Constraining galaxy-halo connection and cosmology using DESI data

by

Hanyu Zhang

B.S., Nanjing University, China, 2015

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Abstract

The Dark Energy Spectroscopic Instrument (DESI) survey presents a unique opportunity to investigate the nature of dark energy and the evolution of the universe, offering an abundance of data that can enhance our comprehension of both the galaxy-halo connection and cosmological parameters. This thesis delves into a comprehensive exploration of the galaxy-halo connection within the context of the DESI survey, emphasizing its pivotal role in interpreting large galaxy surveys and extracting cosmological insights.

We underscore the importance of the Halo Occupation Distribution (HOD) model, a statistical tool designed to populate galaxies from dark matter halos. A profound understanding of this model is essential for accurately interpreting the galaxy clustering data from DESI and for inferring the underlying cosmological model.

In this thesis, we emphasize the significance of higher-order statistics, such as the projected three-point correlation function, for constraining the galaxy-halo connection. While two-point statistics have been extensively utilized, we demonstrate that the additional information provided by higher-order statistics enables us to glean more information on 1-halo scales and constrain the HOD parameters with greater precision. We adopt and generalize a swift HOD fitting pipeline to significantly reduce the required computation time for the three-point analysis.

We present the HOD analysis of Luminous Red Galaxy (LRG) and Quasi-Stellar Object (QSO) samples derived from the One-Percent survey conducted by DESI. We scrutinize the HOD parameters as well as physical parameters such as the satellite fraction and mean halo mass of LRGs within a redshift range of $0.4 < z < 1.1$, and QSOs within a redshift range of $0.8 < z < 2.1$. Notably, for LRGs, we detect clear trends of redshift evolution for these physical parameters.

We demonstrate how a well-constrained galaxy-halo connection model can be employed to extract cosmological information from DESI data, particularly on small scales. Finally, we present a systematic study of the Jackknife resampling technique, which is widely used in HOD fitting, in the estimation of the covariance matrix.

Constraining galaxy-halo connection and cosmology using DESI data

by

Hanyu Zhang

B.S., Nanjing University, China, 2015

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Approved by:

Major Professor
Lado Samushia

Copyright

Abstract

The Dark Energy Spectroscopic Instrument (DESI) survey presents a unique opportunity to investigate the nature of dark energy and the evolution of the universe, offering an abundance of data that can enhance our comprehension of both the galaxy-halo connection and cosmological parameters. This thesis delves into a comprehensive exploration of the galaxy-halo connection within the context of the DESI survey, emphasizing its pivotal role in interpreting large galaxy surveys and extracting cosmological insights.

We underscore the importance of the Halo Occupation Distribution (HOD) model, a statistical tool designed to populate galaxies from dark matter halos. A profound understanding of this model is essential for accurately interpreting the galaxy clustering data from DESI and for inferring the underlying cosmological model.

In this thesis, we emphasize the significance of higher-order statistics, such as the projected three-point correlation function, for constraining the galaxy-halo connection. While two-point statistics have been extensively utilized, we demonstrate that the additional information provided by higher-order statistics enables us to glean more information on 1-halo scales and constrain the HOD parameters with greater precision. We adopt and generalize a swift HOD fitting pipeline to significantly reduce the required computation time for the three-point analysis.

We present the HOD analysis of Luminous Red Galaxy (LRG) and Quasi-Stellar Object (QSO) samples derived from the One-Percent survey conducted by DESI. We scrutinize the HOD parameters as well as physical parameters such as the satellite fraction and mean halo mass of LRGs within a redshift range of $0.4 < z < 1.1$, and QSOs within a redshift range of $0.8 < z < 2.1$. Notably, for LRGs, we detect clear trends of redshift evolution for these physical parameters.

We demonstrate how a well-constrained galaxy-halo connection model can be employed to extract cosmological information from DESI data, particularly on small scales. Finally, we present a systematic study of the Jackknife resampling technique, which is widely used in HOD fitting, in the estimation of the covariance matrix.

Table of Contents

List of Figures

List of Tables

Acknowledgments

Completing this thesis has been a transformative journey and an endeavor that would not have been possible without the guidance and support of many individuals.

First and foremost, I extend my deepest gratitude to my supervisor, Lado Samushia. His unwavering support and invaluable mentorship have been a beacon during the course of my research. I have been privileged to learn from his expertise, profound knowledge, and insightful perspectives. His patience and wisdom have not only guided me in my academic journey but have also shaped my professional growth. His belief in me has been a constant motivation and I am truly grateful for his supervision.

I would also like to extend my sincere thanks to Gongbo Zhao, Shadab Alam, Chia-Hsun Chuang, Ashley Ross, Sihan Yuan, Antoine Rocher. Their comments and useful insights have helped to shape, refine, and focus my research. Their guidance has allowed me to reach new heights of understanding in my field and I am immensely grateful for their involvement.

Special thanks to Shun Satio, Hee Jong Seo, Andrew Hearin, David Pearson, who have been instrumental in expanding my knowledge through the Midwest Journal Club. Their expertise and willingness to share their insights have played a significant role in broadening my academic perspective. Their constructive suggestions have fostered my intellectual growth and have made me a better researcher.

A heartfelt thank you to my committee members, Bharat Ratra, Glenn Horton-Smith, and David Auckly for their invaluable insights, and the time they invested. A special appreciation to Kim Coy, whose warmth, understanding, and supportive nature transformed our department into a place that truly feels like home

I am grateful to my dear friends in Kstate, Hongyu, Keyu, Shulei, Issac, Pavan, Lydia, Chris, Eric, Claudia, Sajed, Abhishek, Jayashree, Giorgi, and many more for the endless fun we had together. I would thank my friends Yuanchen, Hongyi, Hao, and Tiejun for those gaming afternoons and nights during the pandemic that offered me a much-need escape and joy. Thanks to Zerui, Jingsan, Yingying, and Sherry, for the texts chat that were always been a mix of humor and intellect.

I am thankful to all my peers and colleagues, both in the DESI collaboration and Kstate, whose camaraderie and support have made this journey enjoyable and worthwhile. The countless discussions, debates, and shared moments of triumph and struggle have enriched this journey immeasurably.

Finally, I am profoundly grateful to my parents, whose constant love, support, and guidance have shaped me into the person I am today.

Dedication

This thesis is dedicated to my beloved parents, Yingchun and Zheng, whose endless love, unwavering belief in my potential, and invaluable sacrifices have been the foundation of all my achievements.

Chapter 1

Introduction

1.1 Modeling the galaxy-halo connection

The ΛCDM model of modern cosmology is the most widely accepted framework that describes the evolution of the Universe. Planck Collaboration et al.[2] find that under the ΛCDM framework, around 95% of the universe is made of dark matter and dark energy. The distribution of dark matter, therefore, serves as an important source of cosmological information, revealing key properties of the Universe. However, the dark matter can not be directly observed. Galaxies, on the other hand, are the most accessible tracer on cosmological scales that galaxy surveys can observe. The galaxy-halo connection links the observable properties of galaxies with the underlying dark matter halos in which they reside, provides critical insights into the processes governing galaxy formation and evolution, and thus plays a vital role in understanding the large-scale structure of the universe.

The Dark Energy Spectroscopic Instrument[3] (DESI) currently stands as the most extensive ongoing spectroscopic galaxy survey, aiming to unveil the nature of dark energy and achieve percent-level precision in measuring crucial cosmological parameters. Over the next five years, DESI is expected to gather around 40 million galaxy and quasar spectra while spanning over 14,000 square degrees of the sky. This impressive feat represents a 20-fold increase in data collection compared to its predecessors, the Baryon Oscillation Spectroscopic

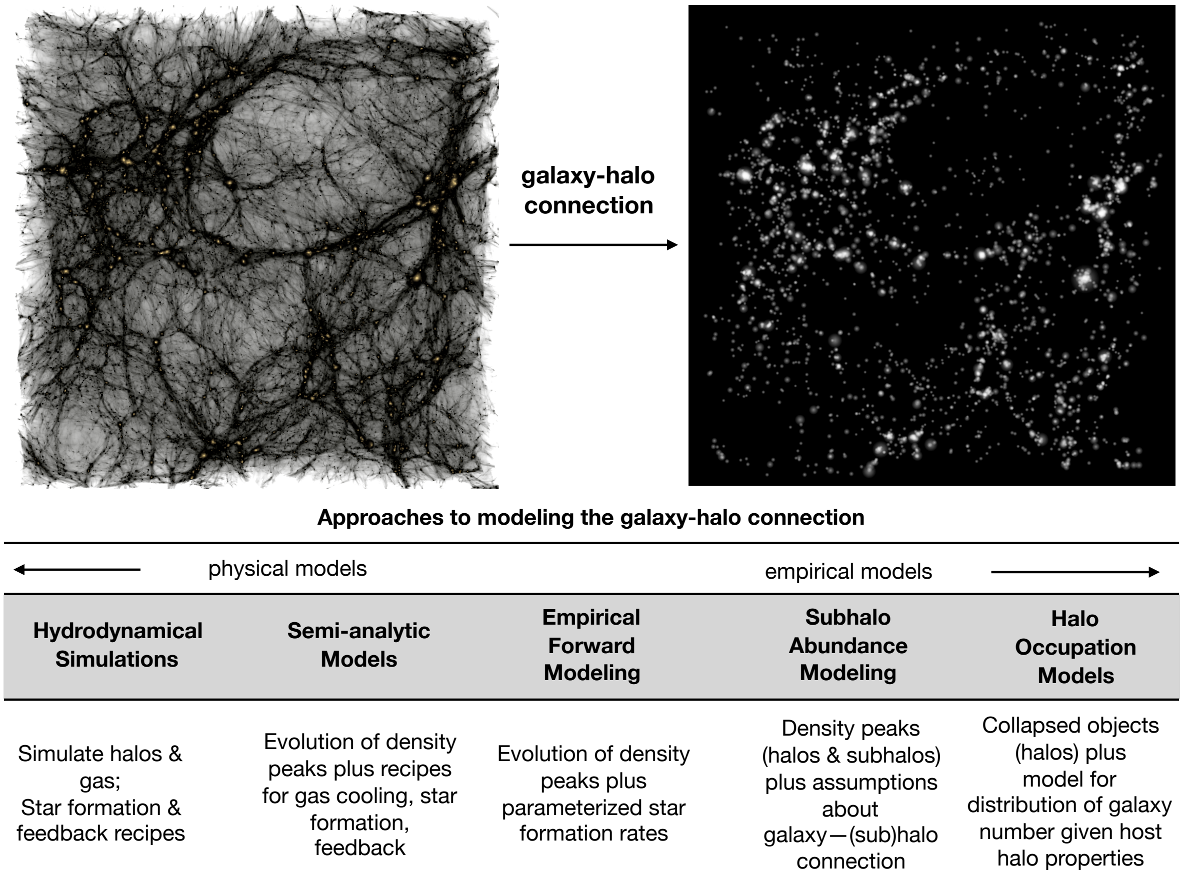


Figure 1.1: Figure from Wechsler and Tinker[1] describe different approaches to model the galaxy-halo connection. The top left panel shows a snapshot of a cosmological simulation, and the top right panel shows galaxy distribution populated using an abundance matching approach. Different models are listed on the bottom from left to right ranging from more physical to more empirical.

Survey (BOSS) and the extended BOSS (eBOSS), both of which were part of the Sloan Digital Sky Survey (SDSS)[4].

To fully capitalize on the galaxy property measurements obtained from galaxy surveys, it is crucial to develop a robust model that can accurately describe the relationship between galaxies and their underlying dark matter. This will enable us to effectively account for uncertainties in the galaxy-halo connection, ensuring more reliable cosmological interpretations of the data. On the other hand, obtaining constraints on the galaxy-halo connection model can also contribute to a deeper understanding of the physics involved in galaxy formation and the evolution of matter distribution.

A variety of methodologies have been developed to investigate this connection, each with its own strengths and limitations. Fig. 1.1 shows a schematic summary of these approaches to the galaxy–halo connection, popular models of the galaxy-halo connection were well-reviewed by Wechsler and Tinker[1], so we will briefly summarize some of them here and refer readers to Wechsler and Tinker[1] for more details.

1. Hydrodynamical simulations are the most detailed and computationally intensive method for modeling the galaxy-halo connection, providing a comprehensive and detailed view of the complex processes involved in galaxy formation and evolution. These simulations incorporate both dark matter and baryonic physics, solving the equations of gravity and hydrodynamics in a cosmological context. They track the evolution of density fluctuations in the universe, leading to the formation of a cosmic web of dark matter halos. Simultaneously, they simulate the gas dynamics within these halos, including cooling, heating, star formation, and feedback processes, which give rise to the diverse population of galaxies we observe. By directly modeling these physical processes, hydrodynamical simulations offer a first-principles approach to the galaxy-halo connection. They provide insights into how the properties of galaxies are shaped by their host halos and the surrounding cosmic environment, and they enable predictions for a wide range of galaxy observables. Hydrodynamical simulations were used to study and help building the model of the halo occupations[5–11], understand the semi-analytic models[12], and test the assumptions of the subhalo abundance matching approach[10;13;14].

2. Semi-analytic models (SAMs)[15–20] offer a more computationally efficient approach to explore the galaxy-halo connection. SAMs incorporate a set of simplified, yet physically motivated, equations to track the evolution of galaxies within the hierarchical structure of the universe. They combine the detailed merger histories of dark matter halos, derived from N-body simulations or extended Press-Schechter formalism[21], with a series of analytic prescriptions for baryonic physics, including gas cooling, star formation, feedback mechanisms, and galaxy mergers. This approach allows SAMs to generate large synthetic galaxy catalogs, linking galaxy properties to their host halo

characteristics. By adjusting the parameters of the analytic prescriptions, SAMs can be tuned to reproduce a wide range of observed galaxy statistics, providing valuable insights into the processes that govern galaxy formation and evolution.

3. The Subhalo Abundance Matching (SHAM)[22–25] model suggests that every observable galaxy is hosted by a dark matter subhalo, and the properties of these galaxies are intrinsically linked to their host subhalo's characteristics. Typically, the most luminous or massive galaxies are associated with the most massive halos or subhalos, or those with the highest maximum circular velocity. This method involves ranking both galaxies and halos/subhalos according to a selected property, such as luminosity, stellar mass, or maximum circular velocity, and subsequently matching them. Central galaxies are generally associated with primary halos, while satellite galaxies are linked to subhalos. The model parameters are then fine-tuned until the simulated galaxy distribution aligns with observed galaxy statistics.

4. Halo occupation models[26–35] describe the statistical relationship between the number of galaxies residing within a dark matter halo and the properties of the halo itself. These models are typically based on the halo occupation distribution (HOD), which quantifies the probability that a halo of a given mass contains a specific number of galaxies. By fitting the parameters of the HOD to match observed galaxy statistics, such as the galaxy two-point correlation function, the HOD models provide a powerful tool for connecting the distribution of galaxies to the underlying dark matter distribution.

Most of the research presented in this thesis will delve into the galaxy-halo connections utilizing the Halo Occupation Distribution (HOD). We provide a detailed discussion of the HOD in the subsequent section.

1.2 Halo occupation distribution

The halo occupation distribution was developed in response to the growing availability of high-quality observational data from large galaxy surveys. This approach offers a probabilistic description of the number of galaxies residing within a dark matter halo based on the halo's properties. At present, the HOD is among the most popular methods for placing galaxies in dark matter halos, utilizing data from dark matter simulations.

1.2.1 Baseline HOD

In baseline HOD models, the probability of a halo hosting a specific number of galaxies depends solely on its mass. This probability distribution is denoted as $P(N_g|M_h)$, where N_g represents the number of galaxies within a given halo and M_h signifies the mass of the halo. Typically, the distribution is calculated separately for central galaxies residing in the halos and satellite galaxies orbiting within them. Central galaxies are assumed to reside in all halos above a certain mass threshold, while satellite galaxies begin to populate larger halos following a Poisson distribution. Among several analytic forms proposed for $\langle N|M\rangle$, the most canonical one is a 5-parameter model originally proposed by Zheng et al. [34]:

$$\langle N_{\rm cen}\rangle(M) = \frac{1}{2}\left(1 + \mathrm{erf}\left[\frac{\log(M) - \log(M_{\rm cut})}{\sigma}\right]\right). \tag{1.1}$$

$$\langle N_{\rm sat}\rangle(M) = \left(\frac{M - M_0}{M_1}\right)^{\alpha}. \tag{1.2}$$

In both formulas, M represents the mass of the host halo. $M_{\rm cut}$ is the characteristic minimum mass required to host a galaxy, while sigma describes the steepness of the probability increase with halo mass around $M_{\rm cut}$. M_0 is a mass threshold for satellite galaxies, and α controls the steepness of the increase in satellite probability with host halo mass. Lastly, M_1 establishes the typical mass scale necessary for a halo to host one satellite. This model has proven effective in accurately describing the Luminous Red Galaxies (LRGs) in both the BOSS and eBOSS surveys[36–40]. For other cosmic markers, such as Bright Galaxy Samples (BGS), Emission Line Galaxies (ELG)[41], and Quasi-Stellar Objects (QSO)[42;43], a similarly formed baseline HOD has also been successfully employed to represent their clustering measurements.

1.2.2 Decorations to the baseline HOD

Assembly bias

Assembly bias refers to the phenomenon where the clustering of dark matter halos depends not only on halo mass but also on other properties related to their formation histories, such as their concentration or age. There is substantial evidence for assembly bias, various studies have been conducted by Hadzhiyska et al.[9],[10]; Croton et al.[44]; Gao et al.[45]; Pujol et al.[46]; Artale et al.[47]; Zehavi et al.[48]. This effect introduces additional complexity to the galaxy-halo connection and can serve as a "decoration" to the baseline halo occupation distribution (HOD) model, which typically only considers halo mass.

In a decorated HOD model incorporating assembly bias, the properties of galaxies are tied not just to the mass of their host halo, but also to these additional halo properties. For instance, at fixed halo mass, older or more concentrated halos might be more likely to host a galaxy or to host a galaxy of a particular type. These additional dependencies can modify the statistical distribution of galaxies within halos, and hence the predicted galaxy clustering, compared to the baseline HOD model.

Decorated HOD models provide a more nuanced view of the galaxy-halo connection and offer the potential to extract more information from galaxy surveys. By fitting these models to the data, we can learn about the impact of assembly bias on galaxy formation and evolution, and test our theoretical understanding of these processes. However, they also pose additional challenges, as they require a more detailed understanding of halo formation histories and their relationship to galaxy properties.

Conformity bias

Conformity bias, or galactic conformity, is another interesting aspect to consider when augmenting the baseline Halo Occupation Distribution (HOD) model. Galactic conformity refers to the observed phenomenon where the properties of neighboring galaxies are more similar than would be expected by chance, implying a correlation in galaxy properties at fixed halo mass. This effect can be seen both within a single dark matter halo (one-halo conformity) and between galaxies in separate halos (two-halo conformity). Incorporating conformity into HOD models, involves modifying the baseline model to capture these correlations, introducing dependencies on additional parameters such as local galaxy density or the properties of neighboring galaxies. For instance, studies based on SAM[49–52] and hydrodynamical simulations[11;53] suggest an increasing satellite occupation at high halo mass for ELGs, this effect could be taken into account by including central-satellite conformity that satellite occupation may be conditioned by the presence of central galaxies[54]. **?** find a satellite occupation of the DESI one-percent ELGs in good agreement when introducing the central-satellite conformity.

Velocity bias

Enhancing the baseline HOD model also requires considering galaxy velocity bias. This term refers to the phenomenon where the velocities of galaxies within a halo do not exactly mirror the velocities of the dark matter particles. This discrepancy can arise due to various factors, including dynamical friction, tidal stripping, and satellite-satellite interactions, leading to systematic differences between the velocity distributions of galaxies and dark matter. Velocity bias has been tested both in hydrodynamical simulations and observational surveys[41;55–57].

By parameterizing velocity bias, we can apply modifications to the 3D velocity of galaxies. However, it's worth noting that only the line-of-sight component is of significance when modeling redshift space clustering. We will briefly introduce how we parameterize velocity bias in Chap. 3 but refer the reader to Yuan et al.[58] for more details.

1.3 Constraining the HOD model

1.3.1 Observational statistics

Two point correlation functions

The two-point correlation function (2PCF), often denoted as $\xi(\mathbf{r})$, is a fundamental statistic in the field of cosmology and galaxy clustering studies. It quantifies the excess probability of finding a pair of galaxies separated by a distance r, compared to what would be expected for a random, unclustered distribution. In essence, the 2PCF provides a measure of the strength of galaxy clustering as a function of scale.

The computation of the 2PCF involves pair counting. For a given separation distance $\mathbf{r}$, the number of galaxy pairs in the observed data (Data-Data pairs, DD) and in a random distribution (Random-Random pairs, RR) are counted. The 2PCF is then typically calculated as the ratio of these pair counts, minus one:

$$\xi(\mathbf{r}) = DD/RR - 1 \tag{1.3}$$

Several estimators have been proposed to compute the 2PCF, aimed at minimizing biases and maximizing the signal-to-noise ratio. The most commonly used estimators include the Peebles and Hauser [59] (PH), Davis and Peebles [60] (DP), Hamilton [61] (HAM), and Landy and Szalay [62] (LS) estimators.

The Landy and Szalay estimator, often used in practice, is given by:

$$\xi(\mathbf{r}) = (DD - 2DR + RR)/RR, \tag{1.4}$$

where DR is the number of cross pairs between the data and the random catalog. This estimator has been shown to have the smallest variance and is unbiased for a sufficiently large random catalog.

Through the 2PCF, astronomers can probe the spatial distribution of galaxies and extract valuable information about the underlying cosmology and galaxy formation processes, such as those encoded in HOD models. These functions have been instrumental in constraining the parameters of HOD models and understanding the galaxy-halo connection.

The 2PCF $\xi(\mathbf{r})$ can be decomposed into $\xi(r_\mathrm{p}, \pi)$, a function of the projected separation, r_p, and the separation along the line-of-sight, π. The projected 2PCF, often denoted as $w_\mathrm{p}^{(2)}(r_\mathrm{p})$, measures the excess probability of finding galaxy pairs as a function of r_p by integrating $\xi(r_\mathrm{p}, \pi)$ over π, thereby integrating out the line-of-sight velocity differences.

$$w_\mathrm{p}^{(2)}(r_\mathrm{p}) = \int_{-\pi^\star}^{\pi^\star} d\pi \xi^{(2)}(r_\mathrm{p}, \pi) \tag{1.5}$$

The value of $\pi^\star$ is usually chosen to be of the order of a few tens of megaparsecs to smooth over peculiar velocity effects. Projected 2PCF $w_\mathrm{p}^{(2)}(r_\mathrm{p})$ is particularly useful as it is less affected by redshift-space distortions and provides a clear measure of the spatial clustering of galaxies. $\xi(r_\mathrm{p}, \pi)$ can also be used directly to constraint HOD parameters, as it contains the full information content of the 2PCF, including velocity distribution and the small-scale finger-of-god effect. Yuan et al. [63] showed that the $\xi(r_\mathrm{p}, \pi)$ have extra constraining power of HOD on small scales and can be used to constraint velocity bias.

