between the medium and high resolution. This systematic difference is at the level of noise introduced by the HOD stochasticity (\( \sim 1/\sqrt{N_{true}(n_{sat} = 1)} = 0.04 \)). As long as the variance in the underlying matter density field introduces comparable or larger variance, the bias in \( b(n_{sat}) \) will have a small effect on the mock catalog statistics. Moreover, \( b(n_{sat}) \) could be brought into better agreement by modifying the form of \( N_{cen}(M) \) (i.e., by only adjusting the low mass end to account for the halo catalog mass limit). However, the drastically larger \( \sigma_{\log M} \) required for the low resolution simulation significantly changes the HOD and resulting higher order density field statistics. Therefore we do not recommend relying on such a low resolution catalog. In Table 3.4 we show that there is very good agreement between the bias factors when comparing equal \( \sigma_{\log M} \) catalogs directly; the HOD parameters between the medium and high resolution simulation in this case are in excellent agreement (Table 3.1). This demonstrates that \( N_{sat} \) dominates the behavior of \( b(n_{sat}) \); the large scale bias of the catalog is not as important. Therefore, we should be able to match both the small scale behavior in \( w_p(r_p) \) and \( b(n_{sat}) \), and the large scale bias in \( P(k) \) between the medium and high resolution catalogs by modifying \( N_{cen} \) asymmetrically. We have demonstrated sufficient agreement (with minor tweaking) between the medium and high resolution catalogs for the large \( \sigma_{\log M} \) catalogs.

### 3.8 Summary of MID sample \( (z = 0.34) \) Results

There is a slight but systematic decrease in power with \( k \) as the simulation resolution decreases. This is a 0.2% effect at \( k = 0.2 \) Mpc/h between the medium and high resolution simulations.

The mass functions from the low, medium, and high resolution simulations are in excel-
Table 3.3: Group biases and errors on the average after 40 HOD realizations with $\sigma_{\log M, low} = 1.8$, $\sigma_{\log M, med} = 0.92$, $\sigma_{\log M, high} = 0.8$. The eighth column shows the average number of one-halo groups with $n_{\text{sat}}$ satellites in the high resolution simulation. The number of CiC groups on average is just column 8 multiplied by column 6.

<table>
<thead>
<tr>
<th>$n_{\text{sat}}$</th>
<th>$b_{\text{low}}$</th>
<th>$\sigma_{b,\text{low}}$</th>
<th>$b_{\text{med}}$</th>
<th>$\sigma_{b,\text{med}}$</th>
<th>$b_{\text{high}}$</th>
<th>$\sigma_{b,\text{high}}$</th>
<th>$\langle N_{\text{HOD, high}} \rangle$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.324</td>
<td>0.006</td>
<td>1.194</td>
<td>0.006</td>
<td>1.155</td>
<td>0.004</td>
<td>698.025</td>
</tr>
<tr>
<td>2</td>
<td>0.931</td>
<td>0.013</td>
<td>0.901</td>
<td>0.011</td>
<td>0.958</td>
<td>0.012</td>
<td>130.550</td>
</tr>
<tr>
<td>3</td>
<td>0.757</td>
<td>0.018</td>
<td>0.791</td>
<td>0.025</td>
<td>0.783</td>
<td>0.029</td>
<td>33.675</td>
</tr>
<tr>
<td>4</td>
<td>0.739</td>
<td>0.041</td>
<td>0.730</td>
<td>0.042</td>
<td>0.683</td>
<td>0.037</td>
<td>9.550</td>
</tr>
<tr>
<td>5</td>
<td>0.589</td>
<td>0.054</td>
<td>0.677</td>
<td>0.096</td>
<td>0.824</td>
<td>0.157</td>
<td>3.125</td>
</tr>
<tr>
<td>6</td>
<td>0.631</td>
<td>0.134</td>
<td>0.403</td>
<td>0.080</td>
<td>0.478</td>
<td>0.097</td>
<td>1.000</td>
</tr>
<tr>
<td>7</td>
<td>0.349</td>
<td>0.065</td>
<td>0.143</td>
<td>0.055</td>
<td>0.083</td>
<td>0.044</td>
<td>0.350</td>
</tr>
</tbody>
</table>

Table 3.4: Group biases and errors on the average after 40 HOD realizations with $\sigma_{\log M, low} = 0.8$, $\sigma_{\log M, med} = 0.8$, $\sigma_{\log M, high} = 0.8$. The eighth column shows the average number of one-halo groups with $n_{\text{sat}}$ satellites in the high resolution simulation. The number of CiC groups on average is just column 8 multiplied by column 6.

<table>
<thead>
<tr>
<th>$n_{\text{sat}}$</th>
<th>$b_{\text{low}}$</th>
<th>$\sigma_{b,\text{low}}$</th>
<th>$b_{\text{med}}$</th>
<th>$\sigma_{b,\text{med}}$</th>
<th>$b_{\text{high}}$</th>
<th>$\sigma_{b,\text{high}}$</th>
<th>$\langle N_{\text{HOD, high}} \rangle$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.183</td>
<td>0.005</td>
<td>1.158</td>
<td>0.005</td>
<td>1.155</td>
<td>0.004</td>
<td>698.025</td>
</tr>
<tr>
<td>2</td>
<td>0.952</td>
<td>0.012</td>
<td>0.950</td>
<td>0.011</td>
<td>0.958</td>
<td>0.012</td>
<td>130.550</td>
</tr>
<tr>
<td>3</td>
<td>0.808</td>
<td>0.028</td>
<td>0.799</td>
<td>0.022</td>
<td>0.783</td>
<td>0.029</td>
<td>33.675</td>
</tr>
<tr>
<td>4</td>
<td>0.854</td>
<td>0.060</td>
<td>0.685</td>
<td>0.054</td>
<td>0.683</td>
<td>0.037</td>
<td>9.550</td>
</tr>
<tr>
<td>5</td>
<td>0.541</td>
<td>0.068</td>
<td>0.558</td>
<td>0.062</td>
<td>0.824</td>
<td>0.157</td>
<td>3.125</td>
</tr>
<tr>
<td>6</td>
<td>0.421</td>
<td>0.086</td>
<td>0.779</td>
<td>0.106</td>
<td>0.478</td>
<td>0.097</td>
<td>1.000</td>
</tr>
<tr>
<td>7</td>
<td>0.321</td>
<td>0.071</td>
<td>0.143</td>
<td>0.055</td>
<td>0.083</td>
<td>0.044</td>
<td>0.350</td>
</tr>
</tbody>
</table>

There is a slight increase (~5%) in the halo mass function as the mass limit $50M_\odot$ of our SO halo catalogs is reached; this should be kept in mind in future analyses, but is not inherently problematic since slight changes in the mass function in this regime are degenerate with small variations in the form of $N_{cen}(M)$.

We split the halos into two mass bins: $50 - 99M_{p,med}$ and $M \geq 100M_{p,med}$. This corresponds to 98 - 195 particles per halo for the low mass bin of the high resolution simulation. We did not detect any significant difference in $P(k)$ and $\xi(r)$ besides a $\lesssim 1\%$ difference
in the overall bias of the mass bins. We also compute the velocity correlation statistic
\[ \langle \vec{v}_i(\vec{r}) \cdot \vec{v}_j(\vec{r}') \rangle / \sqrt{\langle \vec{v}_i^2 \rangle \langle \vec{v}_j^2 \rangle} \]
and find generally good agreement between the medium and high resolution simulations. However, Figure 3.21 indicates a slight bias in this statistic for the mass bin at the \( \sim 0.001 \) level, or few percent level as one reaches \( r \sim 150 \text{ Mpc}/h \) where
\[ \langle \vec{v}_i(\vec{r}) \cdot \vec{v}_j(\vec{r}') \rangle / \sqrt{\langle \vec{v}_i^2 \rangle \langle \vec{v}_j^2 \rangle} = 0.03. \]
Since this statistic is independent of mass bin at large separations (Fig. 3.19), this discrepancy cannot be explained by variation in the bias of the halo mass bins between resolutions. Instead, this could indicate a halo selection effect, and would be of concern if we tried to do precise measurements of the baryon acoustic oscillation (BAO) using halos with mass \( M \lesssim 100 M_p \). We also examined the relative halo velocities along the vector separating the halos:
\[ \langle (\vec{v}_i(\vec{r}) - \vec{v}_j(\vec{r}')) \cdot (\vec{r} - \vec{r}')/|\vec{r} - \vec{r}'| \rangle. \]
The statistics from the medium resolution simulation agree to within 1 km/s with the high resolution simulation results. The slight discrepancy (\( \lesssim 2\% \)) at moderate to large separations could perhaps be explained by a slight difference in halo bias between the medium and high resolution simulations.

For each of the three resolution simulations we have made two mock catalogs that bracket the amplitude of the Tegmark et al. (2006) LRG \( P(k) \) (Fig 3.30). We compare their \( P(k) \), \( w_p(r_p) \), and group multiplicity bias factors \( b(n_{sat}) \) for groups with \( n_{sat} = 1, \ldots, 7 \).

Both the medium and low resolution simulation \( \sigma_{\log M} = 0.3 \) mocks are able to match the high resolution mocks for these statistics. The only hint of discrepancy is in \( w_p(r_p) \) at \( \sim 1 \) Mpc/h, but this is below the expected errors on \( w_p(r_p) \) from the full SDSS survey.

The high \( \sigma_{\log M} \) mocks show some resolution dependence. Figure 3.28 shows that halos at the mass limit of each catalog are contributing a non-negligible fraction of the mock LRGs. We are therefore introducing an artificial cutoff in \( N_{cen}(M) \) at the mass limit. While the medium resolution simulation can produce a catalog in good agreement with \( P(k) \), \( w_p(r_p) \), and multiplicity bias factors \( b(n_{sat}) \) of the high resolution simulation, we will be restricted from exploring the full range of plausible HODs by this mass limit. We plot the cumula-
Figure 3.38: The cumulative distribution of LRGs as a function of their host halo mass in units of $10^{14} M_\odot$ for the $\sigma_{\log M, \text{high}} = 0.8$ (gold), $\sigma_{\log M, \text{med}} = 0.92$ (red), and $\sigma_{\log M, \text{low}} = 1.8$ (blue) catalogs; this is just the integral of Figure 3.28.

tive distribution of LRGs as a function of their host halo mass in Figure 3.38. The high resolution catalog has 7.7% of its LRGs below the halo catalog mass limit of the medium resolution simulation, $50M_{p,\text{med}} = 7.15 \times 10^{12} M_\odot$. The medium resolution catalog has 13% of LRGs in halos with 50-79 particles, and 21% below 100. Though we have shown that there are no serious systematics in the statistics of the $50-100M_p$ halo mass subsample, this demonstrates that we are close to the limit set by the resolution of the medium simulation.

Figure 3.39 shows that $\sigma_{\log M} = 0.6$ is closer to the best fit value for matching the amplitude to the Tegmark et al. (2006) $P(k)$. As we show in Ch. 4, the discrepancy in this Figure at high $k$ is because we have not applied the FOG compression in Tegmark et al. (2006). For this mock catalog from the high resolution simulation, only 2.6% of the mock LRGs occupy halos below the halo catalog mass limit of the medium resolution simulation, and only 11% are below $100M_{p,\text{med}}$. Figure 3.40 shows the cumulative distribution of LRGs as a function of host halo mass for $\sigma_{\log M} = 0.6, 0.7, \text{ and } 0.8$. The halo mass limit of the medium resolution catalog ($7.15 \times 10^{12} M_\odot$) is much less important at $\sigma_{\log M} = 0.6$. We conclude that the medium resolution simulation parameters are close to optimal for the time, desired number of realizations, and resolution constraints for catalogs at $z = 0.34$. 
Figure 3.39: $P(k)$ in units of $(\text{Mpc}/h)^3$ vs $k$ in $(\text{Mpc}/h)^{-1}$ for the high resolution $\sigma_{\log M} = 0.6$ catalog where we have scaled the redshift space $P(k)$ using Eqn. 1.20 to estimate the real space power spectrum, as in Fig. 3.30.