The multipole expansion of the 2PCF, $\xi_\ell(s)$, on the other hand, is a powerful tool for probing redshift-space distortions, which arise due to peculiar velocities of galaxies. The 2PCF $\xi(\mathbf{r})$ can be decomposed into multipoles as a function of the separation s and the cosine of the angle μ between the separation vector and the line-of-sight direction.

$$\xi(s, \mu) = \sum_\ell \xi_\ell(s) P_\ell(\mu), \tag{1.6}$$

where $P_\ell(\mu)$ are the Legendre polynomials. Each of these moments, $\xi_\ell(s)$, can be computed by performing an angular integration:

$$\xi_\ell(s) = \frac{2\ell + 1}{2} \int_{-1}^{1} \xi(s, \mu) P_\ell(\mu) d\mu. \tag{1.7}$$

These multipoles capture different aspects of the galaxy distribution, with $\ell = 0, 2, 4$ corresponding to the monopole, quadrupole, and hexadecapole terms, respectively.

The monopole ($\ell = 0$) measures the isotropic (angle-averaged) clustering, and it's typically affected by both real-space clustering and redshift-space distortions due to coherent infall into overdense regions. The quadrupole ($\ell = 2$) and hexadecapole ($\ell = 4$) are sensitive to the anisotropic clustering in redshift space. The quadrupole moment, in particular, provides information about the"flattening" of the correlation function along the line of sight due to peculiar velocities, known as the Kaiser effect. On large scales, the quadrupole and hexadecapole moments can be used to probe cosmological parameters, such as the growth rate of structure. For HOD fitting, quadrupole and hexadecapole are key sources of constraining power for velocity bias parameters.

Three point correlation functions

The three-point correlation function (3PCF), denoted as $\xi^{(3)}(\mathbf{r}_{12}, \mathbf{r}_{23}, \mathbf{r}_{31})$, is another essential statistic in the field of cosmology and galaxy clustering studies. While the two-point correlation function (2PCF) provides information about the pairwise clustering of galaxies, the 3PCF extends this analysis to triplets of galaxies. It quantifies the excess probability of finding a triplet of galaxies with specified separations $\mathbf{r}_{12}$, $\mathbf{r}_{23}$, and $\mathbf{r}_{31}$, compared to what would be expected for a random, unclustered distribution.

The 3PCF provides additional information beyond the 2PCF, including insights into the shape of galaxy clustering (since a triplet of galaxies defines a triangle) and non-Gaussian features of the galaxy distribution. It is sensitive to the biasing relationship between galaxies and the underlying dark matter distribution, and can thus provide additional constraints on the processes of galaxy formation and evolution[58;64–70].

The computation of the 3PCF, like the 2PCF, involves counting sets of points, but in this case triplets rather than pairs. The challenge lies in efficiently counting these triplets in configuration space and minimizing the computational expense.

For a given set of separations $(\mathbf{r}_{12}, \mathbf{r}_{23}, \mathbf{r}_{31})$, the number of galaxy triplets in the observed data (Data-Data-Data triplets, DDD) and in a random distribution (Random-Random-Random triplets, RRR) are counted. The 3PCF is then typically calculated as the ratio of these triplet counts, minus one:

$$\xi^{(3)}(\mathbf{r}_{12}, \mathbf{r}_{23}, \mathbf{r}_{31}) = DDD/RRR - 1 \tag{1.8}$$

Several estimators have been proposed to compute the 3PCF, including the Szapudi and Szalay and Peebles and Groth estimators. These estimators incorporate cross-correlations between the data and random catalogs, similar to the Landy and Szalay estimator for the 2PCF.

Similar to the 2PCF, the 3PCF $\xi^{(3)}(\mathbf{r}_{12}, \mathbf{r}_{23}, \mathbf{r}_{31})$ can be decomposed into $\xi^{(3)}(r_{\mathrm{p}12}, r_{\mathrm{p}23}, r_{\mathrm{p}31}, \pi_1, \pi_2)$, a function of three perpendicular separations and two relative distances along the line of sight. The projected 3PCF can then be computed by

$$w_{\mathrm{p}}^{(3)}(r_{\mathrm{p}12}, r_{\mathrm{p}23}, r_{\mathrm{p}31}) = \int_{-\pi^\star}^{\pi^\star} d\pi_1 d\pi_2 \xi^{(3)}(r_{\mathrm{p}12}, r_{\mathrm{p}23}, r_{\mathrm{p}31}, \pi_1, \pi_2) \tag{1.9}$$

Chap. 2 of this thesis shows in detail how projected 3PCF can be efficiently taken into account for HOD analysis and be a powerful probe of constraining the HOD parameters.

Other statistics

While two-point correlation functions (2PCF) have traditionally been a workhorse in constraining the parameters of Halo Occupation Distribution (HOD) models, there are other statistical tools that can provide complementary insights into the galaxy-halo connection. These include Counts in Cylinders, Satellite Kinematics, and Galaxy Void Statistics.

Counts in Cylinders (CIC) is a technique that counts the number of galaxies within a cylinder of a given radius and height around a particular galaxy. The distribution of these counts can provide constraints on the number and spatial distribution of satellite galaxies within their host halos, contributing to the parameterization of the HOD model. CIC is popular in constraining galaxy halo connection, recent works include Reid and Spergel [71]; Berrier et al. [72]; Oguri and Lin [73]; Gruen et al. [74]; Wang et al. [75]

Satellite Kinematics provides another avenue for constraining HOD parameters, specifically those related to the velocity distribution of satellite galaxies. By measuring the velocity dispersion of satellite galaxies around their host, we can gain insights into the mass of the host halo. This information complements the spatial distribution information obtained from the 2PCF or Counts in Cylinders, providing a more complete picture of the galaxy-halo connection[76;77].

Galaxy Void Statistics refers to the study of large underdense regions in the distribution of galaxies. The size distribution and abundance of these voids are sensitive to both the matter density and the galaxy bias, making them useful for constraining cosmological parameters as well as the parameters of the HOD model. The population of galaxies within voids and on their edges can also provide additional constraints on how galaxies populate halos[78–81].

1.3.2 Fast HOD analysis pipeline

In the standard procedure for fitting the HOD model, galaxy mock populations are generated for each set of HOD parameters that have been sampled. The clustering statistic of interest is subsequently measured and compared to measurements obtained from the data in order to estimate the likelihood. This procedure is repeated multiple times for various HOD parameters until the posterior likelihood has been thoroughly explored. An alternative approach, known as the tabulation method, was first proposed by Neistein et al. [82] and later expanded upon by Zheng and Guo[83]. This method reverses the order of applying the HOD model and measuring the clustering. Specifically, the clustering of halos is precomputed prior to the Markov Chain Monte Carlo (MCMC) stage, and the HOD population scheme is subsequently applied by combining weights with the halo clustering. This approach significantly improves the efficiency of the HOD fitting process, as the most computationally intensive step is moved outside of the MCMC loop. We summarise the tabulation method for the AbacusSummit simulation below but refer the reader to Zhang et al. [84]. for more details.

We take projected 2PCF w_{p} here as an example and assume our HOD model only depend on the mass of the host halo. To match the behavior of the particle-based HOD, where we populate the satellite with particles, we first divide the halo catalog and particle catalog attached to the halos into N_{b} bins. Then the galaxy correlation function $w_{\mathrm{p,gg}}$ is given by a weighted sum over different mass bin cross-correlations.

$$\begin{aligned} w_{\mathrm{p,gg}}(r_{\mathrm{p}}) = &\sum_{i,j}^{N_{\mathrm{b}}} w_{\mathrm{cent}}(M_i) w_{\mathrm{cent}}(M_j) w_{\mathrm{p,hh}}(r_{\mathrm{p}}, M_i, M_j) \\ &+ 2\sum_{i,j}^{N_{\mathrm{b}}} w_{\mathrm{cent}}(M_i) w_{\mathrm{sat}}(M_j) w_{\mathrm{p,hp}}(r_{\mathrm{p}}, M_i, M_j) \\ &+ \sum_{i,j}^{N_{\mathrm{b}}} w_{\mathrm{sat}}(M_i) w_{\mathrm{sat}}(M_j) w_{\mathrm{p,pp}}(r_{\mathrm{p}}, M_i, M_j), \end{aligned} \tag{1.10}$$

where $w_{\mathrm{p,hh}}(r_{\mathrm{p}}, M_i, M_j)$ is the two-point cross-correlation function of halos in the ith and jth mass bins (similarly for the halo-particle and particle-particle correlation functions) and naively we could take the weight as

$$w_{\mathrm{cent}}(M_i) = \bar{n}_{\mathrm{cent}}(M_i), \tag{1.11}$$

$$w_{\mathrm{sat}}(M_i) = \bar{n}_{\mathrm{sat}}(M_i)\frac{N_{\mathrm{h}}^i}{N_{\mathrm{p}}^i}. \tag{1.12}$$

where N_{h}^i and N_{p}^i are numbers of halos and particles in ith mass bin.

Equation 1.10 gives an average value of clustering expected for a given HOD model instead of a specific realization that includes stochastic noise and further reduces the bias of HOD fitting. We also test and confirm that the mean prediction is consistent with the output of the standard HOD fitting pipeline.

While the tabulation method can accelerate the fitting of HOD, it also limits the flexibility of extending the HOD model. The baseline HOD model relies solely on halo mass, making it easy to prepare tabulated halo and particle pair counts across mass bins. However, when dealing with more complex HOD models that involve velocity bias and assembly bias extensions, additional dimensions of dependency can significantly increase the complexity of preparing tabulated halo and particle pair counts.

1.4 Generating mock catalogs

The HOD model is a powerful tool for generating mock catalogs of galaxies, which are essential for interpreting galaxy surveys. Given a dark matter halo catalog from an N-body simulation, the HOD model can populate these halos with galaxies, producing a mock galaxy catalog. These catalogs can reproduce the observed clustering properties of galaxies and provide robust covariance matrices, which are critical for cosmological parameter estimation.

1.5 Outline of the dissertation

This thesis begins by harnessing the power of high-order statistics, such as the projected 3PCF, to delve deeper into the complex relationship between galaxies and their dark matter halos in Chap. 2. This exploration illuminates the performance of projected 3PCF in constraining the HOD for key tracers in the DESI survey, including Luminous Red Galaxies, Emission Line Galaxies, and Quasi-Stellar Objects. For mock LRG samples, the constraints on some parameters have improved by as much as 70%. Additionally, the development of a fast HOD fitting pipeline, incorporating a tabulation method up to the three-point level, is introduced, paving the way for more efficient analysis.

The exploration continues with an in-depth examination of data from the DESI One-Percent Survey as described in Chap. 3. The focus here is on the HOD parameters for LRGs from the redshift range $0.4 < z < 1.1$ and QSOs from the redshift range $0.8 < z < 2.1$, where we find that the baseline HOD model effectively captures small-scale clustering, and we detect clear trends in the redshift evolution of the HOD parameters and physical parameters for the LRG sample.

Chap. 4 then shifts to the implementation of the HOD framework, demonstrating its utility in extracting cosmological information from small-scale observations. A comprehensive exploration of the degeneracy between HOD parameters and cosmological parameters on small-scale two-point correlation function multipoles is presented.

Following the implementation of the HOD framework, Chap. 5 tests the performance of the jackknife resampling method in estimating covariance matrices for different two-point statistics, including two-point correlation functions, power spectrum, the difference and cross-covariance of them from a pair of simulations. This exploration provides valuable insights into the robustness of the jackknife variance, contingent upon various factors such as the number of jackknife regions and the choice between two-dimensional and three-dimensional regions.

Chap. 6 concludes with a summary of the key insights gained from this intricate exploration of the galaxy-halo connection and cosmology using DESI data. These findings not only contribute to the current understanding of these areas but also provide direction for future work in this field.

Chapter 2

Constraining galaxy-halo connection with high-order statistics

This chapter originally appeared in the literature as

Hanyu Zhang, Lado Samushia, et al. Constraining galaxy–halo connection with high-order statistics.

Chapter Abstract

We investigate using three-point statistics in constraining the galaxy-halo connection. We show that for some galaxy samples, the constraints on the halo occupation distribution parameters are dominated by the three-point function signal (over its two-point counterpart). We demonstrate this on mock catalogs corresponding to the Luminous Red Galaxies (LRGs), Emission-Line Galaxies (ELG), and quasars (QSOs) targeted by the Dark Energy Spectroscopic Instrument (DESI) Survey. The projected three-point function for triangle sides less up to $20h^{-1}$ Mpc measured from a cubic Gpc of data can constrain the characteristic minimum mass of the LRGs with a precision of 0.46 %. For comparison, similar constraints from the projected two-point function are 1.55 %. The improvements for the ELGs and QSOs targets are more modest. In the case of the QSOs it is caused by the high shot-noise of the sample, and in the case of the ELGs, this is caused by the range of halo masses of the host halos. The most

time-consuming part of our pipeline is the measurement of the three-point functions. We adopt a tabulation method, proposed in earlier works for the two-point function, to reduce significantly the required compute time for the three-point analysis.

2.1 Introduction

Simulations of structure formation have become invaluable in analyzing cosmological data[85;86]. They are used for studying nonlinear gravitational evolution, validating and calibrating theoretical models of structure formation, and estimating covariance matrices of clustering measurements. Cold dark matter simulations are the easiest to produce. They provide us with an accurate picture for the clustering of dark matter halos[87;88]. The positions of galaxies cannot be obtained from the cold dark matter simulations. They depend on baryonic physics that is not captured by the cold dark matter simulations[89;90]. In addition, resolving galaxies in large volumes requires a much higher mass resolution that cannot be realized with current computers. Galaxy surveys, on the other hand, measure positions of galaxies rather than their host dark matter halos. It is essential to have an accurate method of placing galaxies in these dark matter simulations for the robust analysis of such data.

The Halo Occupation Distribution (HOD) approach is currently one of the most widely used methods to achieve this goal[26–28;30–35]. In the HOD framework galaxies are placed in halos based on some probabilistic prescription that depends on the properties of the host halo and its neighborhood. In the basic HOD models, the probability of a halo to host a certain number of galaxies only depends on its mass. In more complicated models it can also depend on the local density of halos around the host and some features of the history of the halo formation. Models of various complexity have been offered for where exactly to place the galaxies inside the halo and how to assign velocities to those galaxies.

An alternative approach to connect galaxies and halos is the sub-halo abundance matching (SHAM) method[22;24;25;91–93]. By assuming a monotonic relation between certain halo properties and certain galaxy properties, a galaxy catalog can be generated by matching the observed list of galaxies sorted by galaxy property with a list of halos (and sub-halos) sorted by halo property from simulations. The method based on a Conditional Luminosity Function which models galaxies as a function of both their luminosity and host halo mass is another alternative[94–97].

The HOD models have adjustable parameters that are tuned to obtain galaxies as similar as possible to the observed sample. Traditionally, they are constrained by their 2-Point Correlation Function (2PCF), which is the likelihood of finding a pair of galaxies with a certain separation. The 2PCF for separations up to 20 h^{-1}Mpc is usually used for this purpose[36–42].

The 2PCF alone does not always have enough constraining power. Many different combinations of HOD parameters may result in a 2PCF that is consistent with the data within the measurement errors. One way of improving the constraints is to also fit the observed 3-Point Correlation Function (3PCF), which is a probability of finding a triplet of galaxies with certain side lengths and orientation concerning the line-of-sight with respect to an observer[98;99].

The usage of the 3PCF to constrain galaxy-halo connection has a long history[64–67]. In more recent works, Kulkarni et al.[68] studied the shape dependence of reduced 3PCF and find that signal from reduced 3PCF could help break the degeneracy between HOD parameters. Marín[69] measured the redshift-space 3PCF of LRGs from SDSS on large scales up to $90h^{-1}$Mpc and use the 3PCF to constraint bias parameters, which in turn help estimate the LRG HOD parameters. Guo et al.[70] explored the constraining power of redshift space 3PCF on HOD parameters including the galaxy velocity bias. Yuan et al.[58] tested the potential extra constraining power of HOD parameters from squeezed 3PCF[100].

The top diagram on Fig. 2.1 schematically shows the steps required to constrain the HOD parameters with a 2PCF or a 3PCF. For a set of HOD parameters we populate mocks with galaxies according to that model, we then measure the clustering statistics of interest, it is compared with a similar measurement from the data, and the posterior likelihood is assessed. This process is repeated many times for various HOD parameter sets until the posterior likelihood is well explored. The most time-consuming part of this algorithm is computing the 2PCF and the 3PCF. Computing the three-point correlation function is

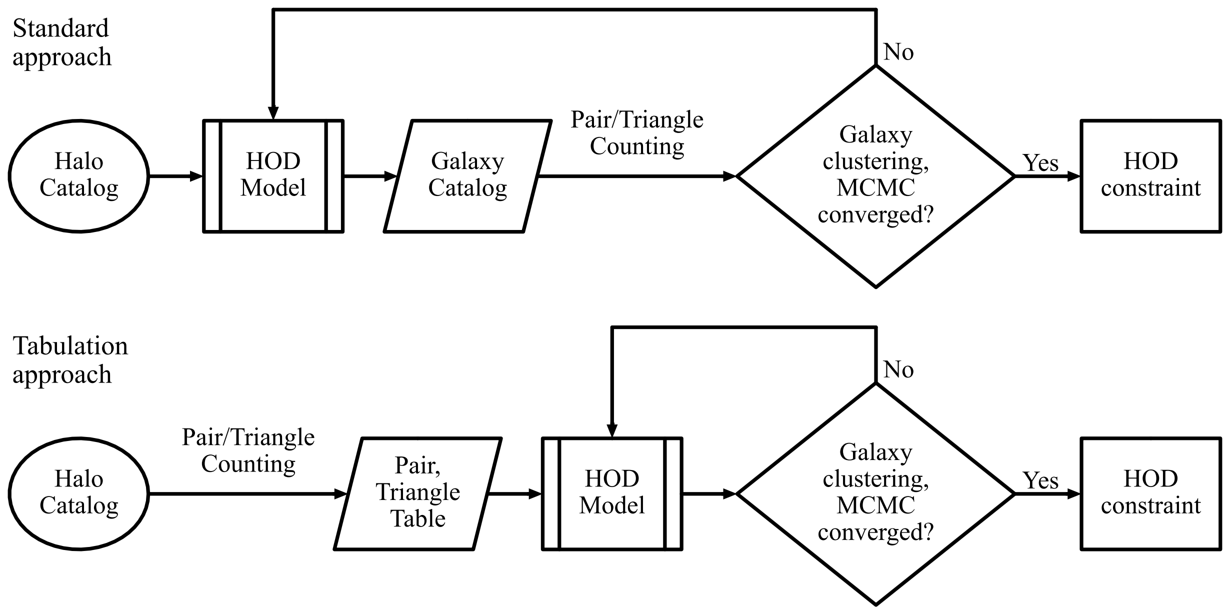


Figure 2.1: The flow chart on top shows the conventional sequence of steps leading to the HOD constraints. The bottom panel shows the same flow chart for the tabulation approach.

especially time-consuming. The number of all possible triplets scales as $N_{\rm gal}^3$, where $N_{\rm gal}$ is the number of galaxies in the sample. For a big sample, this requires looking at many millions of triangular configurations. This computation needs to be performed at each point in the MCMC chain. Recent works have proposed algorithms that make it possible to compute certain combinations of 3PCF with $N_{\rm gal}^2$ complexity[101–103], but even with these algorithmic improvements, this step remains the most computationally expensive piece in the pipeline.

The bottom diagram on Fig. 2.1 shows similar schematics for the tabulation approach that has been first proposed in Neistein et al.[82] and extended by Zheng and Guo[83]. In this approach, the 2PCF of some subsets of halos are precomputed separately before the MCMC stage. These measurements are then combined with certain weights to statistically emulate various HOD population schemes. We describe the tabulation method in detail in Sec. 2.3. This approach saves a lot of computation time since the most time-consuming part of the algorithm is performed only once before launching the MCMC chain.

The tabulation method was initially developed for the 2PCF-based fits but it is trivially generalizable to the 3PCF. Many 3PCF-based results that we present in this paper would have required prohibitive computation times with the traditional approach.

We test our method on the galaxies designed to emulate the Luminous Red Galaxies (LRGs), the Emission line galaxies (ELGs), and the Quasars (QSOs) targeted by the Dark Energy Spectroscopic Survey[104]. We show the 3PCF constraints on the HOD parameters dominate the 2PCF results for the DESI-like LRGs. 3PCF has up to 70 % improvement for a certain parameter. For the ELG and the QSO galaxies, the improvements offered by adding the 3PCF are more modest because of the lower typical host halo mass and lower density of those tracers.

2.2 HOD analysis pipeline

2.2.1 HOD model

We use a HOD prescription in which the expectation value of galaxies hosted by a dark matter halo only depends on the virial mass of the halo. The expectation value is different for central galaxies that occupy the center of the halo, and satellites that are in virial motion around the center.

For the LRGs we use

$$\langle N^c_{\rm lrg}\rangle(M) = \frac{A_c}{2}\left(1+\mathrm{erf}\left[\frac{\log(M)-\log(M_{\rm cut})}{\sigma}\right]\right). \tag{2.1}$$

$$\langle N^s_{\rm lrg}\rangle(M) = A_s\left(\frac{M-M_0}{M_1}\right)^{\alpha} H(M-M_0). \tag{2.2}$$

The central probability increases with mass until it saturates to some high mass value. The satellite probability is zero below some threshold mass but increases as a power law above that mass.

In both formulas, M is the mass of the host halo. A_c, referred to as a duty cycle in the literature, is a maximum probability for high mass halos to host an LRG. $M_{\rm cut}$ is the characteristic minimum mass to host an LRG. σ describes how steeply the probability increases with halo mass around $M_{\rm cut}$. M_0 is a mass threshold for the satellite galaxies. α

controls the steepness of the increase in the satellite probability with the host halo mass. M_1 is the extra mass above the threshold that the halo must have for the expected number of satellites to be equal to one. A_s sets an overall amplitude of the probability. In principle, this parameter is fully degenerate with M_1. We use A_s for convenience when creating mock catalogs because it can be changed independently of other parameters to adjust the overall number density of the galaxies without affecting their distribution across masses. H is a Heaviside step function.

This model has been demonstrated to describe well the LRGs in BOSS and eBOSS surveys e.g.[36–40].

For the ELGs e.g.[41] the central probability is a Gaussian function that decays at both high and low mass ends. The satellite probability is similar to the LRGs.

$$\langle N^c_{\rm elg}\rangle(M) = \frac{A_c}{\sqrt{2\pi}\sigma}\exp\left(-\frac{[\log(M)-\log(M_{\rm cut})]^2}{2\sigma^2}\right) \tag{2.3}$$

$$\langle N^s_{\rm elg}\rangle(M) = A_s\left(\frac{M-M_0}{M_1}\right)^{\alpha} H(M-M_0). \tag{2.4}$$

$M_{\rm cut}$, in this case, describes the most probable halo mass to host a central ELG and σ is the variance in the width of this pdf as a function of mass.

For the QSOs, we use a similar formula for the central probability but a slightly modified formula for a satellite probability.

$$\langle N^c_{\rm qso}\rangle(M) = \frac{A_c}{2}\left(1+\mathrm{erf}\left[\frac{\log(M)-\log(M_{\rm cut})}{\sigma}\right]\right), \tag{2.5}$$

$$\langle N^s_{\rm qso}\rangle(M) = A_s\left(\frac{M}{M_1}\right)^{\alpha}\exp\left(-\frac{M_0}{M}\right). \tag{2.6}$$

The difference from the LRG is that the QSO hosting probability decays exponentially at lower masses instead of having a sharp cutoff. M_0, in this case, controls the decay rate as we go to the lower masses, while M_1 set the normalization e.g.[42;43].

There is substantial evidence that the probability of a halo hosting a certain galaxy may depend on other parameters in addition to the virial mass, a phenomenon called an assembly bias[9;10;44–48]. In this work we ignore the assembly bias. This does not affect our main conclusions, since the main objective of our work is to study a potential improvement in the HOD parameter constraints and we don't expect our conclusions to be sensitive to the exact nature of the HOD model.

2.2.2 Mock galaxy catalog

We use the AbacusSummit cosmological N-body simulation to create mock galaxy catalogs[105–110]. AbacusSummit were designed to meet the cosmological simulation requirements of DESI. Specifically, we use the `AbacusSummit_highbase_c000_ph100` box of AbacusSummit with Planck 2018 cosmology, box size of 1000 h^{-1}Mpc per side, and 3456^3 dark matter particles with the mass of $2.1 \times 10^9\ h^{-1}M_\odot$ per particle. We use cleaned CompaSO[111] halo catalog at the $z = 0.8$, $z = 1.1$, and $z = 1.4$ snapshots to create the LRG, ELG, and QSO samples respectively. These are the redshifts at which the number densities of the tracers are expected to peak.AbacusSummit suite provides multiple nested definitions of halos out of which we use the level two (L2) halos see,[111] for the details of the halo definition We use center of mass position and velocity of the largest L2 subhalo fields, `x_L2com` and `v_L2com`, and generate our mocks in redshift space. AbacusSummit simulations come with a subsample (3 %) of particles that make each halo, which will then be used for satellite population. Based on the HOD parameters we chose, galaxies with host halo mass lower than $10^{11}h^{-1}M_\odot$ barely exist. We then set a cut-off mass and remove all halo with mass smaller than $10^{11}h^{-1}M_\odot$ when populating HOD mock catalogs for all tracers (see App.2.A).

Tab. 2.1 shows the fiducial parameter values that we use to create LRG, ELG, and QSO catalogs. They were obtained by fitting to the early version of the DESI Survey Validation data. These values may change as more DESI data is accumulated. For the purposes of our project, however, the exact fiducial values do not matter. The top panel on Fig. 2.2 shows the expected number of galaxies per halo and as a function of the halo mass, the bottom panel shows the probability distribution of host halo mass for a galaxy normalized as the probability per $\log(M)$, based on fiducial parameter values in Tab. 2.1. The host halo mass

Parameters	LRG	ELG	QSO
$\log(M_{\rm cut})$	12.70	11.70	12.50
σ	0.17	0.08	0.30
$\log(M_1)$	13.80	12.00	15.00
$\log(M_0)$	12.13	11.60	12.00
α	1.28	0.33	1.20
$A_{\rm c}$	0.70	0.025	0.05
$A_{\rm s}$	0.70	0.03	1.00
n	5.14	6.35	0.415
z	0.8	1.1	1.4

Table 2.1: Fiducial values of HOD parameters for each tracer and the resulting comoving number density in units of $10^{-4}(h^{-1}\text{Mpc})^{-3}$.

of ELG is smaller and more concentrated than that of LRG and QSO.

For each dark matter halo, we make a random decision about whether to put a central galaxy in it and how many satellites (if any) we put in the halo. We compute a probability of a central galaxy and make a random draw from Bernoulli distribution $B(1, \langle N^c\rangle)$ and if it results in 1 we put a galaxy in the center of the halo. As for satellites, we chose the particle-based population approach e.g.[58], we compute the average number of satellites then make a random draw from Bernoulli distribution $B(1, \langle N^s\rangle/N_p)$ and if it results in 1 we put a galaxy in the particle position, where N_p is the number of particles attached to the halo. The total number of galaxy distribution is then Poissonian, $p(N|M) = \text{Pois}(\langle N|M\rangle)$.

We assign the velocity of the halo to the central galaxy and the velocity of the particle to the satellite galaxy. Recent works[55;56] have shown that there is evidence for the velocity bias, the velocity of galaxies being systematically different from the velocity of the dark matter field at the same position. We ignore velocity bias in this work. This should not affect our results for the reasons outlined in the previous paragraph.

The procedure for creating mock catalogs is intrinsically stochastic. Depending on the outcomes of random draws we can get many different equivalent realizations of the galaxy population following the same HOD model on average.

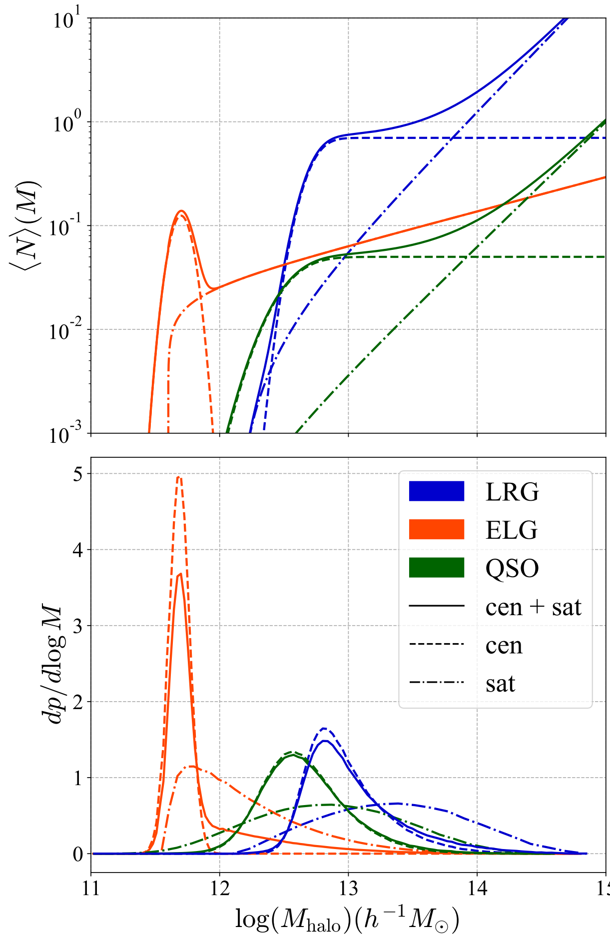


Figure 2.2: The top panel shows the expected number of galaxies hosted by a halo as a function of halo mass for the fiducial HOD parameters. The blue, orange, and green colors are for the LRG, ELG, and QSO respectively. The solid, dash and dash-dotted lines represent the expected number of all (cen+sat), central, and satellite galaxies. The bottom panel shows the probability distribution of host halo mass for a galaxy of each tracer. The solid line shows the host halo mass distribution for all, normalized as the probability per $\log(M)$, dash and dash-dotted line shows central and satellite host halo mass distribution respectively.

2.2.3 Projected correlation functions

2PCF, $\xi^{(2)}(\mathbf{r})$, describes a probability of finding two galaxies in infinitesimal volumes dV_1 and dV_2 that are separated by distance $\mathbf{r}$. This probability is proportional to $dV_1 dV_2[1+\xi^{(2)}(\mathbf{r})]$ and is conventionally normalized so that for particles distributed in space with uniform probability (all spatial points are equally likely to host a particle) $\xi^{(2)}(\mathbf{r})=0$. The 3PCF, $\xi^{(3)}(\mathbf{r}_{12},\mathbf{r}_{23},\mathbf{r}_{31})$, is similarly defined as an excess probability (over spatially uniform distribution) of finding a triplet of galaxies to be separated by $\mathbf{r}_{12}$, $\mathbf{r}_{23}$, and $\mathbf{r}_{31}$ [112]. 2PCF of observed galaxies depends only on the along and across the line-of-sight separations (with respect to the observer) of galaxies instead of the full separation vector, $\xi^{(2)}(\mathbf{r})=\xi^{(2)}(r_p,\pi)$ where r_p is a distance perpendicular to the line-of-sight and π is a distance along the line-of-sight. The 3PCF similarly depends on three perpendicular separations and two relative distances along the line of sight, $\xi^{(3)}(\mathbf{r}_{12},\mathbf{r}_{23},\mathbf{r}_{31})=\xi^{(3)}(r_{p12},r_{p23},r_{p31},\pi_{12},\pi_{23})$. The variations in the line-of-sight separation in these correlation functions depend on the velocities of the galaxies in addition to their positions. To make HOD modeling easier projected correlation functions are often used[60;113]. They are defined by

$$w_{\mathrm{p}}^{(2)}(r_{\mathrm{p}})=\int\limits_{-\pi^\star}^{\pi^\star} d\pi\xi^{(2)}(r_{\mathrm{p}},\pi) \tag{2.7}$$

$$w_{\mathrm{p}}^{(3)}(r_{\mathrm{p}12},r_{\mathrm{p}23},r_{\mathrm{p}31})=\int\limits_{-\pi^\star}^{\pi^\star} d\pi_1 d\pi_2\xi^{(3)}(r_{\mathrm{p}12},r_{\mathrm{p}23},r_{\mathrm{p}31},\pi_1,\pi_2) \tag{2.8}$$

The value of $\pi^\star$ is usually chosen to be of the order of a few tens of megaparsecs. This is done to smooth over peculiar velocity effects that affect the functional dependence of the correlation functions in the parallel to the line-of-sight direction. We derive our main results using the value of $\pi^\star=100h^{-1}$ Mpc for the projected 2PCF. This value is large enough for the residual peculiar velocity effects to be negligible. Although, these projected correlation functions will depend on the velocities of the galaxies unless $\pi^\star\to\infty$ see e.g.[114;115]. Lower values of $\pi^\star$ maybe optimal because they do not depend on large scale correlations that are noisier, but using a lower integration limit would require careful modelling of the peculiar

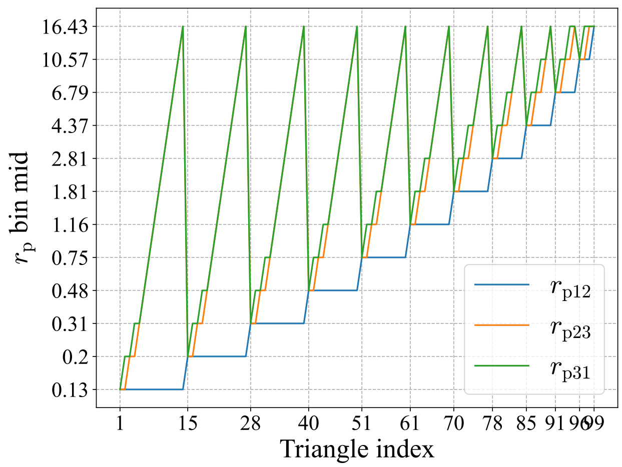


Figure 2.3: Projected separation as a function of the triangular index.

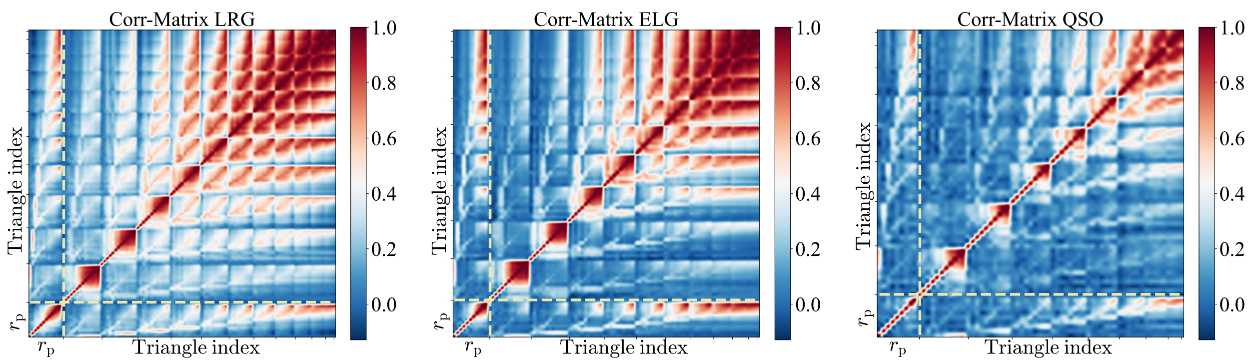


Figure 2.4: Correlation matrices including the projected 2PCF and projected 3PCF (simplified version) for LRGs (left), ELGs(mid), and QSOs(right) used in this analysis derived by Jackknife re-sampling. We include all bins between $r_{\mathrm{p}} = 0.1 - 20h^{-1}\mathrm{Mpc}$ for the projected 2PCF, all triangles of LRGs, triangle index from 15 to 99 of ELGs, and triangle index from 28 to 99 of QSOs for the projected 3PCF. The color indicates the level of correlation, where red represents 100% correlation and dark blue means a low level of anti-correlation.

velocity effects and the difference turns out not to be big enough to affect any of our main conclusions (see App. 2.B).

2.2.4 Measuring projected correlation functions

The 2PCF and 3PCF are usually measured by counting the number of galaxy pairs and triplets for the data and for uniform distribution in the same volume. They can be estimated from these pair and triplet counts by

$$\xi^{(2)}(r_{\rm p},\pi) = \frac{DD(r_{\rm p},\pi)}{RR(r_{\rm p},\pi)} - 1, \tag{2.9}$$

$$\xi^{(3)}(r_{\rm p12},r_{\rm p23},r_{\rm p31},\pi_1,\pi_2) = \frac{DDD(r_{\rm p12},r_{\rm p23},r_{\rm p31},\pi_1,\pi_2)}{RRR(r_{\rm p12},r_{\rm p23},r_{\rm p31},\pi_1,\pi_2)} - 1, \tag{2.10}$$

where DD is the number of pairs of galaxies separated by certain radial and transverse distances, DDD is the number of triplets of galaxies having a specific triangular configuration, RR and RRR are the equivalent number of pairs and triplets from a uniform random distribution. Additive factors of -1 normalize the correlations to be zero when $DD \sim RR$ and $DDD \sim RRR$.

We compute the 2-point projected correlation functions by estimating the 2PCF first and then integrating the estimated correlation functions over π. For the 3-point projected correlation function, we use a slightly modified algorithm. Instead of eq.(2.8), we compute a simplified version (SV),

$$\begin{aligned} w^{(3)}_{\rm p(SV)} &= \frac{\sum\limits_{\pi_1,\pi_2} DDD(r_{\rm p12},r_{\rm p23},r_{\rm p31},\pi_1,\pi_2)}{\sum\limits_{\pi_1,\pi2} RRR(r_{\rm p12},r_{\rm p23},r_{\rm p31},\pi_1,\pi_2)} - 1 \\ &= \frac{DDD(r_{\rm p12},r_{\rm p23},r_{\rm p31})}{RRR(r_{\rm p12},r_{\rm p23},r_{\rm p31})} - 1. \end{aligned} \tag{2.11}$$

This is not the same projected correlation function as the one defined in eq. (2.8). Eq.(2.8) compute the integral over π_1, π_2, which still need triangle counting in a 5 dimension space. Eq. (2.11), on the other hand, compute a ratio of sums over π_1, π_2, which reduced triangle counting to a 3 dimension space only relay on $r_{\rm p12}, r_{\rm p23}, r_{\rm p31}$. What we estimate with eq. (2.11) is still a three-point function that depends on the distribution of triangular configurations and it is projected in a sense that it is insensitive to the radial separation between the three galaxies (and therefore also insensitive to the velocities of the galaxies). The second function is significantly easier and faster to compute. We choose the $\hat{z}$-direction of the AbacusSummit boxes to be the line-of-sight of the observer. This makes the projected distance along z the π and the projected distance in the x-y plane the $r_{\rm p}$. This lets us completely ignore the $\hat{z}$-direction and significantly accelerate the triplet counting part of the algorithm. We use a modified version of GANPCF package [1], which is a GPU accelerated tool for N-point correlation function measurements, to compute the DDD counts defined this way.

We measure the DD counts of the projected 2PCF using CorrFunc package[116;117] setting $\pi^\star = 100h^{-1}$Mpc. A smaller value of $\pi^\star$ would result in a less noisy measurement, but since our main objective is to compare the relative constraining power of the 2 and 3 point clustering, we need to compute both in similar settings.

Our volume is a simple periodic cube and the RR and RRR counts for the uniform distribution can be computed analytically (see App. 2.C for analytical RRR computation).

The correlation functions change more rapidly at small separations. We require narrower bins at smaller separations in order not to lose too much information to the binning effects. To achieve this we measure $w_{\rm p}^{(2)}(r_{\rm p})$ in 12 bins equally spaced in $\log_{10} r_{\rm p}$ between 0.1 h^{-1}Mpc and 20 h^{-1}Mpc.

We use the same binning for the three sides of the $w^{(3)}_{\rm p(SV)}(r_{\rm p12}, r_{\rm p23}, r_{\rm p31})$. We arrange triplets of separations by starting with all possible unique triplets that satisfy $r_{\rm p12} \leq r_{\rm p23} \leq r_{\rm p31}$. We start with the triplet that has all three sides belonging to the shortest separation bin. We then arrange all other triplets so that each following triplet is in increasing order of $r_{\rm p12}$. Triplets that have equal $r_{\rm p12}$ we internally arrange by increasing $r_{\rm p23}$. Finally, the triplets that have both $r_{\rm p12}$ and $r_{\rm p23}$ equal we arrange by increasing $r_{\rm p31}$. We remove triplets for which the midpoints of the bins do not satisfy the triangular condition $r_{\rm p31} \leq r_{\rm p12} + r_{\rm p23}$. Once the triplets are arranged and sorted in this way, we assign to each one of them an integer "triangular index". For our choice of binning, we end up with 99 unique triangular configurations. Fig. 2.3 shows the values of $r_{\rm p12}$, $r_{\rm p23}$, $r_{\rm p31}$ as a function of the triangular index.

[1] https://github.com/dpearson1983/ganpcf

2.2.5 Covariance matrix of projected correlation functions

We use the jackknife re-sampling method to estimate the variance of clustering. We divide the simulation volume into $N_{\rm sub}$ sub-volumes. We compute the projected 2PCF and 3PCF by omitting each one of the subvolumes. This results in $N_{\rm sub}$ measurements corresponding to $(N_{\rm sub}-1)/N_{\rm sub}$ fraction of the origin volume. Covariance matrix from jackknife method is then estimated by:

$$C^{\rm jk}_{i,j} = \frac{(N_{\rm sub}-1)}{N_{\rm sub}} \sum_{k=1}^{N_{\rm sub}} (X_i^k - \bar{X}_i)(X_j^k - \bar{X}_j) \tag{2.12}$$

where X_i^k is the clustering measurements (either 2PCF or 3PCF) in ith bin from the kth jackknife realization. Overline denotes an average measurement over all realizations.

$$\bar{X}_i = \frac{1}{N_{\rm sub}} \sum_{k=1}^{N_{\rm sub}} X_i^k. \tag{2.13}$$

This version of the jackknife realization is referred to as "delete-one" version in the literature.

We set $N_{\rm sub} = 400$ to make sure we have enough sub-volumes to estimate the error of 3PCF. We slice the box Z (LOS) axis into rectangles with equal area squared base on XY plane. Fig. 2.4 shows the correlation matrices for LRGs, ELGs and QSOs measured from the jackknife method using HOD mock catalog we populated as described in Sec. 2.2.2. Each panel shows a matrix with horizontal and vertical division dash lines. The first column displays the correlation between $r_{\rm p}$ bins in the projected 2PCF with itself (bottom), and with the triangles in the projected 3PCF(top). The second column is the correlation between the 2PCF and 3PCF (bottom) and 3PCF with itself. For the correlation matrix of LRGs, there is a strong auto-correlation for the projected 2PCF at scale $r_{\rm p} > 1.4h^{-1}$Mpc (after 6th bin) and for the projected 3PCF with at least two triangle side lengths $r_{\rm p} > 1.4h^{-1}$Mpc. There is a cross-correlation between 2PCF at relatively large scale and 3PCF with at least two triangle side lengths at corresponding scale, while other cross-correlation is quite weak. The correlation matrices of ELGs and QSOs show similar pattern as LRGs, with a weaker

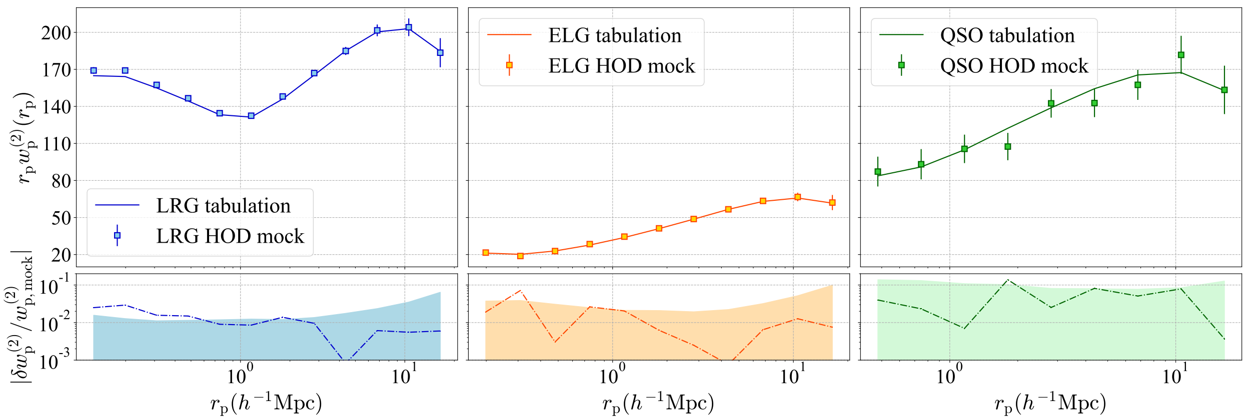


Figure 2.5: Panels on top are the projected 2PCF from HOD mock catalog and tabulation method with fiducial HOD parameters for different tracers. The blue, orange and green colors are for LRGs, ELGs, and QSOs respectively. Filled markers are measurements from the fiducial HOD mock catalog, errors are calculated using the jackknife method. Solid lines represent measurements from the tabulation method. Dot-dash lines on the bottom sub-panel are absolute value of percentage difference the light shaded area represents jackknife error.

correlation.

These covariances correspond to the constraining power of a one cubic Gigaparsec box. The actual DESI samples will cover a much larger volume. For small separations the covariances on both the 2PCF and the 3PCF will scale as an inverse of a volume.

Solid circles in Fig. 2.5 show the projected 2PCF measurements from the mock catalog for different tracer and the jackknife errorbars. The first bin of the ELG and the first three bins of the QSO projected correlation function has been omitted. For ELGs, it is a conservative choice to omit the smallest scale bin (see App.2.A). For QSOs, the number density of QSOs is too small to have a sufficient number of pairs on those small scales. Fig. 2.6 shows a similar plot for the projected 3PCF where all triangles that include the bins omitted for the 2PCF have been removed. This results in 99, 85, 60 triangular configurations for LRGs, ELGs, and QSOs respectively. We keep the original triangular indexes that have been assigned before the removal of the low separation triangles. As a result, the ELG triangular index starts with 15, and the QSO triangular index starts with 40.