Figure 3.40: The cumulative distribution of LRGs as a function of their host halo mass in units of $10^{14} M_\odot$ for the high resolution simulation mocks with $\sigma_{\log M} = 0.6$, $\sigma_{\log M} = 0.7$, and $\sigma_{\log M} = 0.8$ catalogs.
3.9 Catalogs at other redshifts

We assume that the detailed comparisons of halo two-point correlation and velocity statistics at $z = 0.34$ extend to the other simulation outputs at $z = 0.235$ and $z = 0.421$. However, we must demonstrate that the halo catalog mass limit of the medium resolution simulation does not affect mock catalogs at the other redshifts. The number density of LRGs at the NEAR, MID, and FAR sample mean redshifts are $n_{LRG}(z = 0.235) = 10^{-4}$ (Mpc/h)$^3$, $n_{LRG}(z = 0.342) = 10^{-4}$ (Mpc/h)$^3$, and $n_{LRG}(z = 0.421) = 4.3 \times 10^{-5}$ (Mpc/h)$^3$ (Zehavi et al., 2005). Figure 6 of Tegmark et al. (2006) shows that the NEAR, MID, and FAR samples have nearly the same $P(k)$; the increase of $D(z)$ with redshift is balanced by the decrease in overall bias of the sample with redshift. We therefore compare catalogs at $z \approx 0.235$ and $z \approx 0.421$ with the $P(k)$ published for the entire catalog. The FAR sample has a lower number density and is intrinsically more luminous on average; it should therefore have a lower satellite fraction than the NEAR and MID sample. We do not account for this difference in this section because we use the CiC multiplicity function from the NEAR+MID sample in Ch. 2 to derive the FAR sample HOD as well. We measure the CiC multiplicity function of the FAR sample in Ch. 4.

Figure 3.41 shows the power spectra of $\sigma_{\log M} = 0.5$, $\sigma_{\log M} = 0.6$ and $\sigma_{\log M} = 0.8$ mock catalogs from the NEAR sample ($z = 0.233$) medium resolution catalog. 15% of the LRGs are between 50 and 100 $M_p$ in the medium resolution simulation catalog with $\sigma_{\log M} = 0.8$, and 8% for the $\sigma_{\log M} = 0.6$ catalog. This is a smaller fraction than for the MID ($z = 0.33$) catalogs. This mass resolution is sufficient for mock catalogs of the NEAR sample.

For a fixed number density sample of LRGs, the host halo mass should decrease with redshift, making the mass resolution constraints stricter at higher redshifts. However, our FAR sample has less than half the number density of the NEAR and MID samples. We have made catalogs for the FAR sample with $\sigma_{\log M} = 0.9$ and $\sigma_{\log M} = 1.2$. Since the satellite fraction in our fits is likely higher than in the real FAR sample, we show $P(k)$ for
Figure 3.41: \( P(k) \) in \( \text{(Mpc}/\text{h})^3 \) vs \( k \) in \( \text{(Mpc}/\text{h})^{-1} \) for the medium resolution NEAR sample \((z = 0.233)\) mock catalogs with \( \sigma_{\log M} = 0.5 \) (light blue), \( \sigma_{\log M} = 0.6 \) (blue) and \( \sigma_{\log M} = 0.8 \) (red). We have estimated the real space \( P(k) \) from the redshift space \( P(k) \) as in Figure 3.30. Tegmark et al. (2006) shown in green.

the central objects only as well as central and satellite galaxies in Figures 3.42 and 3.43. The \( P(k) \) with the correct satellite fraction should lie somewhere in between. In Figure 3.42 we also show a catalog with \( \bar{n} = 10^{-4} \text{(Mpc}/\text{h})^3 \) and \( \sigma_{\log M} = 0.9 \). Because of the form of \( N_{\text{cen}} \) we have chosen, this catalog has a much lower \( P(k) \) amplitude than the catalogs produced with \( \bar{n} = 4.3 \times 10^{-5} \text{(Mpc}/\text{h})^3 \). If the host halo mass did not depend on LRG luminosity, then this catalog could simply be down-sampled to get a smaller \( \bar{n} \) catalog. This emphasizes a deficiency in the form of \( N_{\text{cen}} \) we have chosen – we are able to get a lower amplitude \( P(k) \) without putting so many LRGs near the catalog mass limit, as is required in our \( \bar{n} = 4.3 \times 10^{-5} \text{(Mpc}/\text{h})^{-3} \) catalogs. In future work we will explore other forms for this function. Figure 3.44 shows \( N_{\text{cen}}(M) \) for these three catalogs (with the \( \bar{n} = 10^{-4} \text{(Mpc}/\text{h})^3 \) catalog multiplied by 0.43 to match the mean sample density). Slightly increasing \( N_{\text{cen}}(M) \) for \( M = 1 - 4 \times 10^{13} M_{\odot} \) should be sufficient to attain the correct \( P(k) \) amplitude without the need for halos at smaller masses. Of course, in the real universe LRGs may indeed inhabit halos at smaller masses.

For the \( \bar{n} = 4.3 \times 10^{-5} \text{(Mpc}/\text{h})^3 \) catalogs, Figure 3.42 shows that \( \sigma_{\log M} = 0.9 \) is a lower bound, since the central objects only have the same amplitude as Tegmark et al. (2006),
Figure 3.42: $P(k)$ in $(\text{Mpc}/h)^3$ vs $k$ in $(\text{Mpc}/h)^{-1}$ for the medium resolution FAR sample ($z = 0.42$) mock catalogs with $\sigma_{\log M} = 0.9$, $\bar{n} = 4.3 \times 10^{-5}$ $(\text{Mpc}/h)^3$ central objects only (blue), central and satellite galaxies (red), and $P(k)$ published in Tegmark et al. (2006) (green). We have estimated the real space $P(k)$ from the redshift space $P(k)$ as in Figure 3.30. The cyan curve shows the power spectrum of a galaxy catalog with $\sigma_{\log M} = 0.9$ and $\bar{n} = 10^{-4}$ $(\text{Mpc}/h)^3$. Because of the form of $N_{\text{cen}}(M)$ we have chosen, this catalog has a lower $P(k)$ amplitude than the lower density catalogs.

while $\sigma_{\log M} = 1.2$ is an upper bound, since all the satellites are necessary to reach the amplitude of $P(k)$ from Tegmark et al. (2006). The cumulative distribution of LRGs for the central and satellite catalogs at $\sigma_{\log M} = 0.9$ and $\sigma_{\log M} = 1.2$ ($\bar{n} = 4.3 \times 10^{-5}$ $(\text{Mpc}/h)^3$) and $\sigma_{\log M} = 0.9$, $\bar{n} = 10^{-4}$ $(\text{Mpc}/h)^3$ are shown in Figure 3.45. For $\sigma_{\log M} = 0.9$ ($\bar{n} = 10^{-4}$ $(\text{Mpc}/h)^3$), 14% of LRGs have mass below $100 M_\odot$; for $\sigma_{\log M} = 1.2$, this mass range holds 22% of the galaxies. Moreover, between $7 \times 10^{12} M_\odot$ and $5 \times 10^{13} M_\odot$, the number of LRGs per logarithmic mass bin in halo mass is roughly constant, suggesting that LRGs may be present in lower mass halos as well. For the $\sigma_{\log M} = 0.9$ ($\bar{n} = 10^{-4}$ $(\text{Mpc}/h)^3$) catalog, there are also 22% below $100 M_\odot$. However, the $P(k)$ amplitude for that catalog is low, so in a better matched catalog, this fraction would decrease.

3.10 Conclusion

The goal of this chapter was to determine the optimal mass resolution for a simulation set with the largest possible total volume, enough realizations to estimate the covariance matrix of the power spectrum, and sufficient mass resolution to avoid systematic errors in
Figure 3.43: $P(k)$ in $(\text{Mpc}/h)^3$ vs $k$ in $(\text{Mpc}/h)^{-1}$ for the medium resolution FAR sample ($z = 0.42$) mock catalogs with $\sigma_{\log M} = 1.2$ central objects only (blue), central and satellite galaxies (red), and $P(k)$ published in Tegmark et al. (2006) (green). We have estimated the real space $P(k)$ from the redshift space $P(k)$ as in Figure 3.30.

Figure 3.44: $N_{\text{cen}}(M)$ vs $M$ in units of $10^{14} M_\odot$ for the three FAR ($z = 0.42$) mock catalogs: $\sigma_{\log M} = 0.9$, $\bar{n} = 4.3 \times 10^{-5} (\text{Mpc}/h)^3$ (red), $\sigma_{\log M} = 1.2$, $\bar{n} = 4.3 \times 10^{-5} (\text{Mpc}/h)^3$ (green), and $\sigma_{\log M} = 0.9$, $\bar{n} = 10^{-4} (\text{Mpc}/h)^3$ downsampling by a 0.43 (blue).

Figure 3.45: The cumulative distribution of LRGs as a function of their host halo mass $M$ in units of $10^{14} M_\odot$ for the medium resolution FAR sample ($z = 0.42$) mock catalogs: $\sigma_{\log M} = 0.9$, $\bar{n} = 4.3 \times 10^{-5} (\text{Mpc}/h)^3$ (blue), $\sigma_{\log M} = 1.2$, $\bar{n} = 4.3 \times 10^{-5} (\text{Mpc}/h)^3$ (red), and $\sigma_{\log M} = 0.9$, $\bar{n} = 10^{-4} (\text{Mpc}/h)^3$ (gold).
statistics affecting the structure of mock catalog power spectra and FOGs. We completed 3 periodic box simulations with $L_{\text{box}} = 558 \, \text{Mpc}/h$ and the same initial conditions, but with different numbers of dark matter particles: $N_p = 384^3$, $512^3$, and $640^3$, corresponding to particle masses $M_p = 3.39 \times 10^{11}$, $1.43 \times 10^{11}$, and $7.33 \times 10^{10} \, M_{\odot}$.

In the resolution study we found very small, if any, discrepancies between resolutions in the two-point and higher order Counts-In-Cylinder statistics. The main concern here is reaching a low enough halo mass limit to sufficiently explore the space of possible distributions of LRG host halo masses. We have shown that for a $z = 0.34$ high resolution mock catalog with a WMAP5 cosmology that best fits the measured amplitude of $P(k)$ and SDSS LRG CiC group multiplicity function, only 2.6% of LRGs are in halos below the 50 $M_p$ mass limit of the medium resolution catalog. Therefore, we deem this mass resolution sufficient for the MID redshift sample of Tegmark et al. (2006). As expected, the $z = 0.233$ catalog performs even better at this mass limit. At $z = 0.42$, the situation is less clear. There will be fewer satellites to boost the overall bias, but this has not been quantified from the Counts-In-Cylinders yet. From the discussion in § 3.9, we should be able to make good catalogs at this redshift, but the form of $N_{\text{cen}}(M)$ may need to be modified. The catalogs at this redshift will be slightly more restricted, since there are fewer halos above the catalog mass limit and they are more highly biased with respect to the density field. We are willing to accept this restriction on catalogs of the FAR sample in order to maximize the simulated volume. We adopt the parameters of the medium resolution simulation for the simulation set in Ch. 4.
Chapter 4

Halo Density Field Reconstruction and Application to Large Scale Structure Measurements

This chapter is based on work done in collaboration with David Spergel and Paul Bode and will be submitted for publication to The Astrophysical Journal. Fig. 4.7 is reproduced from Tegmark et al. (2006) and appears with permission of the author.