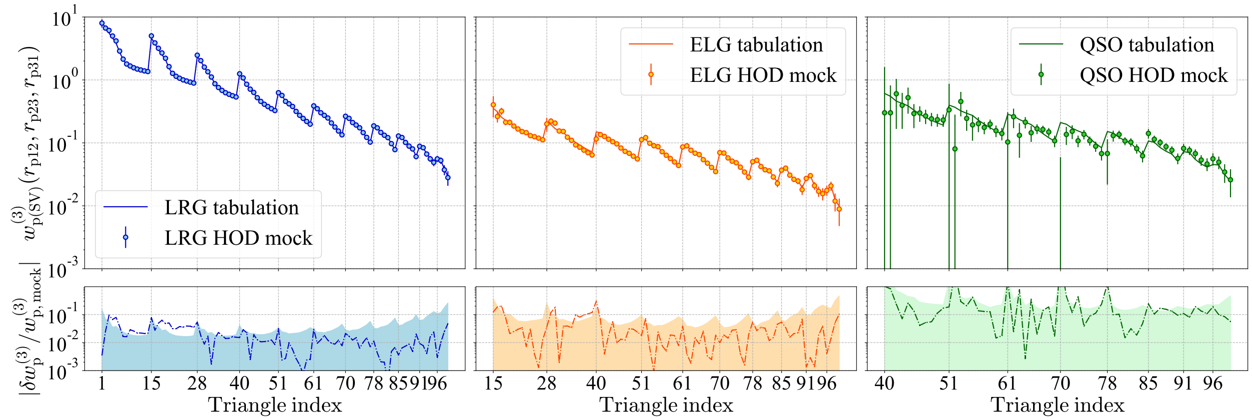


Figure 2.6: A similar plot for 3PCF. Panels on top are the simplified projected 3PCF from HOD mock catalog and tabulation method with fiducial HOD parameters for different tracers. The blue, orange, and green colors are for LRGs, ELGs, and QSOs respectively. Filled markers are measurements from the fiducial HOD mock catalog, errors are calculated using the jackknife method. Solid lines represent measurements from the tabulation method. Dot-dash lines on the bottom sub-panel are absolute value of percentage difference the light shaded area represent jackknife error.

2.2.6 Constraining HOD parameters

Not all three galaxy samples we consider can be constrained equally well with data from a one cubic Gigaparsec box. We find that for the LRGs it is possible to constrain all five parameters: $\log(M_{\mathrm{cut}})$, σ, $\log(M_1)$, $\log(M_0)$ and α (A_{c} and A_{s} are used for the tuning of the number density of the LRG sample and do not affect the 2 and 3 PCF). For ELGs, constraining all five parameters turns out to be more difficult. We only let the $\log(M_{\mathrm{cut}})$, α and A_{s} be free parameters and fix the remaining two to their fiducial values. A_{s} is degenerate in its effects with $\log(M_1)$. We choose to vary A_{s} in our computations for convenience. For QSOs, we need to further reduce the number of free parameters because the QSO sample has a much lower number density. We set $\log(M_{\mathrm{cut}})$ and $\log(M_1)$ as free parameters and fix the remaining three to their fiducial values. We apply flat priors for all free parameters. The intervals are listed in Tab. 2.2. The fiducial values for the fixed parameters are listed in Tab. 2.1.

We perform Markov Chain Monte Carlo (MCMC) to obtain the posterior probability distribution of parameter space. The likelihood function $\mathcal{L} \propto \exp -\chi^2/2$, where χ^2 is given by

$$\chi^2 = \Delta X_i (S')^{-1}_{ij} \Delta X_j, \tag{2.14}$$

where ΔX_i is the difference of binned 2PCF and 3PCF between theory and observation, in our case corresponding to tabulated estimation and HOD mock measurements. $(S')^{-1}$ is the inverse of the re-scaled covariance matrix, here we follow Percival et al.[118] to take into account the error propagation from the error in the covariance matrix into the fitting parameters.

$$S' = \frac{(n_s - 1)[1 + B(n_d - n_p)]}{n_s - n_d + n_p - 1} S \tag{2.15}$$

$$B = \frac{(n_s - n_d - 2)}{(n_s - n_d - 1)(n_s - n_d - 4)}, \tag{2.16}$$

where n_s is the number of jackknife realizations, n_d is the number of data points we are fitting to and n_p is the number of free parameters in the model. S is the original covariance matrix.

We use a modified version of CosmoMC[119] as MCMC engine ,which use the Metropolis-Hastings algorithm, to sample the parameter space and search for the minimum χ^2. We run 16 chains parallel with MPI and we ignore the first 30% of each chain as burn-in. We apply the Gelman and Rubin R statistic[120;121] as convergence criteria, all of our chains have $R - 1 < 0.01$, which represents a good convergence. Since the tabulation method provide us a fast estimation of projected 2PCF and 3PCF (see Sec.2.3), the MCMC stage takes less than half an hour to converge for a joint run using both 2PCF and 3PCF.

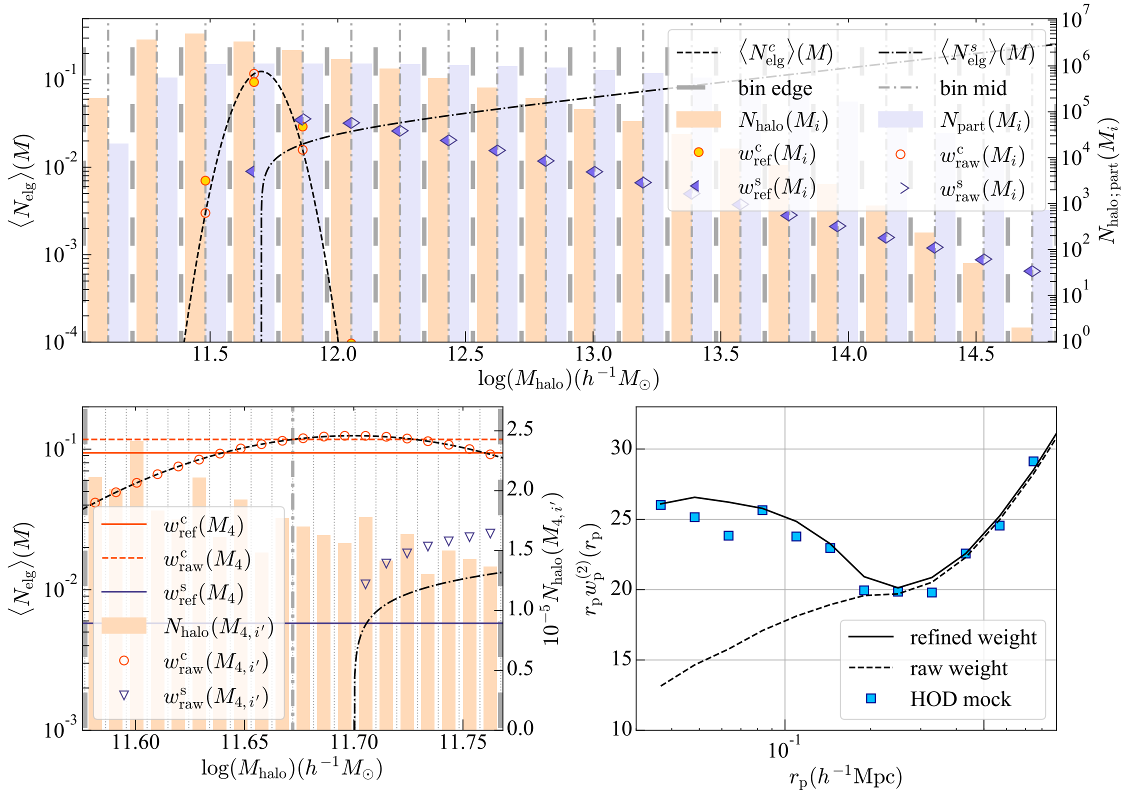


Figure 2.7: Mass binning effects for ELGs when shift satellite parameter $\log(M_0)$ slightly to 11.7 from fiducial, all other parameters remain at fiducial value. Top panel: Gray vertical lines show the edges (dashed) and the middle points (dashed-dotted) of the mass bins. The lines show the expected number of centrals (dashed) and satellites (dashed-dotted). The bars show the available number of halos (peach) and particles (lavender) in the simulation. Triangles show the weight of satellites computed with the raw probabilities (open) and refined probabilities (filled). The open and filled circles show the same information for the centrals. Bottom left panel: A zoom-in version of the 4th mass bin, where raw weights do not work. Thin dotted vertical lines show the edges of sub mass bin. The peach bars show the number of halos in each sub-mass bin. The open circles and triangles show values of raw weights for each sub mass bin. The orange horizontal dash line shows raw central weight for 4th mass bin and the orange solid line shows refined central weight for 4th mass bin. The solid purple line shows refined satellite weight for 4th mass bin and there is no dash purple line for raw satellite weight because the value is zero. Bottom right panel: Blue squares show the measured projected 2PCF of the galaxies from the ELG HOD mock catalog using HOD parameters plotted. Lines show the projected 2PCF computed with the tabulated method using the raw (dashed) and refined (solid) weights.

Parameters	LRG	ELG	QSO
$\log(M_{\mathrm{cut}})$	[12.0, 13.5]	[11.0, 13.5]	[11.5, 13.5]
σ	[0.0001, 1.0]	-	-
$\log(M_1)$	[12.5, 14.5]	-	[13.0, 17.0]
$\log(M_0)$	[11.0, 15]	-	-
α	[0.0, 2.0]	[0.0, 2.0]	-
A_{c}	-	-	-
A_{s}	-	[0.0, 0.2]	-

Table 2.2: The flat prior interval on HOD parameter for different tracer, fitting to LRG, ELG, and QSO HOD mock catalog have 5, 3, and 2 free parameters respectively. Parameters with a dash are fixed to their fiducial values.

2.3 Tabulation method of computing galaxy N-point Correlation Functions

2.3.1 Tabulated 2PCF

The two steps leading to the pair counts of the mock catalog — populating N-body mocks with galaxies and counting pairs of galaxies — can be formally summarized with the equation

$$DD(r_p) = \sum_{ij} \Theta_{ij}^{r_p} D_i D_j, \tag{2.17}$$

where each index in the double summation goes over all halo centers and halo particles,

$$\Theta_{ij}^{r_p} = \begin{cases} 1, & \text{if distance between (i,j) pair falls within the specified } r_p \text{ bin} \\ 0, & \text{otherwise} \end{cases} \tag{2.18}$$

and D is a stochastic variable

$$D_i = \begin{cases} 1, & \text{if the } i^{\text{th}} \text{ halo/particle got populated by a galaxy} \\ 0, & \text{otherwise} \end{cases} \tag{2.19}$$

We rewrite eq. (2.17) as

$$DD = \sum_{ij}^{N_h} \Theta_{ij}^{hh} D_i^h D_j^h + 2 \sum_{i,j}^{N_h,N_p} \Theta_{ij}^{hp} D_i^h D_j^p + \sum_{ij}^{N_p} \Theta_{ij}^{pp} D_i^p D_j^p, \tag{2.20}$$

explicitly separating halos and particles, where superscripts h and p refer to the halos and particles respectively, N_h and N_p are the numbers of halos and particles, and we dropped the r_p label for brevity. In the traditional approach (top panel of Fig. 2.1) the random numbers D have to be drawn for and the double sum over all occupied halos and particles computed for every HOD model under consideration.

The tabulation approach reduces the complexity of this computation by employing the following trick. The expectation value of the pair count is

$$\langle DD \rangle = \sum_{ij}^{N_h} \Theta_{ij}^{hh} \lambda_i^c \lambda_j^c + 2 \sum_{i,j}^{N_h,N_p} \Theta_{ij}^{hp} \lambda_i^c \lambda_j^s + \sum_{ij}^{N_p} \Theta_{ij}^{pp} \lambda_i^s \lambda_j^s, \tag{2.21}$$

where λ^c and λ^s are the expected values of that particular halo or a particle to host a central or satellite galaxy in a given HOD model. Since these numbers only depend on the mass of the host halo, we can simplify the computation by binning the halos and particles into bins of mass in log space narrow enough so that the expected values do not significantly change within it. The pair count can then be rewritten as

$$\langle DD \rangle = \sum_{i,j}^{N_h^k,N_h^\ell} \sum_{k\ell}^{N_b} \Theta_{ij,k\ell}^{hh} \bar{\lambda}_k^c \bar{\lambda}_\ell^c + 2 \sum_{i,j}^{N_h^k,N_p^\ell} \sum_{k\ell}^{N_b} \Theta_{ij,k\ell}^{hp} \bar{\lambda}_k^c \bar{\lambda}_\ell^s + \sum_{i,j}^{N_p^k,N_p^\ell} \sum_{k\ell}^{N_b} \Theta_{ij,k\ell}^{pp} \bar{\lambda}_k^s \bar{\lambda}_\ell^s, \tag{2.22}$$

where the indices k and ℓ now go over N_b number of mass bins, N_h^k is the number of halos in the k^{th} mass bin (similarly for particles) and $\bar{\lambda}^c$ and $\bar{\lambda}^s$ are effective average expected values in each mass bin,

$$\bar{\lambda}_k^c = \langle N^c \rangle (M_k) \tag{2.23}$$

$$\bar{\lambda}_k^s = \langle N^s \rangle (M_k) \frac{N_h^k}{N_p^k}. \tag{2.24}$$

M_k is the representative mass in kth mass bin $\log M_k \pm (\Delta \log M)/2$, where $\Delta \log M$ is the width of log space mass bin. $\langle N^c \rangle (M_k)$ and $\langle N^s \rangle (M_k)$ are expectation number of central and satellite galaxies hosted by halo in kth mass bin. $\bar{\lambda}_k^s$ could be understood this way, $\langle N^s \rangle (M_k) N_h^k$ is the expected total amount of satellite galaxies in kth mass bin, divided by the N_p^k would give us the average expected values for each particle to host a satellite galaxy.

Switching the order of summation we get

$$\langle DD \rangle = \sum_{k\ell}^{N_b} \bar{\lambda}_k^c \bar{\lambda}_\ell^c DD_{k\ell}^{hh} + 2 \sum_{k\ell}^{N_b} \bar{\lambda}_k^c \bar{\lambda}_\ell^s DD_{k\ell}^{hp} + \sum_{k\ell}^{N_b} \bar{\lambda}_k^s \bar{\lambda}_\ell^s DD_{k\ell}^{pp}, \tag{2.25}$$

where

$$DD_{k\ell}^{hh} = \sum_{i,j}^{N_p^k, N_p^\ell} \Theta_{ij,k\ell}^{hh} \tag{2.26}$$

is the number of halo pairs with a separation that falls in the specified bin, where one member of the pair is in mass bin k while another is in mass bin ℓ (similarly for the particle-halo and particle-particle pairs).

Eq. (2.25) is equivalent to eq. (2.21) but has two advantages. First, it gives an average value of the number count expected for a given HOD model instead of a specific realization that includes stochastic noise. Secondly, it has the potential to save a significant amount of computational time. The most time-consuming part of the computation - the double sum over halos and particles - can be performed only once. Changing the HOD model amounts to simply summing up precomputed pair counts with different weights, a procedure that is orders of magnitude faster.

This method was used by[83] to estimate 2PCF from N-body simulations efficiently. The approach we introduced above is more like section 2.2 in Zheng and Guo[83], Case with

Subhaloes, instead of subhaloes, we populate satellite with particles here. This method could be applied to different kinds of galaxy clustering, e.g. real space 2PCF, projected 2PCF, and 2PCF multipole. The correlation function is given by a similar weighted sum over different mass bin cross-correlations and we take the projected correlation function as an example to show in detail.

Halos and particles live in the same periodic box so the RR counts are identical for them. This means that the projected 2PCF is also a weighted average of the cross-2PCF of different mass halos (and particles)

$$\begin{aligned} w^{(2)}_{\mathrm{p,gg}}(r_{\mathrm{p}}) = &\sum_{k,\ell}^{N_b} w^c(M_k) w^c(M_\ell) w^{(2)}_{\mathrm{p,hh}}(r_{\mathrm{p}}, M_k, M_\ell) \\ &+ 2\sum_{k,\ell}^{N_b} w^c(M_k) w^s(M_\ell) w^{(2)}_{\mathrm{p,hp}}(r_{\mathrm{p}}, M_k, M_\ell) \\ &+ \sum_{k,\ell}^{N_b} w^s(M_k) w^s(M_\ell) w^{(2)}_{\mathrm{p,pp}}(r_{\mathrm{p}}, M_k, M_\ell), \end{aligned} \tag{2.27}$$

where $w^{(2)}_{\mathrm{p,hh}}(r_{\mathrm{p}}, M_i, M_j)$ is the two-point cross correlation function of halos in the ith and jth mass bins (similarly for the halo-particle and particle-particle correlation functions) and naively we could take the weight as

$$w^c_{\mathrm{raw}}(M_k) = \bar{\lambda}^c_k = \langle N^c \rangle (M_k), \tag{2.28}$$

$$w^s_{\mathrm{raw}}(M_k) = \bar{\lambda}^s_k = \langle N^s \rangle (M_k) \frac{N^k_h}{N^k_p}. \tag{2.29}$$

We define Eqn. (2.28) & (2.29) as raw weights, raw weights are a good approximation in most cases but have some exceptions, we will further explain this in Sec. 2.3.3.

Solid lines on the top panels of Fig. 2.5 show the galaxy projected 2PCF computed using the tabulation method. The bottom panels show the fractional deviation between the projected 2PCF computed with the tabulated method and a specific realization. The offset is in all cases within the expected standard deviation. The deviations are caused by the stochasticity in the realization, the tabulated method being almost noise-free. There is a very small stochastic noise in the tabulated 2PCF related to the finite number of halos and particles in the box, but it is negligible compared to the noise in a single realization.

2.3.2 Tabulated 3PCF

We further generalize the derivation of the previous section to the 3PCF. Similar arguments lead to the expression

$$\begin{aligned} w^{(3)}_{\rm p,ggg}(\triangle) = &\sum_{i,j,k} w^c(M_i)w^c(M_j)w^c(M_k)w^{(3)}_{\rm p,hhh}(\triangle, M_i, M_j, M_k) \\ &+ 3\sum_{i,j,k} w^c(M_i)w^s(M_j)w^s(M_k)w^{(3)}_{\rm p,hpp}(\triangle, M_i, M_j, M_k) \\ &+ 3\sum_{i,j,k} w^s(M_i)w^c(M_j)w^c(M_k)w^{(3)}_{\rm p,phh}(\triangle, M_i, M_j, M_k) \\ &+ \sum_{i,j,k} w^s(M_i)w^s(M_j)w^s(M_k)w^{(3)}_{\rm p,ppp}(\triangle, M_i, M_j, M_k). \end{aligned} \tag{2.30}$$

Here $r_{\rm p12}, r_{\rm p23}, r_{\rm p31}$ are abbreviated as $\triangle$. Solid lines on top panels of Fig. 2.6 show the galaxy projected 3PCF computed using the tabulation method. Similarly to the 2PCF the measurement from a single realization is noisier but consistent within expected errors for all tracers.

2.3.3 Mass binning effects

We bin the mass of the host halo in 20 bins between around $11 < \log M_\star < 14.8$ (see App.2.A for details about downsampling). The bins are narrow enough so that the hosting probabilities do not change significantly within the bins and assigning to each halo and particle a probability at the middle of the bin (we referred to this practice as a raw probability) is in most cases a good approximation. However, there are some exceptions if we shift satellite parameter $\log(M_0)$ of ELG just a little bit to 11.7, where hosting probability drops very steeply below $\log(M_{\rm halo}) \sim 11.7$, the middle of bin failed to catch satellite information. Fig. 2.7 demonstrates the nature of this problem.

The black dash and dot-dash line on the top panel of Fig. 2.7 show the expected number of the central and satellite ELGs in our fiducial HOD model. Bold gray vertical dash line and gray vertical dot-dash lines show the edges and the middle of the mass bin. The peach and lavender histogram shows the number of halos and particles in each mass bin. Empty triangles pointing to the right show the raw weights based on the value in the middle of the mass bin and the filled triangle pointing to the left shows the refined weights. For most of the mass range, the two are very consistent. The last nonzero bin on the left (4th bin) is the exception. The mean number drops so steeply with the mass that it reaches an extremely low number for the middle mass of that bin. If we applied weight based on that value very few of the halos in that mass bin would acquire a satellite. This would incorrectly down-weight the halos close to the right edge of the mass bin that has a substantial probability of hosting a satellite.

Increasing the number of mass bins is however impractical as it would significantly increase the number of separate cross-correlation functions that we need to keep track of. However, this problem would go away if we used finer binning for the mass of the host halo. We modify our probabilities as

$$w^c_{\rm ref}(M_k) = \sum_{k'}^{N_{\rm sub}} \frac{N_h^{k,k'}}{N_h^k} w^c_{\rm raw}(M_{k,k'}) = \sum_{k'}^{N_{\rm sub}} \frac{N_h^{k,k'}}{N_h^k} \langle N^c \rangle (M_{k,k'}) \tag{2.31}$$

$$w^s_{\rm ref}(M_k) = \sum_{k'}^{N_{\rm sub}} \frac{N_h^{k,k'}}{N_h^k} w^s_{\rm raw}(M_{k,k'}) = \sum_{k'}^{N_{\rm sub}} \frac{N_h^{k,k'}}{N_p^k} \langle N^s \rangle (M_{k,k'}) \tag{2.32}$$

As shown in lower left panel on Fig. 2.7, we further subdivided the mass bin into $N_{\rm sub} = 20$ sub-bins, and take a weighted average of raw weights for each sub mass bin $w_{\rm raw}(M_{k,k'})$ based on the number of halos $N_h^{k,k'}$ in this sub mass bin. The dashed line on this panel shows the raw weight of this mass bin before correction, the solid line shows the refined weight. For central weight, the difference is subtle. For satellite weight, refined weight accurately accounts for the contribution of satellites from this mass bin while raw weight failed to capture a number accurately.

The lower right panel of Fig. 2.7 compares the tabulated projected 2PCF computed with the raw weights and the refined weights. The raw weights clearly fail to describe the 2PCF on the scales of $r_p < 0.1$ where the systematic offsets are more than 50% of the signal.

This problem will appear whenever there are variations in the expected number of galaxies (either satellites or centrals) across the width of the bin. When assigning all galaxies in the bin the weight at the middle of the bin we are making an approximation

$$\int_{M_{\min}}^{M_{\max}} \mathrm{d}M\, w(M) \frac{dN}{dM} \approx w(\overline{M}) N^{\mathrm{tot}}, \tag{2.33}$$

where $\overline{M}$ is the middle of the bin, ΔM is the width of the bin, the integration is between $\overline{M} - \Delta M/2$ and $\overline{M} + \Delta M/2$, and the N^{tot} is the total number of galaxies in that bin. To see when this approximation may fail we can expand the weight function on the left side of the equation in Taylor series around the middle point as

$$w(M) = w(\overline{M}) + \frac{dw}{dM}\Big|_{M=\overline{M}}(M - \overline{M}) + \frac{1}{2}\frac{d^2w}{dM^2}\Big|_{M=\overline{M}}(M - \overline{M})^2 + \mathcal{O}[(M - \overline{M})^3]. \tag{2.34}$$

The integral over the first term results in $w(\overline{M})N^{\mathrm{tot}}$, which matches exactly with our approximation. The second term integrates to zero (as an odd function over symmetric limits). The third term integrates to $w''(\overline{M})(\Delta M)^3 N^{\mathrm{tot}}/24$, here the double prime denotes the second derivative with respect to halo mass, ΔM is the width of the bin.[2] This is the leading error term and it needs to be small for the approximation to be valid. The condition is

$$\frac{w''(\overline{M})}{w(\overline{M})} \ll \frac{24}{(\Delta M)^3} \tag{2.35}$$

If this condition between the second derivative of the weights and the width of the mass bin starts breaking down the weight refining procedure described in this section may be in order. This condition is obviously broken for our HOD models since they contain a

[2] We are assuming that the fundamental quantities such as the number of halos and the bias of halos do not change significantly within the bin. If this is not the case, the binning is obviously too broad and needs to be refined.

Tracer	LRG			ELG			QSO		
Data	$w_{\rm p}^{(2)}$	$w_{\rm p(SV)}^{(3)}$	$w_{\rm p}^{(2)}+w_{\rm p(SV)}^{(3)}$	$w_{\rm p}^{(2)}$	$w_{\rm p(SV)}^{(3)}$	$w_{\rm p}^{(2)}+w_{\rm p(SV)}^{(3)}$	$w_{\rm p}^{(2)}$	$w_{\rm p(SV)}^{(3)}$	$w_{\rm p}^{(2)}+w_{\rm p(SV)}^{(3)}$
$\log M_{\rm cut}$	12.88 ± 0.199	12.73 ± 0.058	12.73 ± 0.059	11.83 ± 0.059	11.74 ± 0.125	11.83 ± 0.054	12.47 ± 0.060	12.43 ± 0.130	12.48 ± 0.058
σ	0.315 ± 0.200	0.151 ± 0.103	0.162 ± 0.105	-	-	-	-	-	-
$\log M_1$	13.93 ± 0.141	13.83 ± 0.053	13.82 ± 0.047	-	-	-	15.49 ± 0.766	15.53 ± 0.851	15.53 ± 0.755
$\log M_0$	11.73 ± 0.431	11.76 ± 0.407	11.78 ± 0.400	-	-	-	-	-	-
α	1.279 ± 0.055	1.300 ± 0.040	1.301 ± 0.038	0.188 ± 0.097	0.268 ± 0.161	0.178 ± 0.095	-	-	-
$A_{\rm c}$	-	-	-	-	-	-	-	-	-
$A_{\rm s}$	-	-	-	0.015 ± 0.007	0.023 ± 0.016	0.015 ± 0.006	-	-	-
χ^2/d.o.f	2.5/ (12 − 5)	36.5/ (99 − 5)	34.9/ (111 − 5)	0.9/ (11 − 3)	57.5/ (85 − 3)	56.7/ (96 − 3)	4.3/ (9 − 2)	39.4/ (60 − 2)	42.6/ (69 − 2)

Table 2.3: The results for the fits to HOD mock catalog of each tracer with three different data sets: 2PCF only ($w_{\rm p}^{(2)}$), 3PCF only ($w_{\rm p(SV)}^{(3)}$) and joint ($w_{\rm p}^{(2)}+w_{\rm p(SV)}^{(3)}$). We show the mean $\pm\,1\sigma$ error for floating HOD parameters and χ^2/d.o.f for each fit. The fiducial (true) value for each parameter is listed in Tab. 2.1

discontinuous function in the satellite probability in eqs. (2.2) and (2.4).

2.4 Results

We create DESI like LRG, ELG, and QSO samples as described in Sec. 2.2.2. We then use the MCMC method to fit the HOD parameters for the model described in Sec. 2.2.6 with the covariance matrix obtained as described in Sec. 2.2.4. The covariance matrix represents the variance in the measurements expected from a cosmic volume of 1 cubic gigaparsec. The actual DESI measurements will be obtained from larger volumes, but since the errors on both the 2PCF and the 3PCF scale are similar to the volume the relative strength of the constraints coming from the two will not change.

Fig. 2.8 shows 1 σ uncertainty band of LRG sample HOD function from 2PCF only fitting and 2PCF+3PCF joint fitting. The light blue band shows 68% CL uncertainty from 2PCF only and the dark blue band shows the band from 2PCF+3PCF joint fitting. The orange line represents the fiducial HOD setting as the truth behind the mock we fit to. It is clear to

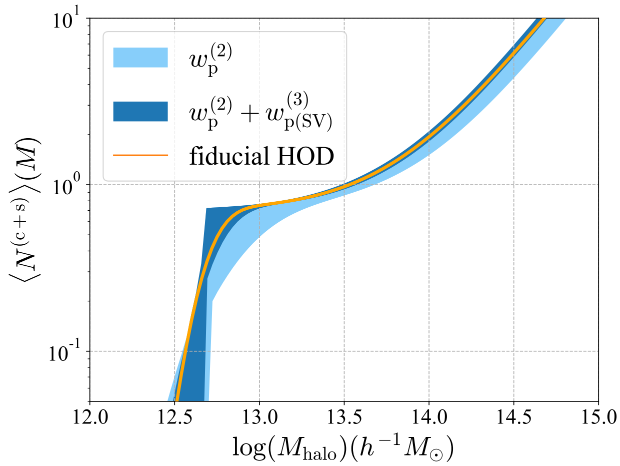


Figure 2.8: 1-σ band of LRG sample HOD. The light blue is the 68% CL uncertainty from projected 2PCF only, the dark blue band is the 68% CL uncertainty from the joint fitting of projected 2PCF and simplified version projected 3PCF. The orange line is the fiducial HOD of the LRG sample.

see joint fitting has a much narrow band compared to the one using 2PCF only, especially for the range $\log(M_{\mathrm{halo}} > 12.7)$, indicating a much better constraint on satellite parameters from joint fitting. Fiducial HOD lies in the 1σ band showing a good recovery for both cases.

Fig. 2.9 shows 1 and 2 σ confidence level contours on the HOD parameters for the LRG sample. These constraints are dominated by the $w_{\mathrm{p(SV)}}^{(3)}$. The improvement is especially large for the parameters $\log M_{\mathrm{cut}}$, σ, and $\log M_1$. The 3PCF constraints on those parameters improve by 70, 49, and 62 percent respectively compared to the 2PCF results. Combined fitting does not significantly differ from the 3PCF only results. Tab. 2.3 summarizes the marginalized statistic for each fit. From the 1D distribution of each parameter on Fig. 2.9, all cases successfully recover the fiducial HOD parameters.

Fig. 2.12 and 2.13 show 1 and 2 σ confidence level contours for the ELG at redshift 1.1 and 0.8 and QSO samples at redshift 1.4 respectively. We only free HOD parameters as shown in the contours for these tracers. For the ELG and QSO, the constraints are dominated by the projected 2PCF. Improvements offered by the addition of the projected 3PCF are negligible.

There could be several reasons why the LRGs benefit greatly from the addition of the 3PCF information while ELGs and QSOs do not. One potential explanation is that the ELGs and QSOs are at higher redshifts where matter underwent less nonlinear evolution and the

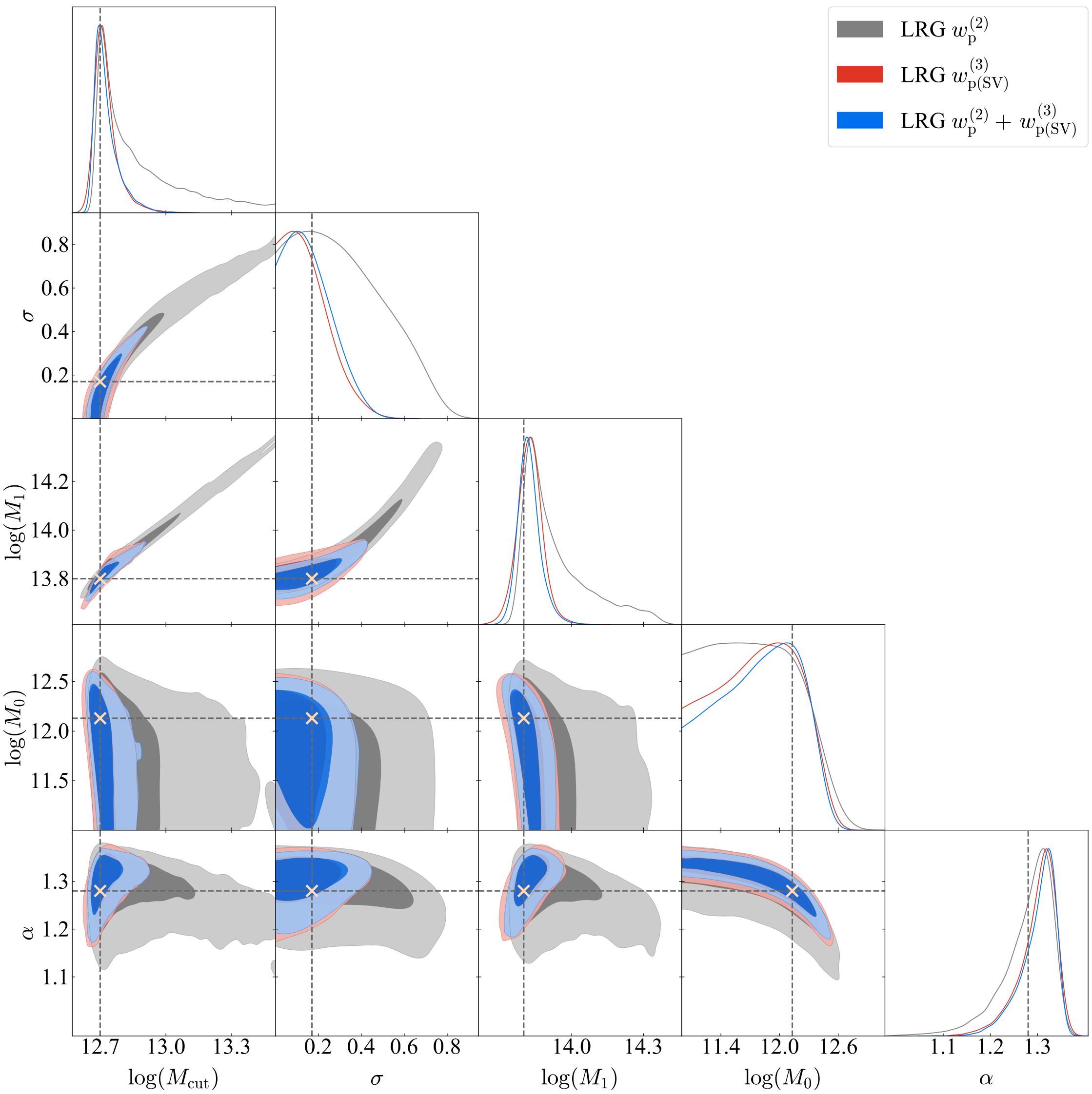


Figure 2.9: Marginalized probability distribution of HOD parameters for DESI like LRG sample at $z = 0.8$. The results from the projected 2PCF and 3PCF are shown in grey and red respectively. Blue shows the joint constraints of the two. The contours represent 68 and 95 percent confidence levels. 1D marginalized distribution for each parameter is shown on top of each column. The dashed line and light yellow cross markers show fiducial HOD parameter values.

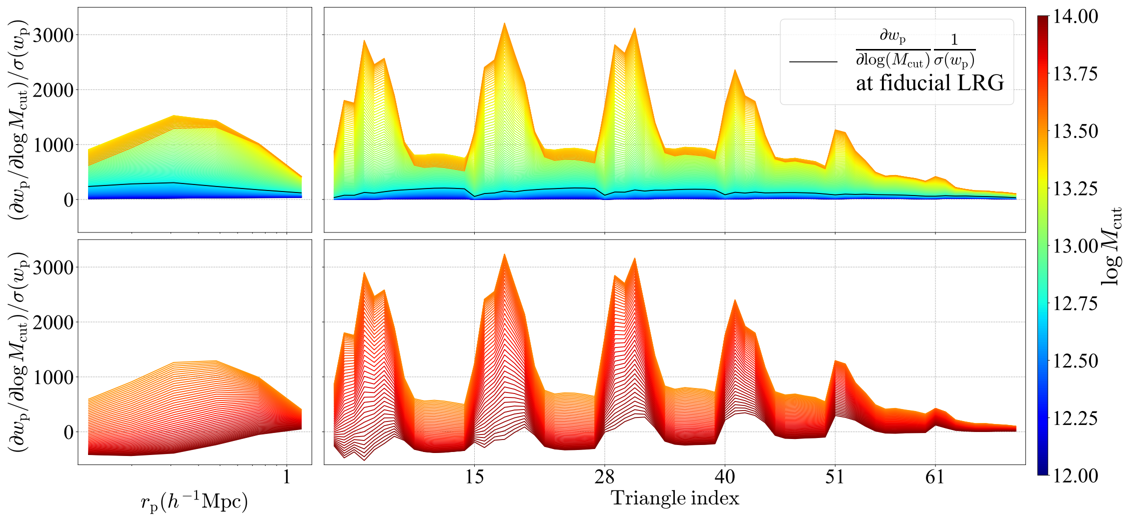


Figure 2.10: Partial derivative of w_p with respect to $\log(M_{cut})$ normalized using error of w_p when fixing other HOD parameters. Plot has been separated to two panels to avoid overlap when partial derivative drop down.

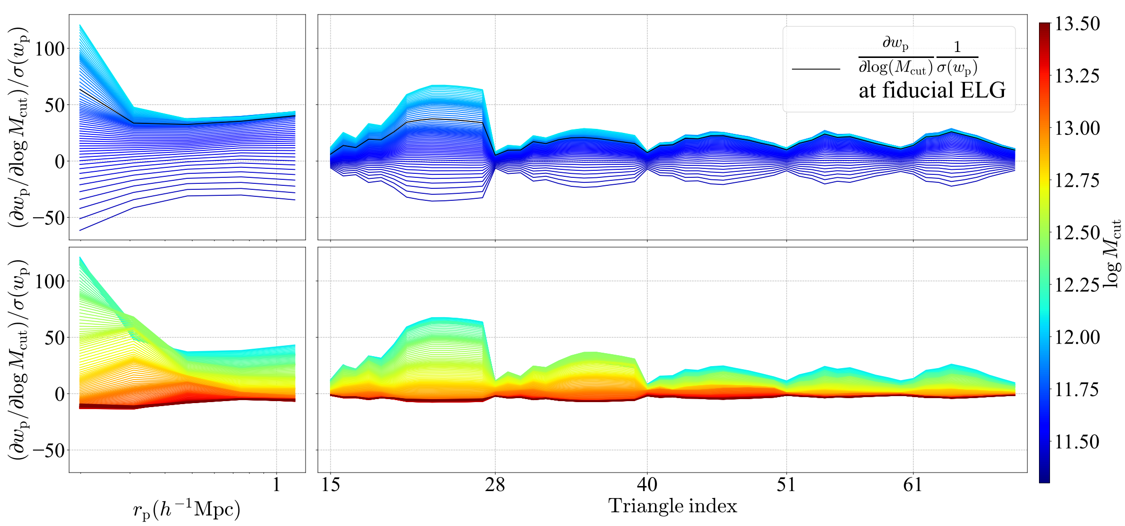


Figure 2.11: Similar plot as Fig. 2.10 for ELG sample at fiducial redshift $z = 1.1$

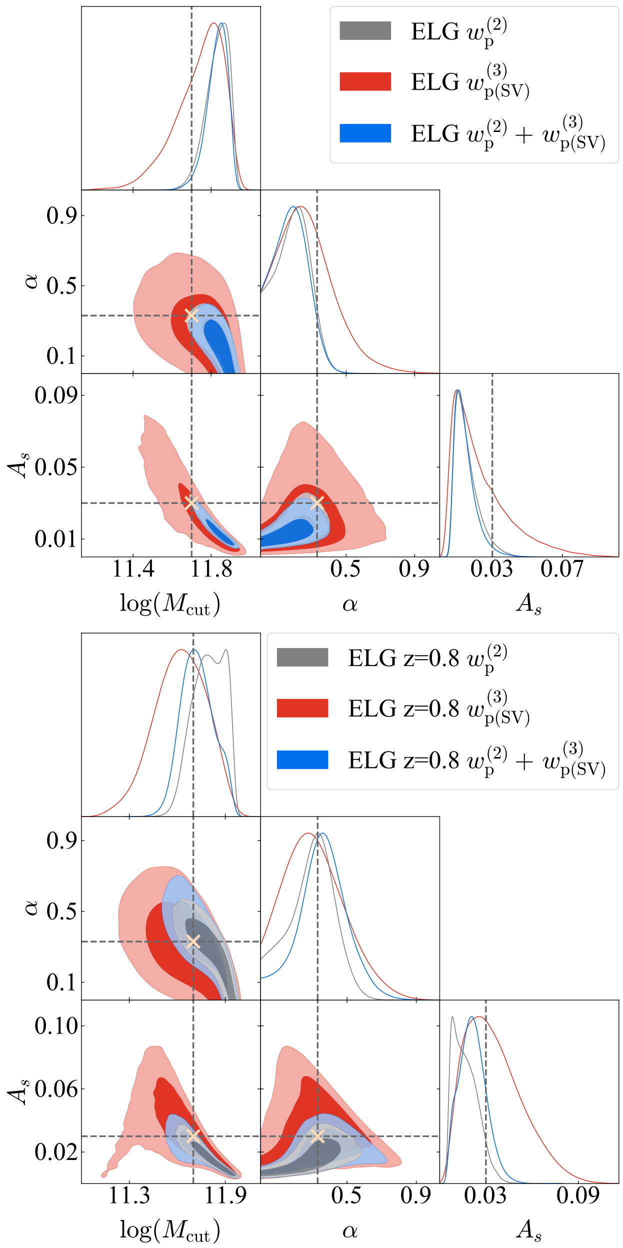


Figure 2.12: Marginalized probability distribution of selected HOD parameters for DESI like ELG sample at $z = 1.1, 0.8$. The results from the projected 2PCF and 3PCF are shown in grey and red respectively. Blue shows the joint constraints between the two. The contours represent 68 and 95 percent confidence levels. 1D marginalized distribution for each parameter is shown on top of each column. The dashed line and light yellow cross markers shows fiducial HOD parameter values.

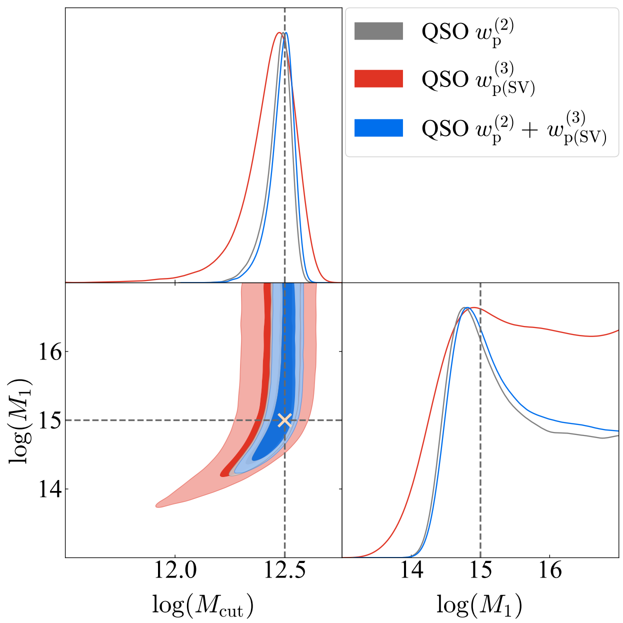


Figure 2.13: Marginalized probability distribution of selected HOD parameters for DESI like QSO sample at $z = 1.4$. The results from the projected 2PCF and 3PCF are shown in grey and red respectively. Blue shows the joint constraints between the two. The contours represent 68 and 95 percent confidence levels. 1D marginalized distribution for each parameter is shown on top of each column. The dashed line and light yellow cross markers shows fiducial HOD parameter values.

three-point signal is not as pronounced. Another potential explanation is that galaxies of different host halo masses are not equally sensitive to the three-point information see e.g.[68].

To study the sensitivity of 2PCF and 3PCF to HOD parameters at different fiducial values we make a plot of the partial derivative of $w_{\mathrm{p}}^{(2)}$ and $w_{\mathrm{p(SV)}}^{(3)}$ with respect to $\log(M_{\mathrm{cut}})$ normalized to the variance in the measurement at the fiducial value. Fig. 2.10 shows partial derivatives of the 2PCF and the 3PCF with respect to $\log(M_{\mathrm{cut}})$ with other parameters fixed to their fiducial value. To make the plot more readable we separate it into two parts. The top panel covers the range $12 < \log(M_{\mathrm{cut}}) < 13.5$ while the bottom panel covers $13.5 < \log(M_{\mathrm{cut}}) < 14$. High values of this derivative mean that the measurement at that specific bin is highly sensitive to small changes in M_{cut}

The derivative of $w_{\mathrm{p}}^{(2)}$ reaches highest value at $\log(M_{\mathrm{cut}}) = 13.28$ then drops back, while derivative of $w_{\mathrm{p(SV)}}^{(3)}$ keeps increasing up until 13.5 and only then drops down. The 3PCF displays a larger cumulative sensitivity in the range $\log(M_{\mathrm{cut}}) > 13.16$, below that range the 3PCF is not as sensitive to small changes in M_{cut} compare to $w_{\mathrm{p}}^{(2)}$.

Another thing apparent from the figure is that the small scale triangles are more sensitive to $\log(M_{\mathrm{cut}})$ compared to their large-scale counterparts (as evident by the local peaks in the right panel). Triangles with all side lengths within the first 6 bins (side lengths $r_{\mathrm{p}} < 1.41h^{-1}$Mpc, corresponding to triangle indices 1-8, 15-21, 28-33, 40-43, 51-53, 61), peak at $\log(M_{\mathrm{cut}}) = 13.5$, while other triangles behave just like $w_{\mathrm{p}}^{(2)}$, drop back at $\log(M_{\mathrm{cut}}) = 13.28$. The different behavior of small scale triangles and small scale pairs leads to a higher sensitivity to HOD parameter changes for small scale $w_{\mathrm{pSV}}^{(3)}$. The normalized derivative of $w_{\mathrm{p(SV)}}^{(3)}$ hit around 3000 while $w_{\mathrm{p}}^{(2)}$ remains at 1500.

Fig. 2.11 shows the similar plots for ELG sample at $z = 1.1$. The sensitivity in both the 2PCF and the 3PCF increases in the range of $11.3 < \log(M_{\mathrm{cut}}) < 11.98$ and then drops in the range of $11.98 < \log(M_{\mathrm{cut}}) < 13.5$. At this redshift, the top sensitivity is achieved at the values of around $\log(M_{\mathrm{cut}}) = 11.98$. The sensitivity of $w_{\mathrm{p}}^{(2)}$ at the top is higher than the sensitivity of the $w_{\mathrm{p(SV)}}^{(3)}$. For ELGs, means a lower sensitivity for 3PCF. The cumulative sensitivity at the peak is also larger for the 2PCF compared to the 3PCF. Small scale triplets do not show the same behavior as the LRG sample, remaining at low sensitivity compare to small scale pairs.