4.1 Introduction

Galaxy power spectra are an essential link in testing our cosmological models and studying the initial conditions of our universe, and we are now limited by the systematics of relating the galaxy distribution to the underlying matter field. Through a comparison of galaxy samples from 2dFGRS and SDSS, Sanchez & Cole (2007) find that differing scale-dependent biasing between red and blue galaxies cause the cosmological constraints derived from the samples to differ at the $\sim 2\sigma$ level. In this work we focus on the SDSS LRG sample. With an effective volume much larger than the MAIN sample (Eqn. 1.26), the
LRG large scale structure statistics should provide the tightest constraints on the shape of the linear power spectrum. However, their large effective volume and thus low number density is a double-edged sword: the LRG power spectrum requires a large nonlinear correction, and marginalizing over the amplitude of the correction can discard information about the linear spectrum. Several authors have encountered problems when combining the LRG power spectrum results of Tegmark et al. (2006) with CMB data sets. Dunkley et al. (2008) find that when combining with the 5 year WMAP results, the best fit $\Omega_m$ varies systematically with the maximum $k$ from $P_{LRG}(k)$ included in the analysis. When Verde & Peiris (2008) reconstruct the primordial power spectrum with minimal assumptions about its shape, they find that the LRG sample has less statistical power than the SDSS MAIN or 2dFGRS, despite its larger effective volume. Since the authors restrict themselves to $k \leq 0.1$, they are unable to constrain both the nonlinear correction and the primordial power spectrum. The difficulties in both analyses stem from the degeneracy between the power spectrum shape and the potentially large nonlinear correction amplitude. Fig. 4.1 shows that the best fit Tegmark et al. (2006) nonlinear correction is above the statistical error in each $k$-bin for $k \gtrsim 0.07$, while $k \lesssim 0.1 - 0.15$ is typically used in cosmological parameter analyses. Moreover, the $k$-dependence of this large nonlinear correction must be accurate to avoid introducing systematic errors. The solution we propose in this Chapter is to first estimate a halo density field from the LRG density field. The resulting field is nearly linearly biased with respect to the dark matter for $k \leq 0.2$. The blue curve in Fig. 4.1 shows that the nonlinear correction between the dark matter and linear spectrum is much smaller; it is also much easier to calibrate using $N$-body simulations.

The Halo Occupation Distribution (HOD) model (Seljak, 2000; Peacock & Smith, 2000; Cooray & Sheth, 2002) allows us to describe the link between galaxies and dark matter, and provides a framework in which to analyze non-linearities contributing to the power spectrum. Pairs of galaxies can be separated as one- or two-halo (i.e., occupying the same or distinct dark matter halos). As discussed in Seljak (2000) and Schulz & White (2006),
scale dependence of the bias between galaxies and dark matter arises primarily because the one-halo galaxy term is enhanced relative to the dark matter more than the two-halo term. To combat this effect Huff et al. (2007) introduce a configuration space band-power estimator which confines the one-halo contribution to $\sim 2 - 3 \, \text{Mpc}/h$. Hamann et al. (2008) find that for measured galaxy power spectra (SDSS MAIN, 2dFGRS, and SDSS LRG samples), the nonlinearity can be sufficiently modeled as excess shot noise. In analyzing large galaxy redshift surveys, there are three main approaches to disentangling the effects of nonlinear evolution in the matter field, nonlinear redshift space distortions, and nonlinear galaxy biasing to extract the underlying linear matter power spectrum. One can model the nonlinearities with a phenomenological fitting formula and marginalize over its free parameters when fitting cosmological parameters; this technique has been employed in analyses of SDSS LRGs (Tegmark et al., 2006) as well as in 2dFGRS (Cole et al., 2005). The work of Yoo et al. (prep) start with the HOD to predict the nonlinear modifications to large-scale clustering statistics. In principal, cosmological and HOD parameters can be fit simultaneously. In this forward approach, the measured large-scale clustering is compared with predicted clustering given the HOD parameters. The method put forth in this work is an inverse method: we make use of phase information in the galaxy density field to estimate a halo density field. The power spectrum of this field is nearly linearly biased with respect to the underlying nonlinear matter density field. A future extension of this work is to also use the phase information to reconstruct the linear density field (Eisenstein et al., 2007a). This combined approach would “undo” two sources of nonlinearity: galaxy bias and nonlinear dynamics.

In this Chapter, we introduce our halo density field reconstruction method and use $N$-body simulations to test the method. In § 4.4 we apply the CiC method of Ch. 2 to produce high fidelity mock catalogs from 42 $N$-body simulations for three LRG redshift subsamples. § 4.5 describes our method for reconstructing a halo density field. In § 4.6 we present the nonlinear dark matter and mock catalog power spectra and covariance matrices. We ana-
lyze in detail the effects of the FOG compression algorithm employed in the Tegmark et al. (2006) analysis of the LRG power spectrum. We also present results for the redshift space monopole spectrum, which is measured in Percival et al. (2007). We demonstrate that the reconstructed halo density field is linearly biased with respect to the dark matter out to $k = 0.2$ at the 2% level for the MID and FAR subsamples and at the 5% level for the NEAR sample. For $k \leq 0.1$, the traditional regime used for cosmological parameter estimation, the discrepancy is $\leq 1\%$. We also show that the beat-coupling model of Hamilton et al. (2006) fits both the dark matter and mock catalog covariance matrices. Finally we examine the structure of the redshift space distortions as a function of $k$. Throughout this chapter we adopt the cosmological parameters recommended from the latest WMAP5 analysis (Komatsu et al., 2008): $(\Omega_m, \Omega_b, \Omega_{\Lambda}, n_s, \sigma_8, h) = (0.2792, 0.0462, 0.7208, 0.960, 0.817, 0.701)$.

4.2 Background

4.2.1 Summary of Previous Analyses of the SDSS LRG Power Spectrum

Two power spectrum analyses were simultaneously published for the fifth data release of the SDSS. Percival et al. (2007) compute the monopole (i.e., angle-averaged) redshift space power spectrum of the combined sample of MAIN and LRG galaxies. The method extends the FKP method (Feldman et al., 1994) to remove differential bias from the recovered spectrum, and optimally weights the galaxies according to their expected bias, which is determined from their luminosity. This approach leads Percival et al. (2007) to use a different selection function for the LRGs than adopted for our analysis. We do not model the luminosities of the galaxies, and so precise evaluation of their method is not possible here. However, we do compute the monopole redshift power spectrum of our mock catalogs. This should compare most closely to the Percival et al. (2007) method, though the weighting of the galaxies according to their luminosity will certainly alter the relative amplitudes of the one- and two-halo contributions to $P(k)$. 
Figure 4.1: Black points show the error bars divided by the bandpower for the LRG power spectrum published in Tegmark et al. (2006). The black curve shows $(1 + Q_{NL}k^2)/(1 + 1.4k) - 1$, the fractional amplitude of the nonlinear $Q_{NL}$ correction to the linear power spectrum for $Q_{NL} = 30$, the best fit value in Tegmark et al. (2006). For comparison, we show the smooth nonlinear correction fit to the dark matter power spectrum of Fig. 4.4. This is the size of the correction for our reconstructed halo density field.
Tegmark et al. (2006) focus their analysis entirely on the LRGs. They examine three redshift subsamples, NEAR, MID, and FAR. We model each of these samples separately in our analysis. Tegmark et al. (2006) use the PKL method (Tegmark et al., 2004) to estimate the real space galaxy-galaxy, galaxy-velocity, and velocity-velocity power spectra. In linear theory the three spectra are related by (Eqns. 2 and 3 in Tegmark et al. (2006)):

\[ P_{gv}(k) = \beta r_{gv} P_{gg}(k) \]
\[ P_{vv}(k) = \beta^2 P_{gg}(k) \]

\[ \beta = 1/b_{gal} \frac{d\ln D}{d\ln a} \] relates the amplitude of galaxy and velocity field fluctuations in linear theory and \( r_{gv} \) is the dimensionless correlation coefficient between the velocity field and gravitational acceleration field. The final estimate of the real space LRG power spectrum is a linear combination of all three, though it is dominated by the scaled redshift space monopole power spectrum. Their method assumes that the parameters relating the galaxy and velocity power spectra, \( \beta \) and \( r_{ge} \), are scale independent, and they only use data with \( k < 0.09 \ h/\text{Mpc} \) to estimate these factors. To make this assumption more accurate for the LRG density field, Tegmark et al. (2006) compress FOGs before estimating the power spectrum. The algorithm is detailed in Tegmark et al. (2004), but to our knowledge has not been extensively tested on accurate LRG mock catalogs. The algorithm finds groups of LRGs by a Friends-of-Friends (FoF) algorithm (Frenk et al., 1988) and then isotropizes them by scaling all radial separations from the group center. Two objects are considered friends when

\[ \left( \frac{r_{\parallel}}{10} \right)^2 + r_{\perp}^2 \right)^{1/2} \leq \left[ \frac{4}{3} \pi \bar{n}(1 + \delta_c) + r_{\perp,max}^{-3} \right]^{-1/3} \]

with \( \delta_c = 200 \) and \( r_{\perp,max} = 5 h^{-1} \) Mpc; the RHS of Eqn. 4.3 is \( 2.3 h^{-1} \) Mpc for the SDSS LRG NEAR and MID subsample number densities, and \( 2.8 h^{-1} \) Mpc for the FAR sample. For comparison, the virial radius of a \( 10^{14} h^{-1}M_\odot \) halo is \( \sim 1 h^{-1} \) Mpc (\( \sim 2 \) for a \( 10^{15} h^{-1}M_\odot \) halo). Since the LOS distance is compressed in the LHS of Eqn. 4.3, the algorithm will compress objects well beyond the virial radius of the typical host halo. After examination of
the NEAR, MID, and FAR subsamples, Tegmark et al. (2006) conclude that the expected decrease in bias and increase in clustering amplitude with redshift cancel within the observational errors on the subsample spectra. They therefore apply no corrections for evolution of the galaxy population with redshift, and compute a single power spectrum for the entire sample.

4.2.2 Nonlinear Power Spectrum Models

The $Q_{NL}$ model was introduced in Cole et al. (2005) as a fitting formula relating the linear and galaxy power spectra:

$$P_{gal}(k) = \frac{1 + Q_{NL}k^2}{1 + Ak} P_{lin}(k).$$

(4.4)

Both from examining mock catalogs and the HOD model analytically, Cole et al. (2005) fix $A = 1.4$ (redshift space) and $A = 1.7$ (real space) to model the ‘previrialization’ suppression of power on large scales (Lokas et al., 1996). $Q_{NL}$ is considered a nuisance parameter whose amplitude is set by both nonlinearities in the matter field and the scale-dependent bias of the galaxy sample.

Eisenstein et al. (2007b) examine in detail the damping of BAO features in the power spectrum through differential motion of pairs of tracers separated by the BAO scale. Displacements are enhanced in redshift space along the line of sight by $(1 + d \ln D/d \ln a)$. For our real space matter power spectrum we adopt

$$P_{\text{smear}}(k) = P_{\text{lin}}(k)e^{-k^2/2k_{BAO}^2} + P_{\text{no wiggles}}(k) \left(1 - e^{-k^2/2k_{BAO}^2}\right).$$

(4.5)

$P_{\text{no wiggles}}(k)$ is defined by Eqn. 29 of Eisenstein & Hu (1998), and is a smooth version of $P_{\text{lin}}(k)$ with the baryon oscillations removed. Therefore the baryon oscillations remaining in $P_{\text{smear}}(k)$ are damped completely for $k \gg k_{BAO}$ and unaltered for $k \ll k_{BAO}$. In § 4.6.1 we find $k_{BAO} = 0.14 \, h$/Mpc for our real space matter power spectrum, in agreement with the value reported in §6 of Eisenstein et al. (2007b).
Percival et al. (2007) assumes that both $P(k)$ and galaxy bias are linear, and shows that the resulting fits to $\Omega_m$ for $0.01 < k < 0.06 \, h/\text{Mpc}$ and $0.01 < k < 0.15 \, h/\text{Mpc}$ are discrepant at the 2-3$\sigma$ level, demonstrating the need for nonlinear modeling to extract robust cosmological information from the observed galaxy power spectrum. The LRG power spectrum measured in Tegmark et al. (2006) requires a large nonlinear correction (see our Fig. 4.1). They adopt the model

$$P_{\text{LRG}}(k) = P_{\text{smear}}(k, k_{\text{BAO}} = 0.1) b_{\text{LRG}}^2 \frac{1 + Q_N k^2}{1 + A_k}$$  \hspace{1cm} (4.6)$$

with $A$ fixed at either 1.4 or 1.7 (the text states both in different locations in the paper). Their best fit value for the nuisance parameter is $Q_{NL} = 30$. This model is only tested on power spectra of mock LRGs in real space and with no satellite LRGs. We will compare this model to our more realistic mock catalogs in redshift space.

### 4.2.3 Modeling the Covariance Matrix

The covariance matrix for a set of band powers $P(k_{i=1..N})$ is defined by

$$C_{ij} = \langle (P(k_i) - \bar{P}(k_i))(P(k_j) - \bar{P}(k_j)) \rangle.$$  \hspace{1cm} (4.7)$$

Hamilton et al. (2006) showed that the largest nonlinear contribution to the covariance matrix is a beat-coupling term proportional to the power on the largest scale of the survey. In any real survey of finite volume, the observed Fourier amplitudes are convolved with the survey window function $W_s$:

$$\delta_{\text{obs}}(\vec{k}) = \int \delta(\vec{k}') W_s(\vec{k} - \vec{k}') \frac{d\vec{k}'}{(2\pi)^3}$$  \hspace{1cm} (4.8)$$

The beat coupling contribution arises because neighboring Fourier modes $\delta(\vec{k}) \delta(-\vec{k} - \vec{\epsilon})$ are coupled by nonlinear growth to the beat mode $\delta(\vec{\epsilon})$. When the DC mode of the survey is positive, all modes are amplified; when it is negative, all modes are suppressed. This term can be large since the linear power spectrum drops so sharply with $k$. Hamilton et al. (2006) emphasize that this term does not contribute to the covariance of power measured
from ensembles of traditional periodic box simulations, where the band power $k_i$ is averaged over $N_i$ complex modes by

$$P(k_i) = \frac{1}{N_i} \sum_{j=1}^{N_i} \delta(k_{ij})\delta^*(\tilde{k}_{ij}). \quad (4.9)$$

Here, the beat mode is $\tilde{k}_{ij} - \tilde{k}_{ij} = 0$, the DC mode. However, since our simulations allow the DC mode to vary (Sirko, 2005), we will capture this term. We therefore model our covariance matrix as the sum of the usual Gaussian and shot noise contributions and the beat-coupling contribution:

$$C_{ij,\text{model}} = \frac{1}{N_i} \left( P(k_i) + \frac{\alpha_{\text{shot}}}{n} \right)^2 \delta_{ij}^{K} + 4\beta_{\text{beat}} R_{\alpha} P(k_i)P(k_j)\delta_{DC}^2(z). \quad (4.10)$$

Here $\delta_{DC}^2(z) = P_L(k = 0, z = 0)/V_{\text{sim}} \times D^2(z)/D^2(0)$ (Sirko, 2005), the variance of the DC mode linearly evolved to redshift $z$ in the simulation volume $V_{\text{sim}} = L^3$. In perturbation theory $R_{\alpha} \approx 2.62$ (Hamilton et al., 2006). We introduce $\alpha_{\text{shot}}$ and $\beta_{\text{beat}}$ to allow for excess shot noise and variation in the amplitude of the beat coupling term, though we expect both parameters to be $\sim 1$. We show in § 4.6 that this model provides a good model for both the dark matter and LRG covariance matrices. For the dark matter, the shot noise contribution is negligible.

Neyrinck et al. (2006) showed that Poisson fluctuations about the mean halo mass function introduce variance in the amplitude of the one-halo contribution to the dark matter power spectrum that can dominate the covariance matrix in the nonlinear regime. Since our reconstruction of the halo density field seeks to eliminate the one-halo contribution from galaxies, we expect the covariance to be smaller for the reconstructed halo density field than for the original galaxy sample.

### 4.2.4 Error Estimates on $P(k_i)$ and $C_{ij}$

For the dark matter, errors on the bandpowers are straightforward. They are simply the diagonal terms of the inverse covariance matrix divided by the number of simulations. We
use the model covariance matrix in Eqn. 4.10 to compute the inverse.

$$\sigma^2_{P(k_i)} = \frac{1}{N_{\text{sim}}}(C^{-1}_{\text{model}})_{ii}$$  

(4.11)

Of course, the beat coupling leads to off-diagonal terms in both the covariance matrix and its inverse; these terms must be included when estimating errors on model parameters.

We also estimate an error on our estimate of the covariance matrix. We ignore correlations between the $C_{ij}$’s in our error estimates, though they are certainly present.

$$\sigma^2_{C_{ij}} = \frac{1}{N_{\text{sim}}} \langle (\Delta P_i \Delta P_j - C_{ij})^2 \rangle$$

(4.12)

The situation for the mock catalogs is more tricky. For each of the $s = 1, \ldots, N_{\text{sim}}$ TPM simulations we produce $k = 1, \ldots, N_{\text{mocks}}$ mock catalogs using our fixed HOD parameters to reduce the shot noise contribution. We define

$$C_{ij,HOD} = \langle (P_{sk}(k_i) - \bar{P}_{sk}(k_i))(P_{sk}(k_j) - \bar{P}_{sk}(k_j)) \rangle$$  

(4.13)

$$C_{ij,\text{reduced}} = \langle (\bar{P}_{sk}(k_i) - \bar{P}(k_i))(\bar{P}_{sk}(k_j) - \bar{P}(k_j)) \rangle$$  

(4.14)

$$C_{ij,\text{tot}} = \langle (P_{sk}(k_i) - \bar{P}(k_i))(P_{sk}(k_j) - \bar{P}(k_j)) \rangle$$  

(4.15)

$$C_{ij,\text{tot}} = C_{ij,HOD} + C_{ij,\text{reduced}}$$  

(4.16)

Here $P_{sk}(k_i)$ denotes the band power for mock catalog $k$ populating simulation $s$, $\bar{P}_{sk}(k_i)$ denotes the band power in a single simulation $s$ averaged over the $k$ mock catalogs populating $s$, and $\bar{P}(k_i)$ denotes a band power averaged over the entire set of $N_{\text{mocks}} \times N_{\text{sim}}$ catalogs. $C_{ij,\text{tot}}$ is the covariance matrix for this set of mock catalogs. $C_{ij,HOD}$ is the variance introduced by sampling the same matter density field with different mock catalog realizations and $C_{ij,\text{reduced}}$ is the covariance of the power spectra from each simulation after averaging over $N_{\text{mocks}}$ catalogs in each simulation. The expected error on bandpower $P(k_i)$ is then

$$P(k_i) = \langle P_{sk}(k_i) \rangle$$

(4.17)

$$\sigma^2_{P(k_i)} = \frac{1}{N_{\text{sim}}}(C^{-1}_{\text{model,\text{reduced}}})_{ii}$$

(4.18)
Finally, we again neglect the covariance of the covariance matrix elements to estimate the errors on $C_{ij,\text{tot}}$, $C_{ij,\text{reduced}}$, and $C_{ij,HOD}$ from their variance over the simulations. These errors are only used for fitting the parameters $\alpha_{\text{shot}}$ and $\beta_{\text{beat}}$, so inaccuracies in the errors should make our estimates noisier, but not biased.

### 4.3 Simulations

We completed 42 TPM periodic box simulations with the parameters selected in Ch. 3: $L_{\text{box}} = 558 \, \text{Mpc}/h$ and $N_p = 512^3$, corresponding to a particle mass $M_p = 1.43 \times 10^{11} M_\odot$. As described in Ch. 3 we use the publicly available ic code (Sirko, 2005) to generate the initial conditions of the simulations. This code was designed to match real-space statistical properties such as the mass variance in spheres and the correlation function $\xi(r)$ by allowing the DC mode in the box to vary. This improvement is crucial to the analysis in this chapter, since we wish to extract accurate matter and halo power spectra as the density field enters the nonlinear regime.

We output the dark matter particle positions and velocities at time steps nearest to universe redshifts $z_{\text{NEAR}} = 0.235$, $z_{\text{MID}} = 0.342$, and $z_{\text{FAR}} = 0.421$. These redshifts are the galaxy-weighted mean redshifts of the NEAR ($0.155 < z < 0.300$), MID ($0.300 < z < 0.380$), and FAR ($0.380 < z < 0.474$) samples analyzed in Tegmark et al. (2006). These samples contain roughly equal numbers of galaxies, and their power spectra have roughly the same amplitude on large scales. Eqn. 3.1 shows that for a snapshot of the universe at fixed redshift, each simulation should be output at a different $z_{\text{box}}$ that increases with $\Delta_o$. The TPM code outputs particle data at the nearest whole time step to the desired redshift. Since the cosmological parameters for each simulation are slightly different in the ic formalism, the allowed output redshifts will vary slightly with $\Delta_o$. We correct for this effect exactly in the linear regime by scaling the power spectra from each simulation by $(D(z_{\text{MID}})/D(z_{\text{obs}}))^2$ before computing averaged quantities. Without this scaling, we get the same average dark
matter power spectra; changes in the covariance estimates are well below 1%. Any nonlinear corrections to this scaling will have an even smaller effect, and so we safely neglect them.

### 4.3.1 Calculating Power Spectra

Because of the low number density of LRGs, the shot noise correction is large compared to the $P(k)$ of the continuous density field sampled by the LRGs. We use an FFT with $N = 512^3$ points, and present the power spectrum out to $k = 0.4 \, h \, \text{Mpc}^{-1} = 0.14k_{\text{Nyquist}}$. Comparison with an $N = 1024^3$ grid showed agreement at the $\sim 10^{-6}$ level. Following the work of Hockney & Eastwood (1981), Jing (2005), and Baugh & Efstathiou (1994), mock LRGs are distributed on the FFT grid using the triangular-shaped cloud (TSC), and the power spectrum is estimated at each $\vec{k}$ by (Jing, 2005)

$$P_{\text{est}}(\vec{k}) = \frac{|\delta_{\text{LRG,FFT}}(\vec{k})|^2 - P_{\text{shot,TSC}}(\vec{k})}{W_{\text{TSC}}^2(\vec{k})}$$

$$P_{\text{shot,TSC}}(\vec{k}) = \frac{1}{\bar{n}} \prod_i \left[ 1 - \sin^2 \left( \frac{\pi k_i}{2k_N} \right) + \frac{2}{15} \sin^4 \left( \frac{\pi k_i}{2k_N} \right) \right]$$

$$W_{\text{TSC}}(\vec{k}) = \prod_i \left[ \text{sinc} \left( \frac{\pi k_i}{2k_N} \right) \right]^3$$

where $\delta_{\text{LRG,FFT}}(\vec{k})$ is the FFT of the LRG overdensity field, $k_i$ are the Cartesian components of $\vec{k}$, and $k_N = \pi / \Delta_{\text{grid}}$ is the Nyquist frequency. Eqn. 4.19 corrects for FFT aliasing in the limit that $P(\vec{k} + \vec{k}_N)$ is shot noise dominated, which is the case here. Finally, we average the values of Eqn. 4.19 over k-bands with $\Delta k_{\text{box}} = 0.0113$.

### 4.4 Mock Catalogs

We use the technique described in Ch. 2 to produce mock catalogs for the NEAR, MID, and FAR samples. We compute $N_{\text{CiC}}(n_{\text{sat}})$ from the SDSS LRG NEAR, MID, and FAR samples separately as in Ch. 2, but with a few minor changes. First, we include a boundary set of galaxies in the redshift direction, since FOGs may widely separate nearby pairs of LRGs;
we verified that this has a small effect on the resulting CiC group multiplicity. This buffer is $\Delta z = 0.007$ at the low redshift end of the sample and $\Delta z = 0.009$ at the high redshift; that is, equal to the maximum redshift separation for a CiC pair of galaxies at the sample boundary. This causes the redshift ranges of the subsamples for which we measure $N_{\text{CiC}}(n)$ to be slightly different than the Tegmark et al. (2006) NEAR, MID, and FAR samples. Because the color cuts used to select the LRG sample produces a complex radial selection function, the effective number density of the two samples is slightly different. As discussed in § 2.2.2, the main difference between our sample number densities in Column 4 of Table 4.2 and the result of integrating the Zehavi et al. (2005) model for the redshift dependence of $n_{\text{LRG}}(z)$ (Column 5) comes from our careful inclusion of objects from the imaging sample. In the MID sample, this is an 8% increase in the number density. We have verified that once the imaging galaxies are accounted for along with the difference in redshift range of the NEAR and FAR samples, our simple number density estimate $n_{\text{LRG}} = N_{\text{sample}} / V_{\text{sample}}$ is in agreement with the expectations from the Zehavi et al. (2005) model.