These two plots show that both the redshift and the typical halo mass are responsible for the difference between the LRG and the ELG cases. The DESI LRGs happen to be in the halo mass range where the 3PCF is more sensitive to the HOD parameters, while ELGs are in halos with the opposite property. This is the main reason why the improvement in our ELG constraints is modest while the improvement in the LRG constraints is significant.

Plots similar to the ones presented in Figs. 2.10 and 2.11 can be used to determine whether the 3PCF is expected to affect the overall HOD constraints in a meaningful way. For example, one could populate the simulation with the best fit HOD parameters resulting from the 2PCF fits and do the same thing with slightly offset values of HOD parameters. Derivatives of 3PCF with respect to HOD parameters can be estimated by computing the 3PCF for these mocks (using finite differences). The covariance matrix (standard error) can be computed using the jackknife method. Unlike MCMC chains, which typically require order of 10,000 computations, this can be done with just a handful 3PCF computations (a few for the numerical derivatives and an order of 100 3PCF computations for the jackknife covariance) and can be done without preparing the tabulated triplet counts. A comparison of these integrated sensitivities of the 3PCF and the 2PCF can then be used to decide whether the additional investment of computational resources for the computation of the tabulated triplet counts is warranted.

To test the pure redshift dependence of the constraining power of the 3PCF we populate our ELG mock catalog at redshift 0.8 with the same HOD parameters (tuned to the same number density, i.e. the ratio of A_c and A_s remain unchanged). The results are presented in the bottom panel of Fig. 2.12. We do not find a significant change in the overall picture. The constraints are still dominated by the 2PCF signal. We do notice however that the addition of the 3PCF makes the likelihood contours more Gaussian and moves the most likely values closer to the true values somewhat debiasing the results.

2.5 Conclusion

We studied the performance of projected 3PCFs in constraining HOD parameters for different mock galaxy samples, which were based on AbacusSummit simulation, targeted by DESI. We generalized the tabulation method to 3PCF computations to make a fast evaluation of the posterior likelihood possible.

We find that the constraints on the basic HOD parameters of the mock LRG sample with input HOD parameter as shown in Tab. 2.1 at redshift $z \sim 0.8$ can be significantly improved by the addition of the 3PCF. The constraints on some parameters have improved by as much as 70 %. For the characteristic minimum mass of the central LRGs we get the constraints $\log(M_{\rm cut}) = 12.88 \pm 0.199$ with the 2PCF and $\log(M_{\rm cut}) = 12.73 \pm 0.058$ with the 3PCF. For the threshold mass of the satellite LRGs we get the constraints $\log(M_1) = 13.93 \pm 0.141$ with the 2PCF and $\log(M_1) = 13.83 \pm 0.053$ with the 3PCF. All at 1σ confidence level.

We also find that the additional constraining power offered by the 3PCF depends on the redshift of the galaxy sample as well as the typical halo mass that its galaxies occupy. The relative strength of the 3PCF increases at lower redshifts. 3PCF is also a more sensitive measurement for the samples that incorporate more massive halos. The ELG samples of DESI are at higher redshifts and occupy less massive halos. This results in the 3PCF not being as efficient in constraining their host halo mass ranges. For the mock ELG sample with input HOD parameter as shown in Tab. 2.1 at redshift $z \sim 1.1$ the constraints of the characteristic minimum mass of the central are $\log(M_{\rm cut}) = 11.83 \pm 0.059$ with the 2PCF and $\log(M_{\rm cut}) = 11.74 \pm 0.125$ with the 3PCF. For the mock QSOs with lower number density compare to the other tracers and even higher redshift $z \sim 1.4$, we get $\log(M_{\rm cut}) = 12.47 \pm 0.060$ with the 2PCF and $\log(M_{\rm cut}) = 12.43 \pm 0.130$ with the 3PCF, 2PCF remaining dominates.

Acknowledgements

We would like to thank Gongbo Zhao, Shun Satio, Hee Jong Seo, Andrew Hearin, Francisco Villaescusa-Navarro, Ashley J. Ross, and Lehman Garrison for their helpful discussion. LS is grateful for support from DOE grants DE-SC0021165 and DE-SC0011840, NASA ROSES grants 12-EUCLID12-0004 and 15-WFIRST15-0008, and Shota Rustaveli National Science Foundation of Georgia grants FR 19-498.

This research is supported by the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy under Contract No. DE–AC02–05CH11231, and by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract; additional support for DESI is provided by the U.S. National Science Foundation, Division of Astronomical Sciences under Contract No. AST-0950945 to the NSF's National Optical-Infrared Astronomy Research Laboratory; the Science and Technologies Facilities Council of the United Kingdom; the Gordon and Betty Moore Foundation; the Heising-Simons Foundation; the French Alternative Energies and Atomic Energy Commission (CEA); the National Council of Science and Technology of Mexico; the Ministry of Economy of Spain, and by the DESI Member Institutions.

We acknowledge the use of the NASA astrophysics data system https://ui.adsabs.harvard.edu/ and the arXiv open-access repository https://arxiv.org/. The software was hosted on the GitHub platform https://github.com/. The manuscript was typeset using the overleaf cloud-based LaTeX editor https://www.overleaf.com.

Data Availability

The data product related to this study, including tabulated 2 point and 3 point counts, HOD mock catalogs, jackknife covariance matrices, and MCMC chains, is available at https://doi.org/10.5281/zenodo.6380446.

The AbacusSummit simulations used in this study are publicly available (https://abacusnbody.org/).

Appendix

2.A Downsampling

The number of halos in simulations increases rapidly towards the lower mass range. However, the contribution of low mass halos to the 2PCF and 3PCF is negligible for the hod parameter range we are interested in (see Fig. 2.2).

When populating mocks, we set a hard cut-off mass for halo samples at the lower mass end. We remove all halos with a mass less than $10^{11}h^{-1}M_{\odot}$.

For the tabulation method, to speed up pair and triangle counting when preparing tables, we downsample the lower mass halos using the filter,

$$\mathrm{frac}_{\mathrm{halos}}(M_{\mathrm{halo}}) = \frac{1}{1 + 10\exp\left(-25\log(M_{\mathrm{halo}}) - 11.2)\right)} \tag{2.36}$$

from the AbacusHOD package[63]. Fig. 2.A.1 shows the halo sample fraction as a function of halo mass following Eqn. 2.36. This filter will further remove excess halos that have trivial contribution to the clustering in low mass bins.

For the particles in table preparation process, we randomly select 0.15% of particles making each halo for the pair and triangle counting. This small percentage is still larger than the mean satellite number for each halo.

We explicitly checked that the clustering from the tabulation method, which applied the filter, could perfectly match the clustering from mocks, which applied only a hard cut-off, around the fiducial HOD parameter area we are interested in (see Fig. 2.5 and 2.6).

ELGs have a lower typical host halo mass compared to other samples and they could in principle be affected by the filter. We omit the first separation bin for the ELGs to protect against this eventuality for some extreme HOD parameter values.

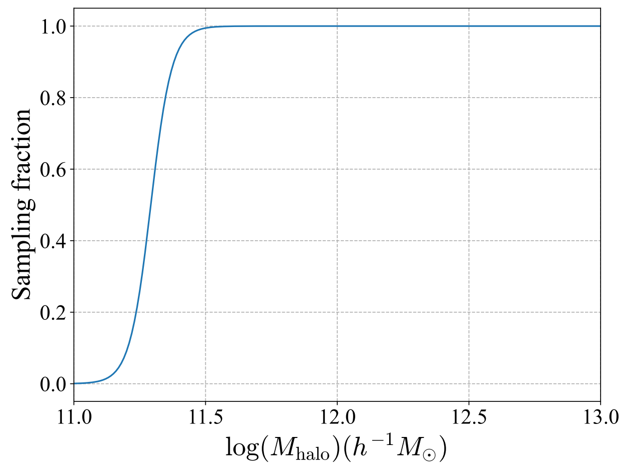


Figure 2.A.1: Fraction of halos we take from all halo as a function of halo mass.

2.B Choice of maximum radial separation

The choice of $\pi^\star$ value affects the resulting constraints on HOD parameters. To minimize RSD effect and make the comparison more direct to simplified projected 3PCF, which ignored line-of-sight separation, we extend this value to $100h^{-1}$ Mpc in the main analysis. To check that this does not significantly alter results we also ran MCMC chains where this value was set to a more conventional $\pi^\star = 40h^{-1}$ Mpc for the projected 2PCF. These results are presented in Fig. 2.B.1. The figure is identical to Fig. 2.9 except we added green contours that correspond to the projected 2PCF results with $\pi^\star = 40h^{-1}$ Mpc. This choice of $\pi^\star$ is indeed more optimal but the likelihood surfaces do not change enough to alter any of our main conclusions. The projected 3PCF(SV) still dominates the joint constraints. Applying full definition projected 3PCF would likely also increase its constraining power.

2.C Analytical Random Triplet Counts

The $RRR\left(r_{12}^{\min}, r_{12}^{\max}, r_{23}^{\min}, r_{23}^{\max}, r_{31}^{\min}, r_{31}^{\max}\right)$ represent the average number of triplets of a random (spatially uncorrelated) distribution of galaxies, where the perpendicular to the line-of-sight distance between the triplet points satisfies conditions $r_{12}^{\min} < r_{p12} < r_{12}^{\max}$, $r_{23}^{\min} < r_{p23} < r_{23}^{\max}$, and $r_{31}^{\min} < r_{p31} < r_{31}^{\max}$. They are usually computed by explicitly count-

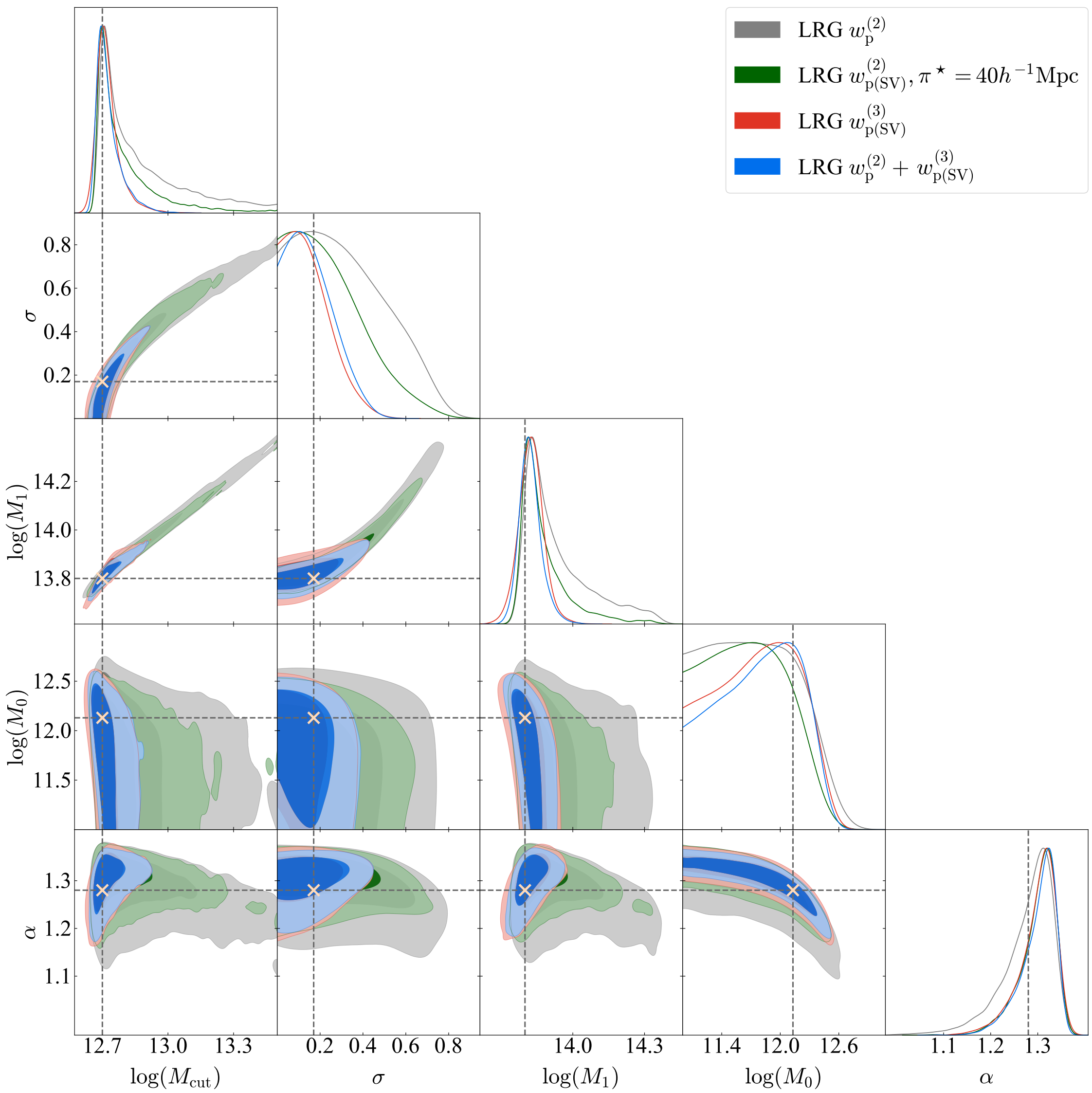


Figure 2.B.1: A similar plot as 2.9 with additional green contours shows results from the projected 2PCF but with a value of $\pi^\star = 40h^{-1}$ Mpc.

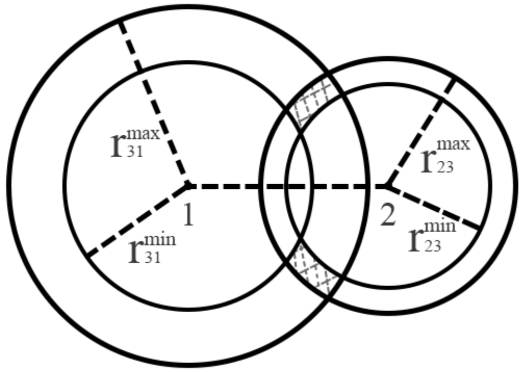


Figure 2.C.1: Geometry of random triplets problem. A fixed pair r_{12} with certain binning setting can only have triplets shown in the shaded area.

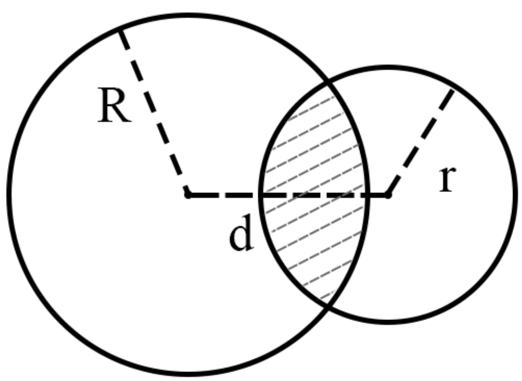


Figure 2.C.2: Intersection of two circles with radius R, r and distance of centers d. Intersection area A has been shaded.

ing triplets of a random distribution of points, but for a box with periodic boundaries these triplet-counts are easy to compute analytically. We follow the approach similar to Pearson and Samushia[122] when computing these triplet counts.

We start by computing a simpler quantity, $RRR^{\star}\left(r_{12}, r_{23}^{\min}, r_{23}^{\max}, r_{31}^{\min}, r_{31}^{\max}\right)$, the average number of third neighbours for a fixed pair separated by an exact perpendicular distance of r_{12}. Fig. 2.C.1 shows the geometry of the problem. For a fixed pair of points, $RRR^{\star}$ is an average number of points falling within the shaded areas.

$$RRR^{\star}\left(r_{12}, r_{23}^{\min}, r_{23}^{\max}, r_{31}^{\min}, r_{31}^{\max}\right) = \overline{\sigma} S^{\star}(r_{12}, r_{23}^{\min}, r_{23}^{\max}, r_{31}^{\min}, r_{31}^{\max}), \tag{2.37}$$

where $\overline{\sigma} = N/L^2$, is the projected density of the points (N being the number of points, and L the side of the cube).

The relationship of this simplified quantity with the full triplet count is,

$$\begin{aligned} &RRR\left(r_{12}^{\min}, r_{12}^{\max}, r_{23}^{\min}, r_{23}^{\max}, r_{31}^{\min}, r_{31}^{\max}\right) = \\ &\quad \int_{r_{12}^{\min}}^{r_{12}^{\max}} RRR^{\star}\left(r_{12}, r_{23}^{\min}, r_{23}^{\max}, r_{31}^{\min}, r_{31}^{\max}\right) N(\rho 2\pi r_{12} \mathrm{d}r_{12}), \end{aligned} \tag{2.38}$$

where N is the total number of possible first particles in the triplet and $(\rho 2\pi r_{12} \mathrm{d}r_{12})$ is the average number of second particles in the triplet as we integrate over the r_{12} bin.

We compute $S^{\star}$ using the expression for the area of the intersection of two circles

$$\begin{aligned} A(d, R, r) = &\, r^2 \arccos\left(\frac{d^2 + r^2 - R^2}{2dr}\right) + R^2 \arccos\left(\frac{d^2 - r^2 + R^2}{2dR}\right) \\ &- \frac{1}{2}\sqrt{(-d + r + R)(d + r - R)(d - r + R)(d + r + R)}. \end{aligned} \tag{2.39}$$

Here, d is the distance between the two circles, R and r are the two radii, and A is the shaded area on Fig. 2.C.2. From Fig. 2.C.1 and 2.C.2 it is clear that

$$S^{\star} = A(r_{12}, r_{23}^{\max}, r_{31}^{\max}) - A(r_{12}, r_{23}^{\max}, r_{31}^{\min}) - A(r_{12}, r_{23}^{\min}, r_{31}^{\max}) + A(r_{12}, r_{23}^{\min}, r_{31}^{\min}). \tag{2.40}$$

In our code we keep track of identical triplets by imposing the condition $r_{p12} < r_{p23} < r_{p31}$. This ensures that we don't count the same physical triplet corresponding to the same particles several times by relabeling the particles 1, 2, and 3. This does not happen (because of the way our code is written) for the triplets for which either three sides or at least two sides of the triangle fall into the same bin, so those triplets are counted more than ones. To correct for this, we apply a permutation factor $N_{\rm perm}$ to our RRR counts. The permutation factor is

$$N_{\rm perm}(r_1, r_2, r_3) = \begin{cases} 6 & \text{for} \quad r_1 \neq r_2 \neq r_3, \\ 3 & \text{for} \quad r_1 = r_2 \neq r_3 \text{ or} \\ & \phantom{\text{for}} \quad r_1 = r_3 \neq r_2 \text{ or} \\ & \phantom{\text{for}} \quad r_2 = r_3 \neq r_1, \\ 1 & \text{for} \quad r_1 = r_2 = r_3. \end{cases} \tag{2.41}$$

Chapter 3

Exploring Halo Occupation Distribution of Luminous Red Galaxies and Quasi-Stellar Objects in DESI One-Percent Survey

This chapter originally appeared in the literature as

Sihan Yuan, Hanyu Zhang et al. The DESI One-Percent Survey: Exploring the Halo Occupation Distribution of Luminous Red Galaxies and Quasi-Stellar Objects with AbacusSummit.

Chapter Abstract

We present the first Halo Occupation Distribution (HOD) analysis of the DESI One-Percent survey Luminous Red Galaxy (LRG) and Quasi Stellar Object (QSO) samples. We measure the redshift-space galaxy 2-point correlation functions from $0.15h^{-1}$Mpc up to $32h^{-1}$Mpc in a set of fiducial redshift bins and use them to constrain the HOD of each sample and test possible HOD extensions. In terms of modeling, we use AbacusSummit cubic boxes at Planck 2018 cosmology and generate galaxy mocks with

the AbacusHOD package. For LRGs, we infer a satellite fraction of $11 \pm 1\%$ and a mean halo mass of $\log_{10} \overline{M}_h = 13.40^{+0.02}_{-0.02}$ in redshift range $0.4 < z < 0.6$, and a satellite fraction of $14 \pm 1\%$ and a mean halo mass of $\log_{10} \overline{M}_h = 13.24^{+0.02}_{-0.02}$ in $0.6 < z < 0.8$. For QSOs, we infer a satellite fraction of $3^{+8}_{-2}\%$ and a mean halo mass of $\log_{10} \overline{M}_h = 12.65^{+0.09}_{-0.04}$ in redshift range $0.8 < z < 2.1$. We achieve good fits with a standard HOD model with velocity bias, and we find no evidence for galaxy assembly bias or satellite profile modulation at the current level of statistical uncertainty. Using these fits, we generate a large suite of galaxy mocks that closely match the observed clustering across all scales. We also present results on the redshift-evolution of the DESI LRG sample from $z = 0.4$ up to $z = 1.1$. Specifically the mean halo mass decreases with redshift whereas the satellite fraction increases with redshift. This paper is the first of a series of detailed galaxy–halo connection studies on the DESI One Percent Survey galaxy samples.

3.1 Introduction

Galaxies are biased tracers of the underlying matter density field of the Universe, and their distribution is an important source of cosmological and astrophysical information. However, while the distribution of dark matter is readily modeled by gravitational collapse, the distribution of galaxies is significantly more complex due to nonlinear evolution and baryonic processes. Thus, to extract cosmology and galaxy physics from the observed galaxy distribution, it is critical to model the connection between galaxies and their underlying dark matter density field.

A key piece of simplification in galaxy–dark matter connection modeling comes in what is known as the halo model, where simulations have shown that galaxies form and evolve in dense dark matter clumps known as halos[32;123]. Within the halo model, we can empirically model the connection between galaxies and halos through a set of probabilistic models known as the Halo Occupation Distribution model HOD; e.g.[5;7;28;30;31;33;34]. The HOD formalism has been highly successful in characterizing magnitude-limited samples of bright galaxies in past galaxy redshift surveys e.g.[38;41;56;63;124–127]. HOD studies are important not only because they reveal aspects of galaxy evolution physics and test assumptions of galaxy–dark matter connection e.g.[38;128–131], but also because they produce mocks that accurately reproduce the observed clustering and thus enable robustness tests of cosmology pipelines e.g.[39;43;132]. Most recently, simulation-based forward modeling approaches have also utilized the flexibility of HODs to constrain cosmology from highly nonlinear scales that are otherwise inaccessible with standard analytical approaches e.g.[133–137].

The Dark Energy Spectroscopic Instrument (DESI) is a stage-IV spectroscopic galaxy survey with the primary goal of determining the nature of dark energy through the most precise measurement of the expansion history of the universe ever obtained[104;138]. The baseline survey will obtain spectroscopic measurements of 40 million galaxies and quasars in a 14,000 deg^2 footprint in five years. This represents an order-of-magnitude improvement both in the volume surveyed and the number of galaxies measured over previous surveys. The DESI large-scale structure samples are divided into 4 target classes: the bright galaxy sample (BGS), the luminous red galaxies (LRG), the emission line galaxies (ELG), and the quasi-stellar objects (QSO). The auto- and cross-correlations of and between the four tracers probe the large-scale structure in increasing high redshift domains and combine to produce the most precise large-scale structure measurement from redshift $z = 0.1$ all the way to $z = 2.1$. Additionally, quasars that have redshifts greater than 2.1 are used as sight-lines for Lyman-α forest absorption, and the combination of Lyα-Lyα, Lyα-QSO, and QSO-QSO correlations probe large-scale structure to $z < 3.5$.