Second, our redshift indicator used for objects without spectra must be adapted as the 4000 Å break moves from the g to the r band at $z \approx 0.4$ (see Figure 4 of Eisenstein et al. (2001)). For our FAR sample we use the $r - i$ color as a redshift indicator. Table 4.1 shows the observed $N_{\text{CiC}}(n_{\text{sat}})$ for the NEAR, MID, and FAR subsamples. Since Tegmark et al. (2006) find that the NEAR, MID, and FAR subsamples have consistent $P(k)$’s, we vary $\sigma_{\log M}$ to match each redshift subsample to the large scale amplitude of the combined $P(k)$ reported in Tegmark et al. (2006). Table 4.2 gives the HOD parameters for each subsample, and Figure 4.6 shows the agreement of the large scale $P(k)$ with Tegmark et al. (2006) when we apply their FOG algorithm to our mock catalogs.

### 4.5 Reconstructing the Halo Density field

In the CiC technique detailed in Ch. 2, two galaxies are considered neighbors when their transverse comoving separation satisfies $\Delta r_{\perp} \leq 0.8 \, \text{Mpc}/h$ and their redshifts satisfy
Table 4.1: The final $N_{CiC} (n_{sat})$ group multiplicity function following the method in Ch. 2 for our NEAR, MID, and FAR LRG subsamples.

<table>
<thead>
<tr>
<th>$n_{sat}$</th>
<th>$N_{CiC,NEAR}(n)$</th>
<th>$N_{CiC,MID}(n)$</th>
<th>$N_{CiC,FAR}(n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>22921.71</td>
<td>24537.81</td>
<td>19109.71</td>
</tr>
<tr>
<td>1</td>
<td>1372.63</td>
<td>1301.29</td>
<td>664.28</td>
</tr>
<tr>
<td>2</td>
<td>170.01</td>
<td>153.40</td>
<td>61.94</td>
</tr>
<tr>
<td>3</td>
<td>41.85</td>
<td>25.59</td>
<td>6.08</td>
</tr>
<tr>
<td>4</td>
<td>15.16</td>
<td>9.07</td>
<td>2.04</td>
</tr>
<tr>
<td>5</td>
<td>2.11</td>
<td>2.04</td>
<td>1.00</td>
</tr>
<tr>
<td>6</td>
<td>1.01</td>
<td>1.01</td>
<td>0.00</td>
</tr>
<tr>
<td>7</td>
<td>1.02</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>8</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 4.2: Mock catalog parameters. Masses are for SO halos with $\Delta = 200$ in units of $10^{14} M_\odot$ and number densities are in units of $10^{-4} \,(Mpc/h)^{-3}$.

<table>
<thead>
<tr>
<th>Sample</th>
<th>$z_{SDSS , CiC}$</th>
<th>$\bar{z}_{, , , , \text{sim}}$</th>
<th>$\bar{n}_{, , , , \text{sim}}$</th>
<th>$\bar{n}_{, , , , \text{model}}$</th>
<th>$\sigma_{\log M}$</th>
<th>$M_{, , , , , \text{min}}$</th>
<th>$M_{, , , , , \text{cut}}$</th>
<th>$M_1$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEAR</td>
<td>0.162 - 0.300</td>
<td>0.2347</td>
<td>1.05</td>
<td>0.982</td>
<td>0.6</td>
<td>0.78</td>
<td>0.49</td>
<td>5.87</td>
<td>1.16</td>
</tr>
<tr>
<td>MID</td>
<td>0.300 - 0.380</td>
<td>0.3420</td>
<td>0.962</td>
<td>0.890</td>
<td>0.6</td>
<td>0.76</td>
<td>0.57</td>
<td>6.29</td>
<td>1.05</td>
</tr>
<tr>
<td>FAR</td>
<td>0.380 - 0.465</td>
<td>0.4216</td>
<td>0.470</td>
<td>0.422</td>
<td>0.9</td>
<td>2.27</td>
<td>1.38</td>
<td>5.97</td>
<td>0.78</td>
</tr>
</tbody>
</table>

$\Delta z/(1 + z) \leq \Delta v_p/c = 0.006$. A cylinder should be a good approximation to the density contours of satellites surrounding central galaxies in redshift space, as long as the satellite velocity is uncorrelated with its distance from the halo center and the relative velocity dominates the separation of central and satellite objects along the line of sight. Galaxies are then grouped with their neighbors by a FoF algorithm. The reconstructed halo density field is defined by the superposition of the centers of mass of each CiC group. The CiC parameters were established as a balance between completeness and contamination in identifying pairs of galaxies at an early stage of this work based on FoF halo catalogs. While the parameters and details of the method described here are sufficient for approximately recovering the halo density field power spectrum, the method could almost certainly be improved to more accurately recover groups of galaxies residing in the same dark matter halo.
4.6 Results

4.6.1 Dark Matter

Mass Function

Fig. 4.2 shows the ratio of volume-weighted SO mass function in our simulation set to the fitting function of Tinker et al. (2008) down to $M = 7.15 \times 10^{12} M_\odot$. There is good agreement at the level of accuracy claimed for their analytic fits ($\sim 5\%$). The overabundance of halos in the lowest mass bins is likely due to the small number of particles per halo in those bins. As discussed in Ch. 3, this slight modification of the mass function is unimportant for our purposes, since there are no satellites in these mass bins, and changes in the mass function are degenerate with changes to the probability of hosting a central galaxy as a function of halo mass, $N_{cen}(M)$.

Matter Power Spectrum

Fig. 4.3 shows the ratio of the matter power spectrum to the power spectrum of the simulation initial conditions scaled by the expected linear growth $D^2(z)$ for our three redshift samples; the results at different redshifts are highly covariant since they are measured from the same set of simulations at relatively small separations in time. Error bars are computed from the inverse of the model covariance matrix in Eqn. 4.10 where $\beta_{beat} = 1$ is fixed and the shot noise is negligible.

As expected the nonlinear correction grows as the redshift decreases. The nonlinear evolution generates power that smoothly increases with $k$ and also damps the baryon oscillations. The halofit nonlinear correction (Smith et al., 2003) (red curve, evaluated at $z_{MID}$) underestimates both the baryon wiggle suppression and the amplitude of the smooth increase in power; Crocce & Scoccimarro (2008) also find a disagreement between in their $N$-body simulation results and the halofit nonlinear correction.
Our fitting function to the nonlinear matter power spectrum $P_{DM}(k)$ aims to capture both the smearing of the acoustic peaks and a smooth increase in power with $k$. We follow Eisenstein et al. (2007b) in defining $P_{\text{smear}}(k)$ (Eqn. 4.5), but adopt a slightly different fitting function for the smooth correction. Our model is

$$P_{DM}(k) = P_{\text{smear}}(k; k_{BAO}) \left( a_0 + a_1 k + a_2 k^2 + a_3 k^3 \right). \quad (4.22)$$

For the MID sample with 35 bandpowers and 5 parameters, $\chi^2 = 21$. We find $k_{BAO} = 0.14$, in good agreement with the values reported in Eisenstein et al. (2007b). In Fig. 4.4 we show that the $P_{\text{smear}}(k)$ term accounts for the baryonic features in $P_{DM}(k)$, and the polynomial in $k$ adequately fits the smooth correction for the MID sample; the NEAR and FAR fits are similar. In Table 4.3 we list fits for the NEAR, MID, and FAR power spectra out to a maximum $k$ of 0.2 and 0.4 $h$/Mpc. When $k_{\text{max,fit}} = 0.2$ only the first three terms in the polynomial expansion are necessary for a good fit.

**Covariance Matrix**

Fig. 4.5 shows diagonal elements of the normalized covariance matrix $\langle \Delta P_i \Delta P_j \rangle / P_i P_j$. We estimate the errors from the diagonal variances (Eqn. 4.12); these may not capture the true errors since off-diagonal elements should be present in the 8-point function as well. Nevertheless, when we use these error estimates to compute $\chi^2$ for the model in Eqn. 4.10, we find $\chi^2 = 1600$ for 1296 degrees of freedom ($0 \leq k \leq 0.4$); if we restrict the covariance matrix to the 196 elements with both $k$ bands between 0.056 and 0.21, we find $\chi^2 = 262$. In this case there are no free parameters and we deem the model a reasonable fit. If we allow the amplitude of the beat coupling term to vary, we find a best fit value 0.96 for the full matrix ($\chi^2 = 1564$) and 0.84 for the 0.056 $\leq k \leq 0.21$ subsample ($\chi^2 = 218$). We note that for the DC mode realizations of our 42 simulations, the variance is a factor of 0.83 lower than the expected variance (in agreement with the expected random variation for a single mode, $N_{\text{sim}}^{-1/2} = 15\%$). We conclude that Eqn. 4.10 is an excellent fit to our dark matter
Figure 4.2: The ratio of the volume-weighted SO mass function from our simulations with the spline fit given in Tinker et al. (2008) for the NEAR (solid line), MID (dashed line), and FAR (dotted line) in bins of $\Delta \log_{10} M = 0.1$. The agreement is within the stated accuracy of the Tinker et al. (2008) spline fit ($\sim 5\%$).
Figure 4.3: The ratio of the volume-averaged matter power spectra at $z_{FAR} = 0.421$ (black), $z_{MID} = 0.342$ (blue), and $z_{NEAR} = 0.235$ (green) to the input power spectrum scaled by $D^2$. The nonlinear correction to the matter power spectrum measured from our simulations is larger than expected from halofit (Smith et al., 2003), shown in red and evaluated at $z_{MID}$. We also overlay $\log k P_{IC}(k)$ to indicate the location of the baryon wiggles in the initial power spectrum. The nonlinear evolution generates power that smoothly increases with $k$ and damps the baryon oscillations. Error bars are shown only for the MID sample for clarity and are estimated from the inverse of the model covariance matrix defined in Eqn. 4.10 where the shot noise is negligible and $\beta_{beat} = 1$; the beat-coupling term in Eqn. 4.10 correlates the bandpowers.
Figure 4.4: $P_{DM}(k)/P_{IC}(k)$ for the MID sample (cyan), $P_{DM}(k)/P_{\text{smeared}}(k)$ for $k_{\text{BAO}} = 0.14$ (blue), and our model fit $1.026 - 1.199k + 11.06k^2 - 8.426k^3$ (red). We also overlay log $kP_{IC}(k)$ to indicate the location of the baryon wiggles in the initial power spectrum. Using error bars derived from the inverse of the model covariance matrix defined in Eqn. 4.10, $\chi^2 = 21$ for 35 bandpowers and 5 parameters. Comparison of the cyan and blue curves shows that $P_{\text{smeared}}(k)$ adequately models the BAO features, and comparison of the blue and red curves demonstrates that our third-order polynomial in $k$ is sufficient for $k \leq 0.4 \, h/\text{Mpc}$. 
covariance matrix.

Another point of interest in Fig. 4.5 is that the inverse of the diagonal elements of the inverse covariance matrix are nearly equal to the Gaussian expectation (black curve). This means that while the beat-coupling does not introduce additional errors on the measurement of the bandpowers, the measured values will be covariant.

4.6.2 Mock Catalogs

Power Spectra

Fig. 4.6 shows the agreement between our NEAR, MID, and FAR power spectra and the measurements presented in Tegmark et al. (2006). We used this comparison to set by eye the large scale bias of our mock catalogs through $\sigma_{\log M}$, the width of $N_{\text{cen}}(M)$ in Eqn. 2.8. The small discrepancy between the normalizations of the FAR and other catalogs at small $k$ could be eliminated with a slight variation in $\sigma_{\log M,\text{FAR}}$. However, the FAR sample clearly has a different shape, and is consistent with trends in Fig. 6 of Tegmark et al. (2006), reproduced here in Fig. 4.7. Our FAR sample has $\sim 3\%$ more power at $k = 0.09$ and $\sim 10\%$ more power at $k = 0.2$ than the MID and FAR samples. Similar trends are evident in Fig. 4.7, though their significance is difficult to assess.

In Fig. 4.8 we fit the NEAR, MID, and FAR FOG-compressed samples to the nonlinear correction model of Tegmark et al. (2006) in Eqn. 4.6. This model accurately fits the nonlinear corrections for each of the subsamples. $P(k)$ computed over the entire sample as in Tegmark et al. (2006) will be a weighted average over these points. We first note that the nonlinear correction for the FAR sample is significantly larger than for the MID and NEAR. Much of this discrepancy can be attributed to the change in $\bar{n}$ in Eqn. 4.3: the size of FOG groups in the Tegmark et al. (2006) algorithm is allowed to increase by 30% as $\bar{n}$ drops, and this more aggressive compression adds more small scale power, even though
Figure 4.5: The blue curve shows the measured diagonal elements of the covariance matrix. Error bars are estimated from the variance of $\langle (\Delta P_i)^2 \rangle$ across our simulations. The red curve shows the Gaussian prediction for the diagonal covariance, the inverse of the number of complex modes in each linearly-spaced $k$-bin. The green curve shows the sum of the Gaussian term and the beat-coupling term in Hamilton et al. (2006), which provides a good model for the full covariance matrix. In black we show the inverse of the diagonal elements of the inverse covariance matrix. We find that while the errors on the bandpowers are very close to the Gaussian expectation, the beat-coupling terms in the covariance matrix mean the bandpowers are correlated.
we have shown that there are fewer satellite galaxies in the FAR sample (see Table 4.1), so that the one-halo power should actually be smaller. The two cyan curves show the best fit vanilla model of Tegmark et al. (2006). In both curves we set $Q_{NL} = 31$; the upper curve has $A = 1.4$ and the lower curve has $A = 1.7$ in Eqn. 4.4; both values were given in different places in Tegmark et al. (2006), so we were unsure of which to adopt. In both cases, the shape of the model fit is consistent with an average of our NEAR, MID, and FAR samples. The black curve shows the ratio of their best fit linear model with $P_{IC}(k)$ from our simulations. These two models differ at a level larger than the statistical error bars shown in Fig. 4.1 even if we restrict the analysis to $k < 0.1$ and certainly if we can use the measurements between $k = 0.1$ and $k = 0.2$. However, the $Q_{NL}$ parameter is degenerate with this difference in power spectra, so that with this nonlinear correction model we cannot distinguish the two linear spectra.

In Figs. 4.9 through 4.11 we attempt to isolate the sources of nonlinearity in the LRG power spectrum, and demonstrate that the power spectrum of the reconstructed halo density field is the best tracer of the underlying linear spectrum. In each of these figures we show the real space and redshift space power spectra of the central mock LRGs only (red) and the central and satellite LRGs (blue). For the MID and FAR subsamples, there is no detectable deviation from a constant bias between the central galaxies and the dark matter in real space for $k \leq 0.1$, and the deviation at $k = 0.2$ is $\leq 2\%$; the discrepancy for the NEAR sample is $\sim 1\%$ for the NEAR sample at $k = 0.1$ and $5\%$ at $k = 0.2$. Therefore the halo field traced by LRGs is linearly biased in real space to a very good approximation. Including the satellite galaxies in real space introduces a scale-dependent bias marginally evident even below $k = 0.1$; the effect is mitigated in the monopole redshift space power spectrum by the FOGs. The FOG-compressed catalog using the algorithm of Tegmark et al. (2006) has more nonlinear power than the LRGs in real space; this is because the algorithm compresses many pairs of galaxies in distinct dark matter halos because the LHS of Eqn. 4.3 is larger than the size of a typical LRG host halo. In black we show the power
spectrum of the reconstructed halo density field. In all three subsamples it is very similar to the redshift space power spectrum of the central LRGs, as one would expect if the reconstruction is sufficiently accurate. In the redshift space power spectra there is evidence for further suppression of the BAOs; the ratio of $P_{LRG}(k)/P_{DM}(k)$ shown here does not account for this since $P_{DM}(k)$ is measured in real space from our simulations. However, Eisenstein et al. (2007b) have accurately modeled the BAO suppression in redshift space, and the effect can be simply included by decreasing $k_{BAO}$ in Eqn. 4.5.

For $k \leq 0.1$ the reconstructed halo density field does not detectably deviate from the shape of the underlying matter power spectrum for the NEAR and MID samples; for the FAR sample, the deviation is $\sim 1\%$. In contrast, the FOG-compressed sample deviates from a constant bias at the 6%, 7%, and 10% level between $k = 0.05$ and $k = 0.1$. The reconstructed halo density field has only small deviations from a scale-independent bias out to $k = 0.2$: the deviation is 4%, 2.8%, and 2.5% for the NEAR, MID, and FAR sample. The systematics out to $k = 0.2$ should therefore be small enough for use in cosmological analyses. In contrast, the FOG-compressed spectra deviate from a constant bias at the 19%, 20% and 30% levels for the NEAR, MID, and FAR samples; this is a factor of $\sim 5$ larger than the statistical errors on Tegmark et al. (2006)’s bandpowers between $k = 0.1$ and $k = 0.2$. The nonlinear correction must therefore be extremely well-calibrated to extract any cosmological information in this regime.

We have shown that the nonlinear correction between the reconstructed halo density field and the matter power spectrum is very small. The small corrections we have found are well fit away from $k = 0$ by a polynomial. We list best fit polynomial parameters in Table 4.3 in the Appendix. The amplitude of this correction in our fiducial cosmology should provide a conservative estimate of the uncertainty in the correction as the cosmology is varied within the space allowed by the latest WMAP analysis Komatsu et al. (2008): $\sim 1\%$ for $k \leq 0.1$ and $\sim 4\%$ between $k = 0.1$ and $k = 0.2$. Finally, to recover the linear power spectrum,
one must know the amplitude of the BAO suppression and the degree of nonlinearity in the matter power spectrum as a function of the cosmological parameters. Several groups are addressing this issue (e.g., Eisenstein et al. (2007b) and Habib et al. (2007)).

Covariance Matrices

The model that we fit for the mock catalog covariance matrices is

\[ B_{ij} = 4R_a P(k_i)P(k_j)\delta_{DC}^2(z) \] (4.23)

\[ C_{ij,\text{tot}} = \frac{1}{N_i} \left( P(k_i)P(k_j) + \frac{\alpha_{\text{tot}}}{n} \right)^2 \delta_{ij}^K + \beta_{\text{tot}} B_{ij} \] (4.24)

\[ C_{ij,\text{reduced}} = \frac{1}{N_i} \left( P^2(k_i) + \frac{\alpha_{\text{red}} P(k_i)}{n} + \frac{1}{n^2 N_{\text{mocks}}} \right) \delta_{ij}^K + \beta_{\text{red}} B_{ij} \] (4.25)

\[ C_{ij,HOD} = \frac{1}{N_i} \left( \frac{\alpha_{\text{HOD}} P(k_i)}{n} + \frac{N_{\text{mocks}}}{n^2 (N_{\text{mocks}} - 1)} \right) \delta_{ij}^K + \beta_{\text{HOD}} B_{ij} \] (4.26)

Note we use the measured \( P(k_i) \) rather than the linear or model \( P(k_i) \) in Eqns. 4.23-4.26. In the reduced covariance we expect the Poisson noise to be suppressed by \( 1/N_{\text{mocks}} \), the number of mock catalogs we produce for each TPM simulation. Eqn. 4.16 implies that \( \alpha_{\text{HOD}} + \alpha_{\text{red}} = 2\alpha_{\text{tot}} \) for \( \alpha_{\text{tot}} \approx 1 \) and \( \beta_{\text{HOD}} + \beta_{\text{red}} = \beta_{\text{tot}} \). Best fit values for each mock catalog type and each redshift subsample are listed in Table 4.4 where the fit included all modes with \( k \leq 0.4 \) and in Table 4.5 where the fit included all modes between \( k = 0.05 \) and \( k = 0.2 \). We present both in order to assess whether the non-Gaussian terms are growing with \( k \). In Fig. 4.12 we show the measured and model \( C_{ij} \) values along the diagonal and three rows in the matrix for the MID reconstructed halo density field covariance matrix; others are similar.

First, \( \alpha_{\text{tot}} \approx 1 \) for all mocks and redshift subsamples, indicating that the standard shot noise contribution to the covariance is approximately correct for our mock catalogs. However, the relative distribution of this cross term between \( C_{ij,HOD} \) and \( C_{ij,\text{reduced}} \) varies; \( \alpha_{\text{HOD}} \) is large for the FAR sample where the number density is \( \sim \) half that in the MID and NEAR sample, and is also larger in the samples including satellites relative to the
Figure 4.6: $kP_s(k)$ for the NEAR (blue), MID (green), and FAR (red) mock galaxy samples with the FOG compression algorithm of Tegmark et al. (2006) applied compared with their observed power spectrum (black points with errors). Our NEAR and MID samples are virtually identical, so we show error bars only for the MID and FAR samples. Our error bars are derived from the diagonal elements of the inverse $C_{ij,\text{red}}$ matrix. We scale the real space $P(k)$ of Tegmark et al. (2006) by $(0.8)^{-1}$ to approximate the redshift space monopole power spectrum; this is the relation between the real and redshift monopole spectra if one neglects the small contributions from the quadrupole and hexadecapole detailed in their §A3. For comparison, the jagged cyan curve shows our simulation initial conditions drawn from the convolved power spectrum according to the $ic$ algorithm, and the smooth cyan curve shows the linear power spectrum for the cosmological parameters adopted in this study.
Figure 4.7: Figure 6 of Tegmark et al. (2006) showing the power spectra of the NEAR, MID, and FAR subsamples along with the MAIN galaxies. Our mock FAR sample in Fig. 4.6 also has slightly more power at small scales compared to the MID and NEAR subsamples.
Figure 4.8: We separately fit our NEAR, MID, and FAR FOG-compressed samples to the nonlinear correction model of Tegmark et al. (2006) in Eqn. 4.6. Best fit parameters are $b^2_{\text{NEAR}} = 5.61$, $Q_{\text{NL,NEAR}} = 17.4$; $b^2_{\text{MID}} = 6.15$, $Q_{\text{NL,MID}} = 18.4$; and $b^2_{\text{FAR}} = 6.67$, $Q_{\text{NL,FAR}} = 23.3$. Each subsample is scaled by the input power spectrum and the best fit $b^2$ value to accentuate the variation in the power spectrum shapes for these three samples. The best fit vanilla model in Tegmark et al. (2006) is shown by two cyan curves, where we have adjusted the normalization to agree at $k \sim 0.05$. The top curve has $A = 1.4$ in Eqn. 4.4 and the bottom curve has $A = 1.7$, and both have $Q_{\text{NL}} = 31$. The black curve shows the ratio of the best fit linear power spectrum of Tegmark et al. (2006) and the linear power spectrum adopted in this work, demonstrating that $Q_{\text{NL}}$ is degenerate with changes in the spectral shape, so that cosmological parameter information is lost to $Q_{\text{NL}}$. 
Figure 4.9: The ratio of the LRG to nonlinear dark matter spectrum for the MID sample. The lower red curve shows only central galaxies with the power spectrum computed in real space. The upper red curve is for the same galaxy sample but measured in redshift space. The lower and upper blue curves are for the sample of both central and satellite galaxies in real and redshift space. The green curve shows the redshift space power spectrum of the central and satellite galaxies after undergoing Tegmark et al. (2006) FOG compression. The black curve is the power spectrum of the CiC groups in redshift space. The nonlinear correction associated with this sample is much smaller than the FOG-compressed sample.
Figure 4.10: Same as Figure 4.9, but for the FAR sample.
Figure 4.11: Same as Figure 4.9, but for the NEAR sample.
samples with only central objects. All three matrices have significant off-diagonal terms. \( \beta_{\text{tot}} \) is larger than one for all samples, which may be expected since the LRGs are biased tracers, so the DC mode variance will increase by the factor \( b_{\text{LRG}}^2 \). However, the amplitude does not scale with \( b^2 \), but is much larger for the FAR sample. The best fit parameters do not change between real and redshift space for the same sample, but \( \beta_{\text{tot}} \) is larger for the samples including satellites. The reconstructed halo density field best fit parameters are consistent with the parameters of the sample of central galaxies, and the FOG-compressed sample is consistent with the sample of central and satellite galaxies. \( \beta_{\text{tot}} \) decreases slightly when fitting only to the \( k = 0.05 \) to \( k = 0.2 \) subsample, but the effect is small and so we adopt the parameters reported in Table 4.5 to carry out our covariance matrix calculations for the error bars on \( P_{\text{LRG}}(k) \).