The Early Data Release (EDR) of the DESI survey consists of data in the so-called One Percent Survey, collected during the Survey Validation campaign[139] before the start of the main survey operations. The One-Percent survey covered 20 fields totaling 140 deg^2 with final target selection algorithms similar to those of the main survey[140–147]. The One-Percent survey reaches higher completeness than the main survey and produces the first clustering measurements from DESI. Specifically, more than 95% targets received fibers in the ELG sample, while more than 99% of targets in each of the BGS, LRG, and QSO samples received fibers.

In this paper, we present a comprehensive HOD analysis of the DESI One Percent Survey LRG and QSO samples. This paper is amongst a series of papers analyzing galaxy–halo connection models with DESI One-Percent Survey data. This paper addresses the more well-understood samples of LRG and QSO, while the more novel ELG sample is analyzed in a dedicated paper, Rocher et al.[148]. In parallel, there are also several Subhalo-abundance matching (SHAM) analyses. Specifically, Prada et al.[149] provides an overview of the Uchuu-based SHAM analyses[150]. Yu et al.[151] presents SHAM analyses based on the UNIT simulation[152]. Beyond the single-tracer analyses, Gao et al.[153] and Yuan et al. in prep analyze the cross-correlation functions between the ELG and LRG tracers with multi-tracer SHAM and HOD models, respectively. These papers together present a significant variety of methodologies and mock products appropriate for a large scope of applications.

This paper is structured as the following. In section 3.2, we introduce the observed samples and present their clustering measurements. In section 3.3 and 3.4, we introduce the simulation suite and the HOD models. In section 3.6, we present LRG fits on both the projected clustering measurements and the full-shape redshift-space clustering measurements, and present the corresponding model constraints. We present the QSO fits in section 3.7. In section 3.8, we present a series of mock products as a result of this analysis. Finally, we conclude in section 3.9.

Throughout this paper, we adopt the Planck 2018 ΛCDM cosmology, specifically the mean estimates of the Planck TT,TE,EE+lowE+lensing likelihood chains: $\Omega_c h^2 = 0.1200$, $\Omega_b h^2 = 0.02237$, $\sigma_8 = 0.811355$, $n_s = 0.9649$, $h = 0.6736$, $w_0 = -1$, and $w_a = 0$[2].

3.2 Data

In this section, we describe the LRG and QSO samples and present their respective clustering measurements.

DESI observed its One Percent Survey as the third and final phase of its Survey Validation program in April and May of 2021. Observation fields were chosen to be in twenty non-overlapping 'rosettes', where a high completeness was obtained by observing in each rosette at least 12 times. See[139] and[154] for more details.

Prior to beginning SV, the DESI instrument[3] had proven its ability to simultaneously measure spectra at 5000 specific sky locations, with fibers placed accurately using robotic positioners populating the DESI focal plane[155]. During SV, the DESI data and operations teams[156] proved their ability to efficiently process the spectra through the DESI spectroscopic pipeline[157]. Thus, DESI was able to start from an initial target list[158] and quickly obtain a highly complete One Percent Survey.

The redshift measurements we use are available in the DESI EDR[154] [1]. These were input to the large-scale structure (LSS) catalogs, also described in the EDR[154]. Briefly, these LSS catalogs apply quality cuts to the data samples and provide matched random catalogs that trace the angular footprint and dN/dz of the data, at a total density of $4.5\times10^4\,\mathrm{deg}^{-2}$.[159] describes how we simulated 128 alternative realizations of the DESI One Percent Survey fiber assignment in order to encode via bits the realizations where each target was assigned and thus any joint probabilities of observation for a given set of targets. We use this information to determine the pairwise-inverse-probability[160] weights to use in our clustering measurements. We further apply angular up-weighting (PIP+ANG)[161].[162] showed that this weighting scheme provides an unbiased clustering down to $0.1h^{-1}$Mpc.

The One Percent Survey LSS catalogs also include the so-called 'FKP'[163] weights in order to properly weight each volume element with respect to how each sample's number density changes with redshift,

$$w_{\mathrm{FKP}} = 1/(1 + n(z) \cdot P_0) \tag{3.1}$$

where $n(z)$ is the weighted number per volume, and P_0 is a fiducial power-spectrum amplitude. We use $P_0 = 10^4 (h^{-1}\mathrm{Mpc})^3$ for LRG and $P_0 = 6 \times 10^3 (h^{-1}\mathrm{Mpc})^3$ for QSO. For a detailed description of the weights and systematics treatment, we refer the readers to[154].

[1] https://data.desi.lbl.gov/public/edr/spectro/redux/fuji

3.2.1 DESI One Percent Survey LRG and QSO samples

LRGs are an important type of galaxy for large-scale structure studies, and are specifically selected for observations due to two main advantages: 1) they are bright galaxies with the prominent 4000Å break in their spectra, thus allowing for relatively easy target selection and redshift measurements; and 2) they are highly biased tracers of the large-scale structure, thus yielding a higher S/N per-object for the BAO measurement compared to typical galaxies. The LRG SV target selection is defined in Zhou et al.[140]. The sample has a target density of 605 deg^{-2} in $0.4 < z < 0.8$, significantly higher than previous LRG surveys BOSS and eBOSS[164;165], while the sample also extends to $z \sim 1$. Within EDR, the LRG main sample consists of 89,059 galaxies, 43,269 in the northern footprint and 45,790 in the southern footprint.

Quasi-stellar objects (a.k.a. Quasars, or QSOs) are the tracers of choice to study large-scale structures at high redshift due to the fact that they are some of the most luminous extragalactic sources. DESI aims to obtain spectra of nearly three million quasars, reaching limiting magnitudes $r \sim 23$ and an average density of $\sim$310 targets per deg^2. Within EDR, the QSO selection yields 24,182 quasars within the redshift range $0.8 < z < 2.1$, and an additional 11,603 Ly-α quasars at higher redshift. For this study, we focus on the quasars at $z < 2.1$ which will be used for quasar clustering analysis.

Figure 3.2.1 shows the DESI One Percent Survey LRG and QSO mean density as a function of redshift $n(z)$. The vertical dashed lines correspond to fiducial bin edges defined for DESI cosmology studies. For the LRG sample, the number density remains fairly constant from $z = 0.4$ to $z = 0.8$ at approximately $5 \times 10^{-4} h^3$Mpc^{-3}. At $z > 0.8$, the LRG density drops off quickly, suggesting increasing incompleteness and strong redshift evolution. For the fiducial HOD analysis presented in section 3.6, we examine the sample in two redshift bins: $0.4 < z < 0.6$ and $0.6 < z < 0.8$. Section 3.6.3 presents a preliminary analysis of the redshift evolution of the LRGs at $z > 0.8$.

The QSO sample delivers roughly constant number density from $z = 0.8$ to $z = 2.1$, at $2 \times 10^{-5} h^3$Mpc^{-3}. In this analysis, we treat this entire redshift range as one single bin to

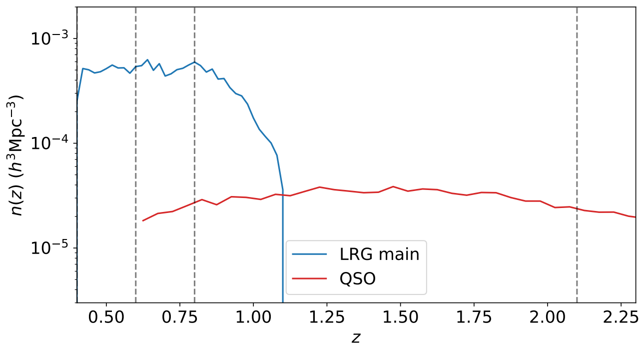


Figure 3.2.1: The DESI One Percent Survey LRG and QSO mean number density as a function of redshift. The dashed vertical lines show the fiducial LRG redshift bin edges of $z = 0.6$, $z = 0.8$, and the maximum redshift we consider for the QSO sample $z = 2.1$.

achieve a reasonably large sample size for clustering measurements. We nevertheless expect at least some degree of redshift evolution, but we defer the analysis of QSO redshift evolution to a future paper when a larger sample becomes available.

3.2.2 Clustering measurements

For this analysis, we consider the 2-point correlation function (2PCF) as our summary statistic of the galaxy clustering. We start by introducing the redshift-space 2PCF $\xi(r_p, r_\pi)$, which can be computed using the Landy and Szalay[62] estimator:

$$\xi(r_p, r_\pi) = \frac{DD - 2DR + RR}{RR}, \tag{3.2}$$

where DD, DR, and RR are the normalized numbers of data-data, data-random, and random-random pair counts in each bin of (r_p, r_π). r_p and r_π are transverse and line-of-sight (LoS) separations in comoving units. The redshift-space $\xi(r_p, r_\pi)$ in principle represents the full information content of the 2PCF. The dependence on transverse separation r_p describes the transition from 1-halo clustering to 2-halo clustering, whereas the dependence on LoS separation r_π details the velocity distributions and the small-scale finger-of-god effect.[63] showed that the $\xi(r_p, r_\pi)$ on small scales yield strong constraints on the HOD. In this paper, we consider $\xi(r_p, r_\pi)$ as our primary summary data vector for constraining the LRG and QSO HOD.

However, $\xi(r_p, r_\pi)$ is often compressed to the projected galaxy 2PCF w_p, which is the line-of-sight integral of $\xi(r_p, r_\pi)$,

$$w_p(r_p) = 2\int_0^{r_{\pi,\max}} \xi(r_p, r_\pi) dr_\pi, \tag{3.3}$$

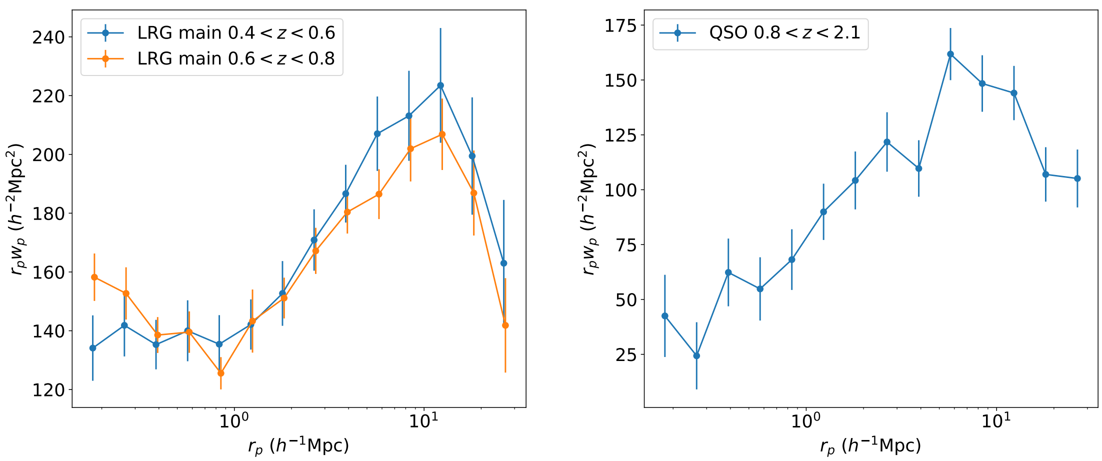
By definition, w_p is strictly less informative than $\xi(r_p, r_\pi)$ as it loses out on the velocity information that is encoded in the LoS clustering. However, w_p also offer several key advantages: it is easy to visualize as a 1D function; it is easy to obtain covariance matrix for; analyzing w_p avoids the complexities of modeling galaxy velocities. For these reasons, we present w_p-only results alongside the $\xi(r_p, r_\pi)$ results in the following analysis.

Figure 3.2.2 shows the projected auto-correlation function of the DESI One Percent Survey LRG and QSO samples, using the fiducial redshift bins we defined above. Throughout the rest of this analysis, we adopt 14 equally spaced logarithmic bins along the projected separation r_p from $0.15h^{-1}$Mpc to $32h^{-1}$Mpc. The projected scale range is designed to capture both the 1-halo regime and the 1 to 2 halo transition regime while limiting our exposure to large-scale modes due to the small footprint in the One-Percent Survey. Along the Line-of-sight (LoS) direction, we adopt a linear binning scheme from 0 to $32h^{-1}$Mpc with bin size $\Delta r_\pi = 4h^{-1}$Mpc to capture the structure of the finger-of-god effect without blowing up the size of the data vector. For w_p, we set $r_{\pi,\max} = 32h^{-1}$Mpc. The redshift multipole measurements are visualized in later figures (Figure 3.6.4 and 3.7.2). The errorbars displayed alongside the data measurements are calculated with 128 jackknife regions of the One Percent Survey footprint. All clustering measurements on DESI One Percent Survey data are done using the PYCORR package [2] [166].

3.3 Simulations

To model the underlying dark matter density field, we use the AbacusSummit simulation suite, which is a set of large, high-accuracy cosmological N-body simulations using the

[2] https://github.com/cosmodesi/pycorr



(a) LRG sample (b) QSO sample

Figure 3.2.2: The DESI One Percent Survey LRG and QSO projected auto-correlation functions. Here we are only showing LRG clustering in two fiducial redshift bins: $0.4 < z < 0.6$ and $0.6 < z < 0.8$. For QSOs, we consider one large redshift bin $0.8 < z < 2.1$ to achieve a reasonable sample size.

Abacus N-body code[105;107;167]. This suite is designed to meet the cosmological simulation requirements of DESI. AbacusSummit consists of over 150 simulations, containing approximately 60 trillion particles at 97 different cosmologies. A base simulation box contains 6912^3 particles within a $(2h^{-1}\text{Gpc})^3$ volume, which yields a particle mass of $2.1 \times 10^9 h^{-1} M_\odot$. [3]

The simulation output is organized into discrete redshift snapshots. Specifically, we use the $z = 0.5$ and $z = 0.8$ snapshots for LRGs at $z < 0.8$, and the $z = 0.8$ and $z = 1.1$ snapshots for LRGs at $z > 0.8$. For the QSO analysis, due to the very limited sample size, we choose not to divide the sample into multiple redshift bins. Instead, we use the $z = 1.4$ snapshot for the single redshift bin $0.8 < z < 2.1$. A more nuanced analysis of the QSO sample is planned when more data become available. All fits are done at Planck cosmology with the `AbacusSummit_base_c000_ph000` box.

The dark matter halos are identified with the CompaSO halo finder, which is a highly efficient on-the-fly group finder specifically designed for the AbacusSummit simulations[111]. CompaSO builds on the existing spherical overdensity (SO) algorithm by taking into consideration the tidal radius around a smaller halo before competitively assigning halo mem-

[3] For more details, see https://abacussummit.readthedocs.io/en/latest/abacussummit.html

bership to the particles in an effort to more effectively deblend halos. Among other features, the CompaSO finder also allows for the formation of new halos on the outskirts of growing halos, which alleviates a known issue of configuration-space halo finders of failing to identify halos close to the centers of larger halos. We also run a post-processing "cleaning" procedure that leverages the halo merger trees to "re-merge" a subset of halos. This is done both to remove over-deblended halos in the spherical overdensity finder, and to intentionally merge physically associated halos that have merged and then physically separated[106].

In addition to periodic boxes, the simulation suite also provides a set of simulation lightcones at fiducial cosmology[168]. The basic algorithm associates the halos from a set of coarsely-spaced snapshots with their positions at the time of light-cone crossing by matching halo particles to on-the-fly lightcone particles. The resulting halo catalogs are reliable at $M_{\rm halo} > 2.1 \times 10^{11} h^{-1} M_\odot$, more than sufficient for LRGs and QSOs. As part of the data products, we take the best-fit HODs across different redshift snapshots and construct redshift-dependent LRG mocks on the 25 base lightcones. Each lightcone covers an octant of the sky (~ 5156 deg^2) up to $z \sim 0.8$. We clarify that in this analysis, we only use the cubic boxes to conduct our analysis, the lightcones are only used to produce redshift-dependent mocks as part of the data products.

3.4 Halo Occupation Distribution (HOD)

To propagate the simulated matter density field to galaxy distributions, we adopt the Halo Occupation Distribution model (HOD), which probabilistically populate dark matter halos with galaxies according to a set of halo properties. Statistically, the HOD can be summarized as a probabilistic distribution $P(n_g|\boldsymbol{X}_h)$, where n_g is the number of galaxies of the given halo, and $\boldsymbol{X}_h$ is some set of halo properties.

In the vanilla HOD model, halo mass is assumed to be the only relevant halo property $\boldsymbol{X}_h = M_h$[33;34]. This vanilla HOD separates the galaxies into central and satellite galaxies and assumes the central galaxy occupation follows a Bernoulli distribution whereas the satellites follow a Poisson distribution. Beyond the vanilla model, galaxy occupation can also depend on secondary halo properties beyond halo mass, a phenomenon commonly referred to as galaxy assembly bias or galaxy secondary bias See[1] for a review. While galaxy assembly bias is well physically motivated, many studies have looked for it both in simulations and data e.g.[9;44;130;169–176] with mixed results.

For this analysis, we use the AbacusHOD code to find best-fit HODs and sample HOD posteriors. AbacusHOD is a highly efficient HOD implementation that enables a large set of HOD extensions[63]. The code is publicly available as a part of the abacusutils package at http://https://github.com/abacusorg/abacusutils. An example usage can be found at https://abacusutils.readthedocs.io/en/latest/hod.html.

3.4.1 Baseline model

For LRG sample, the HOD is well approximated by a vanilla model given by (originally shown in Zheng et al.[34] and referred to as Z07 or vanilla later in the text):

$$\bar{n}_{\mathrm{cent}}^{\mathrm{LRG}}(M) = \frac{f_{\mathrm{ic}}}{2}\mathrm{erfc}\left[\frac{\log_{10}(M_{\mathrm{cut}}/M)}{\sqrt{2}\sigma}\right], \tag{3.4}$$

$$\bar{n}_{\mathrm{sat}}^{\mathrm{LRG}}(M) = \left[\frac{M-\kappa M_{\mathrm{cut}}}{M_1}\right]^{\alpha}\bar{n}_{\mathrm{cent}}^{\mathrm{LRG}}(M), \tag{3.5}$$

where the five vanilla parameters characterizing the model are $M_{\mathrm{cut}}, M_1, \sigma, \alpha, \kappa$. M_{cut} sets the minimum halo mass to host a central galaxy. M_1 roughly sets the typical halo mass that hosts one satellite galaxy. σ controls the steepness of the transition from 0 to 1 in the number of central galaxies. α is the power law index on the number of satellite galaxies. κM_{cut} gives the minimum halo mass to host a satellite galaxy. We have added a modulation term $\bar{n}_{\mathrm{cent}}^{\mathrm{LRG}}(M)$ to the satellite occupation function to mostly remove satellites from halos without centrals[4]. We have also included an incompleteness parameter f_{ic}, which is a downsampling factor controlling the overall number density of the mock galaxies. This parameter is relevant when trying to match the observed mean density of the galaxies in addition to clustering measurements. By definition, $0 < f_{\mathrm{ic}} \leq 1$.

[4]There is evidence that such central-less satellites may exist in a realistic stellar-mass selected catalog[177]. We include this term for consistency with previous works, but it should have minimal impact on clustering.

For QSOs, we adopt essentially the same HOD model except we remove the central modulation term in the satellite occupation as there is no evidence that the existence of satellite QSOs is strongly associated with central QSOs. Thus, for satellite QSOs, we have

$$\bar{n}_{\mathrm{sat}}^{\mathrm{QSO}}(M) = \left[\frac{M - \kappa M_{\mathrm{cut}}}{M_1}\right]^{\alpha}. \tag{3.6}$$

In addition to determining the number of galaxies per halo, the standard HOD model also dictates the position and velocity of the galaxies. In the vanilla model, the position and velocity of the central galaxy are set to be the same as those of the halo center, specifically the L2 subhalo center-of-mass for the CompaSO halos (see Hadzhiyska et al.[111]). For the satellite galaxies, they are randomly assigned to halo particles with uniform weights, each satellite inheriting the position and velocity of its host particle.

Because we are modeling the full-shape $\xi(r_\mathrm{p}, r_\pi)$, we also include an additional level of flexibility in the velocity model known as velocity bias in the baseline model. Velocity bias essentially parametrizes any biases in the velocities of the central and satellite galaxies relative to their respective host halos and particles. This is shown to be a necessary ingredient in modeling BOSS LRG redshift-space clustering on small scales e.g.[55;63]. Velocity bias has also been identified in hydrodynamical simulations and measured to be consistent with observational constraints e.g.[57;176].

We parameterize velocity bias through two additional parameters:

- $\alpha_{\mathrm{vel,c}}$ is the central velocity bias parameter, which modulates the peculiar velocity of the central galaxy relative to the halo center along the LoS. Specifically in this model, the central galaxy velocity along the LoS is thus given by

$$v_{\mathrm{cent,z}} = v_{\mathrm{L2,z}} + \alpha_{\mathrm{vel,c}}\delta v(\sigma_{\mathrm{LoS}}), \tag{3.7}$$

  where $v_{\mathrm{L2,z}}$ denotes the LoS component of the central subhalo velocity, $\delta v(\sigma_{\mathrm{LoS}})$ denotes the Gaussian scatter, and $\alpha_{\mathrm{vel,c}}$ is the central velocity bias parameter. By definition, $\alpha_{\mathrm{vel,c}} = 0$ corresponds to no central velocity bias. We also define $\alpha_{\mathrm{vel,c}}$ as non-negative, as negative and positive α_c are fully degenerate observationally.

- $\alpha_{\rm vel,s}$ is the satellite velocity bias parameter, which modulates how the satellite galaxy's peculiar velocity deviates from that of the local dark matter particle. Specifically, the satellite velocity is given by

$$v_{\rm sat,z} = v_{\rm L2,z} + \alpha_{\rm vel,s}(v_{\rm p,z} - v_{\rm L2,z}), \tag{3.8}$$

  where $v_{\rm p,z}$ denotes the line-of-sight component of particle velocity, and $\alpha_{\rm vel,s}$ is the satellite velocity bias parameter. $\alpha_{\rm vel,s} = 1$ indicates no satellite velocity bias, i.e. satellites perfectly track the velocity of their underlying particles.

To summarize, the baseline HOD model for both LRGs and QSOs is fully specified with the following 8 parameters: (1) 5 vanilla HOD parameters $M_{\rm cut}$, M_1, σ, α, κ; (2) an incompleteness parameter $f_{\rm ic}$; (3) velocity bias parameters $\alpha_{\rm vel,c}$ and $\alpha_{\rm vel,s}$.

3.4.2 Model extensions

AbacusHOD also enables additional physically motivated HOD extensions. In the following analysis, we will test whether the data favor the inclusion of such extensions. We summarize the relevant extensions for LRGs here and refer the readers to[63] for more details:

- $A_{\rm cent}$ or $A_{\rm sat}$ are the concentration-based secondary bias parameters for centrals and satellites, respectively. Also known as galaxy assembly bias parameters. $A_{\rm cent} = 0$ and $A_{\rm sat} = 0$ indicate no concentration-based secondary bias in the central and satellite occupation, respectively. A positive A indicates a preference for lower-concentration halos, and vice versa.