There is a modest decrease in the normalized diagonal covariance \( \Delta P^2(k_i)/P^2(k_i) \) (~5%) when satellite galaxies are included. However, this decrease is offset by the fact that the \( P^2(k_i) \) has a larger nonlinear component, so not all of the available information will be about the linear component; furthermore, the large nonlinear correction introduces uncertainty in extracting the linear component. Moreover, \( \beta_{\text{tot}} \) is larger for samples which include satellite galaxies, so the bandpowers will be more highly correlated for those mocks. We therefore conclude that the reduction in large scale bias by eliminating the satellite galaxies and reconstructing the halo density field is offset in the error budget by smaller off-diagonal covariance and smaller uncertainties in the nonlinear correction.
Figure 4.12: We plot the normalized covariance matrix elements $C_{ij}/P(k_i)P(k_j)$ scaled by $\sqrt{N_i N_j}$, the number modes in bins $i$ and $j$. The solid lines show our measurements and the dashed lines show the model fits from Table 4.4. We plot $C_{ij,\text{tot}}$ (black), $C_{ij,\text{reduced}}$ (blue), and $C_{ij,\text{HOD}}$ (red). The top left panel shows the diagonal elements, and the other panels show cross sections of the covariance matrix with $k_i = 0.09, 0.15, \text{ and } 0.3 \, h/\text{Mpc}$. 
The Redshift Space Quadrupole

In linear theory the redshift space power spectrum can be decomposed into a monopole, quadrupole, and hexadecapole:

\[
\beta = \frac{1}{b_{gall}} \frac{d \ln D}{d \ln a} \quad (4.27)
\]

\[
P_s(\vec{k}) = P(k) \left[ (1 + \frac{2}{3} \beta + \frac{1}{5} \beta^2) L_0(\mu_{\vec{k}}) + \left( \frac{4}{3} \beta + \frac{4}{7} \beta^2 \right) L_2(\mu_{\vec{k}}) + \left( \frac{8}{35} \beta^2 \right) L_4(\mu_{\vec{k}}) \right] \quad (4.28)
\]

where \( L_i \) is the Legendre polynomial of order \( i \). The method of Tegmark et al. (2006) assumes Eqn. 4.28 with \( \beta \) independent of \( k \) to obtain the best estimate of the real space power spectrum from \( P_{gg} \), \( P_{gv} \), and \( P_{vv} \). In Fig. 4.13 we examine the structure of the redshift space distortions in our mock catalogs. The first three panels show the quadrupole to monopole ratio for central LRGs (black), the reconstructed halo density field (blue), and the FOG-compressed density field (green). These are all similar and show a modest increase in the quadrupole to monopole ratio with \( k \). This demonstrates that the FOG compression technique of Tegmark et al. (2006) successfully removes the effects of the FOGs induced by satellite galaxies. The oscillations probably result from the extra suppression of BAO features in redshift space (Eisenstein et al., 2007b). The central and satellite LRG sample (red) quadrupole to monopole ratio lies below the expected value at all \( k \) values, indicating that the FOG suppression of power is evident even in the linear regime. Therefore a sample where the effects of satellite FOGs have not been removed by either halo density field reconstruction or FOG compression will provide a biased estimate of \( \beta \).

The bottom right panel shows the ratio of the redshift space monopole to the real space power spectrum for the sample of central galaxies only (black) and central and satellite galaxies (red). The suppression of power by the satellite FOGs is also evident here. For central LRGs only, the ratio is consistent with the expected value from Eqn. 4.28, and between \( k = 0 \) and \( k = 0.2 \) this ratio varies by only \( \sim 3\% \). While there is a clear scale dependence in the quadrupole to monopole ratio for the central LRG sample, the large bias of LRGs means that the monopole redshift space power spectrum is nearly insensi-
tive to the redshift space nonlinearities in the halo density field. The cyan curve shows $P_{s,sat}/P_{real,cen} \times b_{cen}^2/b_{sat}^2$. The increase in power in real space when the satellites are included is larger than the suppression of power in redshift space by their FOGs, so that the satellite monopole spectrum has more power at high $k$ than the real space central LRGs.

### 4.7 Conclusions

This chapter introduced and tested our algorithm for using the halo density field to measure the power spectrum. We found the nonlinear correction between this field and the underlying matter density field to be both smaller and more robust to the satellite fraction variation in the sample than both the FOG-compressed density field used in the analysis of Tegmark et al. (2006) and the redshift space monopole power spectrum used in the analysis of Percival et al. (2007).

The parameters of our simulation set were selected in Ch. 3 to provide accurate two point halo density and velocity statistics to assure accurate representations of FOG features in our mock catalogs. To our knowledge this is the first attempt to study the detailed effects of FOG treatment on the LRG power spectrum. We find that their treatment can affect bandpowers even in the linear regime $k \leq 0.1$. We first examined the nonlinear matter power spectrum of our 42 simulations. The ratio of the nonlinear to input matter power spectra is well described by a smearing of the BAOs as modeled by Eisenstein et al. (2007b) and a smooth increase in power with $k$ that can be fit by a third order polynomial out to $k = 0.4 \ h/\text{Mpc}$ or second order polynomial out to $k = 0.2 \ h/\text{Mpc}$. We detect a substantial deviation from the predictions of halofit (Smith et al., 2003).

Figure 4.9 demonstrates the main point of this work: satellite galaxies systematically alter the shape of the power spectrum even at $k < 0.1 \ h/\text{Mpc}$. Extraction of cosmological information from the broadband shape of the power spectrum is already limited by system-
Figure 4.13: The bottom right panel shows the ratio of the redshift space monopole to the real space power spectrum for the MID sample. The straight solid lines show the prediction of Eqn. 4.28 for both samples. We fit the bias in Eqn. 4.27 to the large scale ratio $P_{LRG}/P_{DM}$. The black curve is for the sample of central galaxies only ($b_{cen} = 2.11$) and the red includes both central and satellite galaxies ($b_{sat} = 2.23$). The cyan curve shows the ratio of the redshift space monopole of the central and satellite galaxies to the real space power spectrum of the central objects (scaled by $b_{cen}^2/b_{sat}^2$). The other three panels show the quadrupole to monopole ratio in redshift space. Error bars are only shown for the central galaxy sample (black) for clarity. The central and satellite sample is shown in red, the reconstructed halo density field in blue, and the FOG-compressed sample in green. The expected values from Eqn. 4.28 are shown as the solid lines for the central (black) and central and satellite (red) samples.
atics (Sanchez & Cole, 2007) which we (and others) suggest can be attributed primarily to differences in the satellite contribution to the power spectrum. In this paper we demonstrate that while the FOG compression scheme in Tegmark et al. (2006) exacerbates these issues and requires a large nonlinear correction, the FOG features in the density field can be used to reconstruct the halo density field with high fidelity. The power spectrum of this field deviates from the dark matter power spectrum at the $\leq 4\%$ level for $k \leq 0.2 \ h/Mpc$ and $\leq 1\%$ level for $k \leq 0.1 \ h/Mpc$, where cosmological analyses usually restrict themselves. Moreover, we have shown that this correction changes only slightly between the NEAR, MID, and FAR LRG reconstructed halo density field while the FOG compressed mocks have a much larger variation between samples. Therefore, we can hope to push cosmological analyses to larger $k$ using the reconstructed halo density field as a tracer of the underlying matter density field fluctuations, particularly for galaxy samples like the LRGs which are spread over a large redshift range and are not volume-limited and thus have substantial variation in the satellite contribution to the power spectrum with redshift.

While we have not addressed the variation of the nonlinear correction to the reconstructed halo density field as a function of cosmological parameters, we have designed the form of our correction to minimize the variation with cosmology. Other researchers (Habib et al., 2007) are studying the dark matter power spectrum as a function of cosmology. We expect that our nonlinear correction $P_{LRG}/P_{DM}$ will remain small (of order 4% below $k = 0.2 \ h/Mpc$) as the cosmology is varied, so the variation of this small correction should be even smaller. Therefore, instead of introducing a nuisance parameter for the nonlinear correction, we propose that the amplitude of the correction should be taken as the error on its value in cosmological parameter analyses, or relatively strong priors on the amplitude of the correction to the reconstructed halo density field be introduced.

In this work we have also investigated the properties of the covariance matrix of the dark matter power spectrum as well as the mock galaxy catalogs divided into 6 different
samples: central galaxies in real and redshift space, central and satellite galaxies in real and redshift space, and our reconstructed halo density field in redshift space, and the Tegmark et al. (2006) FOG compressed galaxy density field in redshift space. All of these samples were well-modeled by a diagonal matrix with the usual Gaussian and Poisson shot noise terms plus the beat-coupling term presented in Hamilton et al. (2006). We expect that this will be a useful model for fitting the survey covariance matrix, where $P(k_b)$ is replaced by $P(k_{\text{survey}})$ with $k_{\text{survey}}$ determined by some effective survey size.

Finally we examined the structure of the redshift space distortions as a function of $k$ using the quadrupole. Both the reconstructed halo density field and the FOG-compressed mock LRG samples reproduce the modest $k$ dependence of the halo density field quadrupole to monopole ratio. Since the LRGs are so highly biased, this scale dependence causes a $\lesssim 3\%$ deviation in the redshift space monopole to real space power spectrum ratio out to $k = 0.2$. When satellites are included in the sample without FOG compression, the quadrupole to monopole ratio is lower than the linear value for the entire $k$ range accessible in our simulations and will significantly bias the estimate of $\beta$.

4.8 Appendix: Power Spectra and Covariance Matrix Best Fit Parameters
Table 4.3: Fits to nonlinear power spectra. $k_{max;fit}$ and $k_{BAO}$ are in units of $h$/Mpc. For $P_{DM}(k)/P_{IC}(k)$ we fit to Eqn. 4.22. When $k_{max;fit} = 0.2$ we hold $a_3 = 0$. Some of the redshift space mock catalog fits are not well-behaved at very small $k$’s, and the corrections to a constant should be suppressed at $k \lesssim 0.05$. This also makes $a_0$ deviate from the large scale bias, given in the last column and fit using only $k < 0.06$. $P_{sat,z}(k)$ denotes the power spectrum of a catalog containing central and satellite galaxies in redshift space, $P_{CIC,z}(k)$ denotes the power spectrum of the reconstructed halo density field in redshift space, and $P_{teg,z}(k)$ denotes an FOG compressed following Tegmark et al. (2006).
<table>
<thead>
<tr>
<th>Catalog</th>
<th>$\alpha_{HOD}$</th>
<th>$\alpha_{red}$</th>
<th>$2\alpha_{tot}$</th>
<th>$\beta_{HOD}$</th>
<th>$\beta_{red}$</th>
<th>$\beta_{tot}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEAR cen, real</td>
<td>1.07</td>
<td>1.10</td>
<td>2.00</td>
<td>0.26</td>
<td>0.94</td>
<td>1.26</td>
</tr>
<tr>
<td>NEAR sat, real</td>
<td>1.30</td>
<td>0.89</td>
<td>1.98</td>
<td>0.36</td>
<td>1.58</td>
<td>1.97</td>
</tr>
<tr>
<td>NEAR cen, redshift</td>
<td>1.11</td>
<td>0.99</td>
<td>2.02</td>
<td>0.26</td>
<td>0.95</td>
<td>1.29</td>
</tr>
<tr>
<td>NEAR sat, redshift</td>
<td>1.27</td>
<td>0.63</td>
<td>2.01</td>
<td>0.29</td>
<td>1.28</td>
<td>1.64</td>
</tr>
<tr>
<td>NEAR CiC, redshift</td>
<td>1.11</td>
<td>0.92</td>
<td>2.02</td>
<td>0.26</td>
<td>0.94</td>
<td>1.30</td>
</tr>
<tr>
<td>NEAR Teg, redshift</td>
<td>1.39</td>
<td>0.74</td>
<td>1.98</td>
<td>0.35</td>
<td>1.60</td>
<td>1.99</td>
</tr>
<tr>
<td>MID cen, real</td>
<td>1.14</td>
<td>1.0</td>
<td>2.00</td>
<td>0.34</td>
<td>1.21</td>
<td>1.62</td>
</tr>
<tr>
<td>MID sat, real</td>
<td>1.32</td>
<td>0.77</td>
<td>1.99</td>
<td>0.42</td>
<td>1.71</td>
<td>2.20</td>
</tr>
<tr>
<td>MID cen, redshift</td>
<td>1.17</td>
<td>0.93</td>
<td>2.01</td>
<td>0.34</td>
<td>1.19</td>
<td>1.63</td>
</tr>
<tr>
<td>MID sat, redshift</td>
<td>1.30</td>
<td>0.61</td>
<td>2.00</td>
<td>0.36</td>
<td>1.47</td>
<td>1.92</td>
</tr>
<tr>
<td>MID CiC, redshift</td>
<td>1.16</td>
<td>0.86</td>
<td>2.01</td>
<td>0.34</td>
<td>1.18</td>
<td>1.63</td>
</tr>
<tr>
<td>MID Teg, redshift</td>
<td>1.41</td>
<td>0.58</td>
<td>1.97</td>
<td>0.41</td>
<td>1.73</td>
<td>2.20</td>
</tr>
<tr>
<td>FAR cen, real</td>
<td>1.65</td>
<td>0.50</td>
<td>2.05</td>
<td>1.32</td>
<td>1.62</td>
<td>3.13</td>
</tr>
<tr>
<td>FAR sat, real</td>
<td>1.85</td>
<td>0.38</td>
<td>2.07</td>
<td>1.71</td>
<td>2.67</td>
<td>4.50</td>
</tr>
<tr>
<td>FAR cen, redshift</td>
<td>1.66</td>
<td>0.42</td>
<td>2.05</td>
<td>1.25</td>
<td>1.64</td>
<td>3.13</td>
</tr>
<tr>
<td>FAR sat, redshift</td>
<td>1.79</td>
<td>0.16</td>
<td>2.07</td>
<td>1.37</td>
<td>2.08</td>
<td>3.70</td>
</tr>
<tr>
<td>FAR CiC, redshift</td>
<td>1.66</td>
<td>0.37</td>
<td>2.06</td>
<td>1.25</td>
<td>1.69</td>
<td>3.21</td>
</tr>
<tr>
<td>FAR Teg, redshift</td>
<td>1.90</td>
<td>0.13</td>
<td>2.06</td>
<td>1.54</td>
<td>2.53</td>
<td>4.19</td>
</tr>
</tbody>
</table>

Table 4.4: Fits for the mock catalog covariance matrices using Eqns. 4.23 - 4.26 using all bandpowers between $k = 0.01$ and $k = 0.4$. ‘cen, real/redshift’ denotes a sample of only central LRGs in real/redshift space; ‘sat, real/redshift’ denotes a sample with both central and satellite galaxies in real/redshift space; ‘CiC, redshift’ denotes the reconstructed halo density field using the CiC method, and ‘Teg, redshift’ denotes the FOG-compressed sample according to Tegmark et al. (2006).
<table>
<thead>
<tr>
<th>Catalog</th>
<th>$\alpha_{HOD}$</th>
<th>$\alpha_{red}$</th>
<th>$2\alpha_{tot}$</th>
<th>$\beta_{HOD}$</th>
<th>$\beta_{red}$</th>
<th>$\beta_{tot}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEAR cen, real</td>
<td>1.05</td>
<td>0.85</td>
<td>1.96</td>
<td>0.23</td>
<td>0.86</td>
<td>1.14</td>
</tr>
<tr>
<td>NEAR sat, real</td>
<td>1.22</td>
<td>0.62</td>
<td>1.94</td>
<td>0.29</td>
<td>1.43</td>
<td>1.76</td>
</tr>
<tr>
<td>NEAR cen, redshift</td>
<td>1.07</td>
<td>0.71</td>
<td>1.90</td>
<td>0.23</td>
<td>0.94</td>
<td>1.23</td>
</tr>
<tr>
<td>NEAR sat, redshift</td>
<td>1.22</td>
<td>0.51</td>
<td>1.86</td>
<td>0.26</td>
<td>1.33</td>
<td>1.66</td>
</tr>
<tr>
<td>NEAR CiC, redshift</td>
<td>1.07</td>
<td>0.68</td>
<td>1.90</td>
<td>0.24</td>
<td>0.97</td>
<td>1.27</td>
</tr>
<tr>
<td>NEAR Teg, redshift</td>
<td>1.28</td>
<td>0.45</td>
<td>1.86</td>
<td>0.29</td>
<td>1.47</td>
<td>1.82</td>
</tr>
<tr>
<td>MID cen, real</td>
<td>1.08</td>
<td>0.84</td>
<td>1.96</td>
<td>0.28</td>
<td>1.14</td>
<td>1.49</td>
</tr>
<tr>
<td>MID sat, real</td>
<td>1.21</td>
<td>0.62</td>
<td>1.93</td>
<td>0.35</td>
<td>1.60</td>
<td>2.02</td>
</tr>
<tr>
<td>MID cen, redshift</td>
<td>1.11</td>
<td>0.74</td>
<td>1.93</td>
<td>0.30</td>
<td>1.20</td>
<td>1.58</td>
</tr>
<tr>
<td>MID sat, redshift</td>
<td>1.23</td>
<td>0.54</td>
<td>1.88</td>
<td>0.33</td>
<td>1.55</td>
<td>1.96</td>
</tr>
<tr>
<td>MID CiC, redshift</td>
<td>1.10</td>
<td>0.72</td>
<td>1.91</td>
<td>0.29</td>
<td>1.24</td>
<td>1.61</td>
</tr>
<tr>
<td>MID Teg, redshift</td>
<td>1.29</td>
<td>0.46</td>
<td>1.88</td>
<td>0.36</td>
<td>1.66</td>
<td>2.10</td>
</tr>
<tr>
<td>FAR cen, real</td>
<td>1.57</td>
<td>0.36</td>
<td>2.00</td>
<td>1.08</td>
<td>1.66</td>
<td>2.87</td>
</tr>
<tr>
<td>FAR sat, real</td>
<td>1.68</td>
<td>0.21</td>
<td>2.00</td>
<td>1.35</td>
<td>2.51</td>
<td>3.98</td>
</tr>
<tr>
<td>FAR cen, redshift</td>
<td>1.57</td>
<td>0.26</td>
<td>1.99</td>
<td>1.06</td>
<td>1.70</td>
<td>2.94</td>
</tr>
<tr>
<td>FAR sat, redshift</td>
<td>1.67</td>
<td>0.13</td>
<td>1.98</td>
<td>1.20</td>
<td>2.28</td>
<td>3.65</td>
</tr>
<tr>
<td>FAR CiC, redshift</td>
<td>1.58</td>
<td>0.26</td>
<td>2.00</td>
<td>1.06</td>
<td>1.84</td>
<td>3.09</td>
</tr>
<tr>
<td>FAR Teg, redshift</td>
<td>1.73</td>
<td>0.06</td>
<td>1.99</td>
<td>1.31</td>
<td>2.47</td>
<td>3.93</td>
</tr>
</tbody>
</table>

Table 4.5: Fits for the mock catalog covariance matrices using Eqns. 4.23 - 4.26 using all bandpowers between $k = 0.05$ and $k = 0.2$. Samples are the same as in Table 4.4.
References


Schneider, D. P., Scranton, R., Seljak, U., Sheldon, E., Shimasaku, K., Silvestri, N. M.,
Smith, J. A., Smolčić, V., Snedden, S. A., Stebbins, A., Stoughton, C., Strauss, M. A.,
SubbaRao, M., Suto, Y., Szalay, A. S., Szapudi, I., Szkody, P., Tegmark, M., Thakar,


Blanton, M. R., Lin, H., Lupton, R. H., Maley, F. M., Young, N., Zehavi, I., & Loveday, J.
2003, AJ, 125, 2276

Blanton, M. R., Schlegel, D. J., Strauss, M. A., Brinkmann, J., Finkbeiner, D., Fukugita,
M., Gunn, J. E., Hogg, D. W., Ivezić, Ž., Knapp, G. R., Lupton, R. H., Munn, J. A.,


Bland-Hawthorn, J., Bridges, T., Cannon, R., Colless, M., Collins, C., Couch, W., Cross,
N. J. G., Dalton, G., Eke, V. R., De Propris, R., Driver, S. P., Efstathiou, G., Ellis, R. S.,
Glazebrook, K., Jackson, C., Jenkins, A., Lahav, O., Lewis, I., Lumsden, S., Maddox, S.,


Dodelson, S. 2003, Modern Cosmology (San Diego, California: Elsevier)


Hamann, J., Hannestad, S., Melchiorri, A., & Wong, Y. Y. Y. 2008, Journal of Cosmology and Astro-Particle Physics, 7, 17


Huff, E., Schulz, A. E., White, M., Schlegel, D. J., & Warren, M. S. 2007, Astroparticle Physics, 26, 351


Matsumoto, M., & Nishimura, T. 1998, ACM Transactions on Modeling and Computer Simulation, 8, 3


Schulz, A. E., & White, M. 2006, Astroparticle Physics, 25, 172


Verde, L., & Peiris, H. 2008, Journal of Cosmology and Astro-Particle Physics, 7, 9


Wright, E. L., Meyer, S. S., Bennett, C. L., Boggess, N. W., Cheng, E. S., Hauser, M. G.,
Kogut, A., Lineweaver, C., Mather, J. C., Smoot, G. F., Weiss, R., Gulkis, S., Hinshaw,
G., Janssen, M., Kelsall, T., Lubin, P. M., Moseley, S. H., Murdock, T. L., Shafer, R. A.,

Yoo, J., Weinberg, D., et al prep

Bakken, J. A., Barkhouser, R., Bastian, S., Berman, E., Boroski, W. N., Bracker, S.,
Briegel, C., Briggs, J. W., Brinkmann, J., Brunner, R., Burles, S., Carey, L., Carr,
Csabai, I., Czarapata, P. C., Davis, J. E., Doi, M., Dombeck, T., Eisenstein, D., Ellman,
N., Elms, B. R., Evans, M. L., Fan, X., Federwitz, G. R., Fiscelli, L., Friedman, S.,
Frieman, J. A., Fukugita, M., Gillespie, B., Gunn, J. E., Gurbani, V. K., de Haas, E.,
Haldeman, M., Harris, F. H., Hayes, J., Heckman, T. M., Hennessy, G. S., Hindsley,
R. B., Holm, S., Holmgren, D. J., Huang, C.-h., Hull, C., Husby, D., Ichikawa, S.-I.,
Ichikawa, T., Ivezić, Ž., Kent, S., Kim, R. S. J., Kinney, E., Klaene, M., Kleinman,
A. N., Kleinman, S., Knapp, G. R., Korienek, J., Kron, R. G., Kunszt, P. Z., Lamb,
D. Q., Lee, B., Leger, R. F., Limmongkol, S., Lindenmeyer, C., Long, D. C., Loomis,
C., Loveday, J., Lucinio, R., Lupton, R. H., MacKinnon, B., Mannoni, E. J., Mantsch,
P. M., Margon, B., McGeehee, P., McKay, T. A., Meiksin, A., Merelli, A., Monet, D. G.,
Nichol, R. C., Nicinski, T., Nonino, M., Okada, N., Okamura, S., Ostriker, J. P., Owen,
R., Pauls, A. G., Peoples, J., Peterson, R. L., Petravick, D., Pier, J. R., Pope, A.,
Pordes, R., Prosapio, A., Rechenmacher, R., Quinn, T. R., Richards, G. T., Richmond,
M. W., Rivetta, C. H., Rockosi, C. M., Ruthmansdorfer, K., Sandford, D., Schlegel,