- $B_{\rm cent}$ or $B_{\rm sat}$ are the environment-based secondary bias parameters for centrals and satellites, respectively. The environment is defined as the mass density within a $r_{\rm env} = 5h^{-1}$Mpc tophat of the halo center, excluding the halo itself. $B_{\rm cent} = 0$ and $B_{\rm sat} = 0$

indicate no environment-based secondary bias. A positive B indicates a preference for halos in less dense environments, and vice versa.

- s is the satellite profile bias parameter, which modulates how the radial distribution of satellite galaxies within haloes deviates from the radial profile of the halo (potentially due to baryonic effects). $s = 0$ indicates no radial bias, i.e. satellites are uniformly assigned to halo particles. $s > 0$ indicates a more extended (less concentrated) profile of satellites relative to the halo, and vice versa.

For this paper, we will add each of these extensions on to the 8-parameter baseline HOD model and conduct fits on the data. We compare the fits to study whether any of these extensions are favored. However, we only test these extensions on the LRG sample. While similar extensions might also apply for QSOs, we lack sufficient sample size to meaningfully constrain such effects.

3.4.3 Redshift-space distortion

Having generated the mock galaxy catalogs with each HOD prescription, we need to compute the 2PCF to compare to the data. However, because the data is in redshift space, meaning the observed LoS positions of galaxies are shifted by their peculiar velocity divided by the Hubble constant, we need to incorporate this effect in our model too. Thus, we impose redshift-space distortion (RSD) on the z-axis positions of the mock galaxies by the amount

$$Z_{\text{redshift}} = Z_{\text{real}} + \frac{v_{\text{pec,Z}}(1+z)}{H(z)}, \tag{3.9}$$

where Z_{real} and Z_{redshift} are the real and redshift-space z-axis positions of the galaxies. $v_{\text{pec,Z}}$ is the galaxy peculiar velocity projected along the z-axis. $H(z)$ is the Hubble parameter at redshift z. The $1 + z$ scaling converts the coordinates into comoving units.

Finally, we compute the model predicted $\xi(r_\text{p}, r_\pi)$directly from mocks, assuming z-axis as the LoS direction. We use the grid-based 2PCF calculator Corrfunc[116] for efficiency.

3.5 Likelihood model and covariance matrix

To perform the subsequent optimizations and sampling of the HOD parameters, we need to construct a likelihood function. In this analysis, we assume a simple Gaussian likelihood and utilize the χ^2 statistic:

$$\chi^2_\xi = (\xi_{\rm model} - \xi_{\rm data})^T \boldsymbol{C}^{-1} (\xi_{\rm model} - \xi_{\rm data}), \tag{3.10}$$

where the $\xi_{\rm model}$ is the model predicted $\xi(r_p, \pi)$ and $\xi_{\rm data}$ is the DESI measurement. $\boldsymbol{C}$ is the covariance matrix.

We also include in the likelihood model an additional term related to the mean number density of the sample,

$$\chi^2_{n_g} = \begin{cases} \left(\frac{n_{\rm mock} - n_{\rm data}}{\sigma_n}\right)^2 & (n_{\rm mock} < n_{\rm data}) \\ 0 & (n_{\rm mock} \geq n_{\rm data}). \end{cases} \tag{3.11}$$

σ_n is the uncertainty of the galaxy number density. The $\chi^2_{n_g}$ is a half normal around the observed number density $n_{\rm data}$. When the mock number density is less than the data number density ($n_{\rm mock} < n_{\rm data}$), we set the completeness to $f_{\rm ic} = 1$ and give a Gaussian-type penalty on the difference between $n_{\rm mock}$ and $n_{\rm data}$. When the mock number density is higher than data number density ($n_{\rm mock} \geq n_{\rm data}$), then we set $f_{\rm ic} = n_{\rm data}/n_{\rm mock}$ such that the mock galaxies catalog is uniformly downsampled to match the data number density. In this case, we impose no penalty. This definition of $\chi^2_{n_g}$ allows for incompleteness in the observed galaxy sample while penalizing HOD models that produce insufficient galaxy number density. For the rest of this paper, we assume $\sigma_n = 0.1 n_{\rm data}$.

Finally, the full χ^2 is given by

$$\chi^2 = \chi^2_\xi + \chi^2_{n_g}. \tag{3.12}$$

To obtain the covariance matrix, one could divide the observed sample into jackknife regions and compute the clustering in each assuming they are independent realizations.

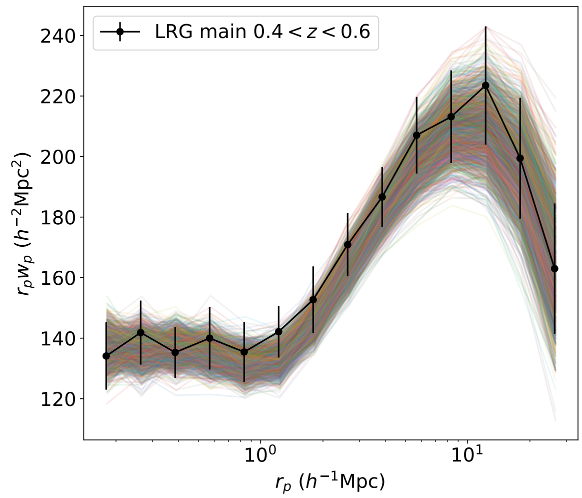


Figure 3.5.1: The projected auto-correlation function w_p of the 1800 boxes after tuning vanilla HOD parameters to match the clustering of One Percent Survey LRGs in $0.4 < z < 0.6$. See section 3.5 for details.

However, given the finite size of the One Percent Survey footprint and the relatively large number of bins in the $\xi(r_\mathrm{p}, r_\pi)$statistic, the resulting jackknife covariances are noisy and close to singular. Instead, we opt to use the 1800 $500h^{-1}$Mpc boxes with varying phases in the AbacusSummit suite to generate mock-based covariance matrices. Specifically, each small box shares the same particle resolution as the base boxes and is $500h^{-1}$Mpc per side, which is sufficient for the scales we analyze.

First, we generate mocks on the 1800 boxes that produce the same clustering as that measured in the data. Specifically, we take a baseline HOD model and fit the observed $\xi(r_\mathrm{p}, r_\pi)$with just the jackknife errors measured on the data, assuming all off-diagonal terms in the covariance matrices are zero. We achieve good fits for both tracers and in all redshift bins. We do not present the values of this fit to avoid confusion with the final "full covariance" fit presented in Table 3.6.3, but the parameter values are consistent with the "full covariance" fits. We then take the best-fit HOD and populate the 1800 covariance boxes, from which compute the covariance matrices. Finally, we renormalize the covariance matrix by keeping the mock-based correlation matrix and using the data-based jackknife diagonal errors to convert the correlation matrix to the final covariance matrix.

Figure 3.5.1 serves as a visualization of the 1800 realizations after tuning to match the

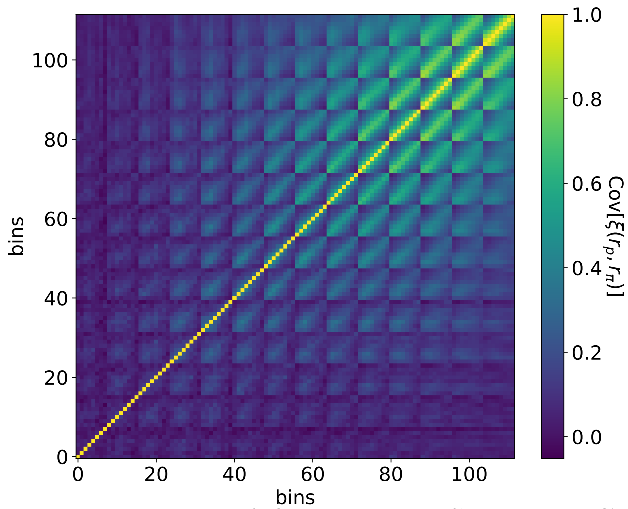


Figure 3.5.2: The correlation matrix of One Percent Survey LRG $\xi(r_\mathrm{p}, r_\pi)$in $0.4 < z < 0.6$, generated from the 1800 boxes after tuning to match the observed clustering. The 2D bins of $\xi(r_\mathrm{p}, r_\pi)$are collapsed in this representation as a column-wise stack (the bin number is strictly increasing in r_p and periodic in r_π)

observed $\xi(r_\mathrm{p}, r_\pi)$, where we overlay the projected auto-correlation functions of the 1800 boxes on the observation. We see that the mock realizations do well in producing the observed clustering, and the spread in the mock clustering is consistent in trend with the data jackknife error bars.

Figure 3.5.2 shows the resulting correlation matrix computed from the 1800 mocks for the LRG $\xi(r_\mathrm{p}, r_\pi)$in $0.4 < z < 0.6$. The 2D bins of $\xi(r_\mathrm{p}, r_\pi)$are collapsed in this representation as a column-wise stack (the bin number is strictly increasing in r_p and periodic in r_π). We see that the off-diagonal terms are determined with high signal-to-noise, a result of the large number of realizations and the large volume available with the 1800 boxes. The large-scale bins are correlated as they become sample variance dominated by large-scale structure, whereas the small-scale bins are largely independent, as they are dominated by the shot noise of galaxy occupation.

We similarly generate mock covariance matrices for the LRG sample in $0.6 < z < 0.8$ and the QSO sample. We omit those plots for brevity. The covariance matrix for the LRG sample in $0.6 < z < 0.8$ shares essentially the same structure as the lower redshift LRG sample, with shot noise dominating the smallest scales and sample variance becoming significant at larger scales. For QSOs, all scales are dominated by shot noise and the covariance matrix is essentially diagonal. Throughout the rest of this paper, we use this set of mock covariance matrices for model comparison and posterior sampling.

Having defined the likelihood function, we use optimization routines and posterior samplers to evaluate the best-fits and posterior constraints, respectively. Finally, we note that a correction term is applied to correct for the finite number of independent realizations used to calculate the covariance matrix[178]. Due to a large number of mock realizations (~1800), the correction factor is small but not negligible ($\sim 6\%$).

3.6 LRG HOD Results

In this section, we present the results of the One Percent Survey LRG and QSO HOD analysis by deriving the HOD best-fits, testing possible model extensions, and presenting the posterior constraints.

3.6.1 LRG at $z < 0.8$

We first examine the LRG main sample at $z < 0.8$, where the number density remains relatively constant. This is also the LRG sample that will be used for DESI Y1 fiducial cosmology analyses. We analyze this sample in two separate redshift bins: $0.4 < z < 0.6$ and $0.6 < z < 0.8$, using the $z = 0.5$ and $z = 0.8$ snapshots, respectively. We target the $\xi(r_\mathrm{p}, r_\pi) + n(z)$ data vector in each bin and incorporate the full covariance matrix built on mocks. Figure 3.6.1 shows the LRG $\xi(r_\mathrm{p}, r_\pi)$data vector in $0.4 < z < 0.6$. We omit the other redshift bin for brevity. We only utilize the transverse scales 0.15–32h^{-1}Mpc and LoS separation from 0 to 32h^{-1}Mpc.

We first test for potential extensions to the 8-parameter baseline HOD model (5-parameter vanilla HOD plus velocity bias plus incompleteness). Specifically, we run optimizations with extended models that include either galaxy assembly bias or satellite radial profile parameters in addition to the baseline parameters. We use a global optimization routine called the Covariance matrix adaptation evolution strategy (CMA-ES) with 400 random walkers.

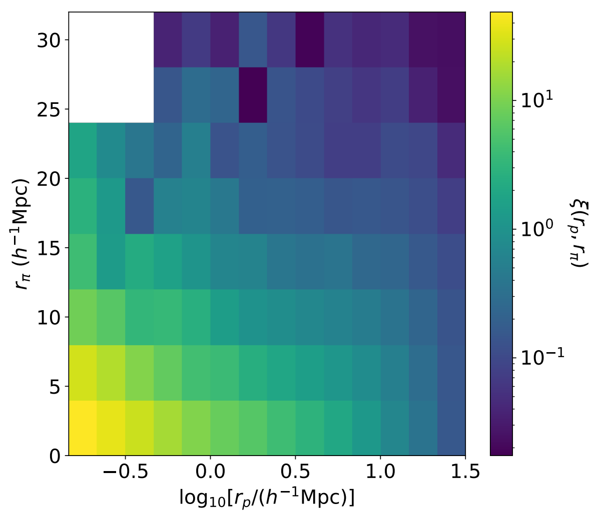


Figure 3.6.1: The One Percent Survey LRG $\xi(r_\mathrm{p}, r_\pi)$in $0.4 < z < 0.6$. r_p and r_π denote the transverse and LoS separation of the galaxies in comoving units. For this analysis we only utilize the transverse scales 0.15–32h^{-1}Mpc. The white section corresponds to negative values, which do not show up on the log scale.

We compute the model Akaike Information Criterion (AIC) scores from the best-fit χ^2 and summarize the results in Table 3.6.1. The AIC scores essentially calculate the best-fit χ^2 but compensate for the number of parameters. Models with lower AIC scores are preferred by the data, and a ΔAIC roughly corresponds to 1σ significance.

Our tests show no evidence for either flavors of galaxy assembly bias or a satellite radial profile parameter. We conclude that the current data vectors favor the baseline model, and we will conduct the rest of this analysis with just the baseline HOD model. However,[63] found significant evidence for galaxy assembly bias in a similar HOD analysis of BOSS CMASS LRGs. This discrepancy is explained by the significantly larger sample size in CMASS ($\sim 600,000$). The factor 10 decrease in sample size translates to a factor 3 increase in the statistical error, which in turn downgrades an assembly bias signal as detected in[63] to less than 1σ significance.

Next, we obtain the parameter posteriors for the baseline model in each redshift bin. To sample the HOD posterior, we use the efficient DYNESTY nested sampler[179;180]. Again, we use $\xi(r_\mathrm{p}, r_\pi)$as our primary data vector in each redshift bin and include the full covariance matrix in the likelihood evaluation. We incorporate broad multivariate Gaussian priors as quoted in

LRG	$0.4 < z < 0.6$		$0.6 < z < 0.8$	
Model	χ^2/d.o.f.	AIC	χ^2/d.o.f.	AIC
Baseline	1.03	130	0.99	125
Baseline+A	1.05	133	0.98	125
Baseline+B	1.05	133	0.99	126
Baseline+s	1.06	131	1.01	126

Table 3.6.1: Comparing LRG HOD model extensions in both redshift bins. The baseline model refers to the standard vanilla 5-parameter model plus velocity bias and incompleteness. A refers to galaxy assembly bias parameterized in terms of halo concentration. B refers to galaxy assembly bias parameterized in terms of the local environment. s refers to 1-halo satellite profile modulations. The AIC scores suggest that none of the extended models are preferred over the baseline model.

Params	LRG		QSO	
	prior	bounds	prior	bounds
$\log M_{\mathrm{cut}}$	$\mathcal{N}(13.0, 1.0)$	[12,13.8]	$\mathcal{N}(12.7, 1)$	[11.2, 14.0]
$\log M_1$	$\mathcal{N}(14.0, 1.0)$	[12.5,15.5]	$\mathcal{N}(15.0, 1)$	[12.0, 16.0]
σ	$\mathcal{N}(0.5, 0.5)$	[0.0,3.0]	$\mathcal{N}(0.5, 0.5)$	[0.0, 3.0]
α	$\mathcal{N}(1.0, 0.5)$	[0.0,2.0]	$\mathcal{N}(1.0, 0.5)$	[0.3, 2.0]
κ	$\mathcal{N}(0.5, 0.5)$	[0.0,10.0]	$\mathcal{N}(0.5, 0.5)$	[0.3, 3.0]
α_{c}	$\mathcal{N}(0.4, 0.4)$	[0.0, 1.0]	$\mathcal{N}(1.5, 1.0)$	[0.0, 2.0]
α_{s}	$\mathcal{N}(0.8, 0.4)$	[0.0, 2.0]	$\mathcal{N}(0.2, 1.0)$	[0.0, 2.0]

Table 3.6.2: Priors used for LRG and QSO HOD fits. We use broad Gaussian priors on all parameters. We also quote the bounds we impose in addition to the Gaussian priors. Units of mass are in $h^{-1}M_\odot$.

Tracer	LRG $0.4 < z < 0.6$		LRG $0.6 < z < 0.8$		QSO $0.8 < z < 2.1$	
Model	Z07 + $f_{\rm ic}$	Z07 + $f_{\rm ic}$ $+\alpha_{\rm c}+\alpha_{\rm s}$	Z07 + $f_{\rm ic}$	Z07 + $f_{\rm ic}$ $+\alpha_{\rm c}+\alpha_{\rm s}$	Z07 + $f_{\rm ic}$	Z07 + $f_{\rm ic}$ $+\alpha_{\rm c}+\alpha_{\rm s}$
Data	$w_{\rm p}+n_z$	$\xi(r_{\rm p},r_\pi)+n_z$	$w_{\rm p}+n_z$	$\xi(r_{\rm p},r_\pi)+n_z$	$w_{\rm p}+n_z$	$\xi(r_{\rm p},r_\pi)+n_z$
$\log M_{\rm cut}$	$12.89^{+0.11}_{-0.09}$	$12.79^{+0.15}_{-0.07}$	$12.78^{+0.10}_{-0.08}$	$12.64^{+0.17}_{-0.05}$	$12.67^{+0.71}_{-0.36}$	$12.2^{+0.6}_{-0.1}$
$\log M_1$	$14.08^{+0.10}_{-0.10}$	$13.88^{+0.11}_{-0.11}$	$13.94^{+0.14}_{-0.11}$	$13.71^{+0.07}_{-0.07}$	$15.00^{+0.62}_{-0.64}$	$14.7^{+0.6}_{-0.6}$
σ	$0.27^{+0.17}_{-0.17}$	$0.21^{+0.11}_{-0.10}$	$0.23^{+0.14}_{-0.15}$	$0.09^{+0.09}_{-0.05}$	$0.58^{+0.37}_{-0.35}$	$0.12^{+0.28}_{-0.06}$
α	$1.20^{+0.15}_{-0.19}$	$1.07^{+0.13}_{-0.16}$	$1.07^{+0.16}_{-0.21}$	$1.18^{+0.08}_{-0.13}$	$1.09^{+0.43}_{-0.37}$	$0.8^{+0.4}_{-0.2}$
κ	$0.65^{+0.45}_{-0.39}$	$1.4^{+0.6}_{-0.5}$	$0.55^{+0.42}_{-0.34}$	$0.6^{+0.4}_{-0.2}$	$0.74^{+0.41}_{-0.29}$	$0.6^{+0.8}_{-0.2}$
α_c	-	$0.33^{+0.05}_{-0.07}$	-	$0.19^{+0.06}_{-0.09}$	-	$1.54^{+0.17}_{-0.08}$
α_s	-	$0.80^{+0.07}_{-0.07}$	-	$0.95^{+0.07}_{-0.06}$	-	$0.6^{+0.6}_{-0.3}$
$f_{\rm ic}$	$0.92^{+0.08}_{-0.17}$	$0.70^{+0.15}_{-0.09}$	$0.89^{+0.11}_{-0.14}$	$0.62^{+0.07}_{-0.06}$	$0.041^{+0.066}_{-0.016}$	$0.019^{+0.029}_{-0.004}$
$f_{\rm sat}$	$0.089^{+0.013}_{-0.010}$	$0.106^{+0.011}_{-0.012}$	$0.104^{+0.013}_{-0.010}$	$0.136^{+0.011}_{-0.010}$	$0.05^{+0.26}_{-0.05}$	$0.03^{+0.08}_{-0.02}$
$\log \overline{M}_{\rm h}$	$13.42^{+0.02}_{-0.02}$	$13.40^{+0.02}_{-0.02}$	$13.26^{+0.02}_{-0.02}$	$13.24^{+0.02}_{-0.02}$	$12.74^{+0.05}_{-0.05}$	$12.65^{+0.09}_{-0.04}$
$b_{\rm lin}$	$1.94^{+0.04}_{-0.04}$	$1.93^{+0.06}_{-0.04}$	$2.11^{+0.03}_{-0.04}$	$2.08^{+0.03}_{-0.03}$	$2.56^{+0.22}_{-0.10}$	$2.63^{+0.37}_{-0.26}$
χ^2/d.o.f	4.5/(14-5)	108/(112-7)	19.6/(14-5)	104/(112-7)	16.0/(14-5)	101/(112-7)

Table 3.6.3: LRG and QSO marginalized posteriors, with different models and different measurements. The error bars are 1σ uncertainties. We also quote the marginalized satellite fraction $f_{\rm sat}$ and the sample completeness $f_{\rm ic}$. Units of mass are given in $h^{-1}M_\odot$.

Table 3.6.2. We specifically test that our choice of Gaussian priors will not significantly bias our main results. (see App. 3.A for detail) We also impose bounds to limit the range explored by the sampler (also documented in Table 3.6.2). The resulting marginalized posteriors are presented in Figure 3.6.2 and summarised in Table 3.6.3. We visualize the corresponding HOD posteriors in Figure 3.6.3, where the shaded bands denote the 1 and 2σ constraints (68% and 95% intervals).

Figure 3.6.2 shows several interesting degeneracies. Perhaps the most prominent degeneracy is between parameter $\log M_{\rm cut}$ and σ. Both parameters control the occupation distribution of the centrals, which for LRGs translates to controlling the clustering amplitude of the 2-halo term on large scales. $\log M_{\rm cut}$ controls the mass scale whereas σ controls the slope of the $N_{\rm cent}$ turn-on. Thus, it makes sense that the two are somewhat degenerate as either a slower turn-on (larger σ) or a lower mass scale would decrease the mean bias of the sample.

Similarly, we see a degeneracy between the completeness parameter $f_{\rm ic}$ and the central occupation parameters $\log M_{\rm cut}$ and σ. This is due to the constraints on the average density of the galaxy sample. Lower typical halo masses for the galaxies would mean lower completeness. There is also an interesting degeneracy between $\log M_{\rm cut}$ and $\log M_1$, which has been previously discussed in Avila et al.[41]. This is interesting because $\log M_1$ controls the halo mass of the satellite galaxies. Thus, a degeneracy between these two parameters suggests a strong constraint on the satellite fraction, and by extension a strong constraint on the relative amplitude of the 1-halo clustering and 2-halo clustering.

The inferred satellite fraction is $11 \pm 1\%$ for LRGs in $0.4 < z < 0.6$ and $14 \pm 1\%$ in $0.6 < z < 0.8$, consistent with 11% inferred for the CMASS sample in $0.45 < z < 0.6$[63] and $13 \pm 3\%$ inferred with eBOSS LRGs between $0.6 < z < 0.9$[37]. The mean halo mass of the LRGs is strongly constrained. We find $\log_{10} \overline{M}_h = 13.40^{+0.02}_{-0.02}$ for $0.4 < z < 0.6$ and $\log_{10} \overline{M}_h = 13.24^{+0.02}_{-0.02}$ for $0.6 < z < 0.8$. In comparison,[129] found $\log_{10} \overline{M}_h = 13.60$ for the CMASS sample, and[37] found $\log_{10} \overline{M}_h = 13.4$ for the higher redshift eBOSS sample. Both values are higher than the average halo mass inferred for DESI One Percent Survey LRGs. This is expected as the DESI sample is fainter and higher number density, thus occupying

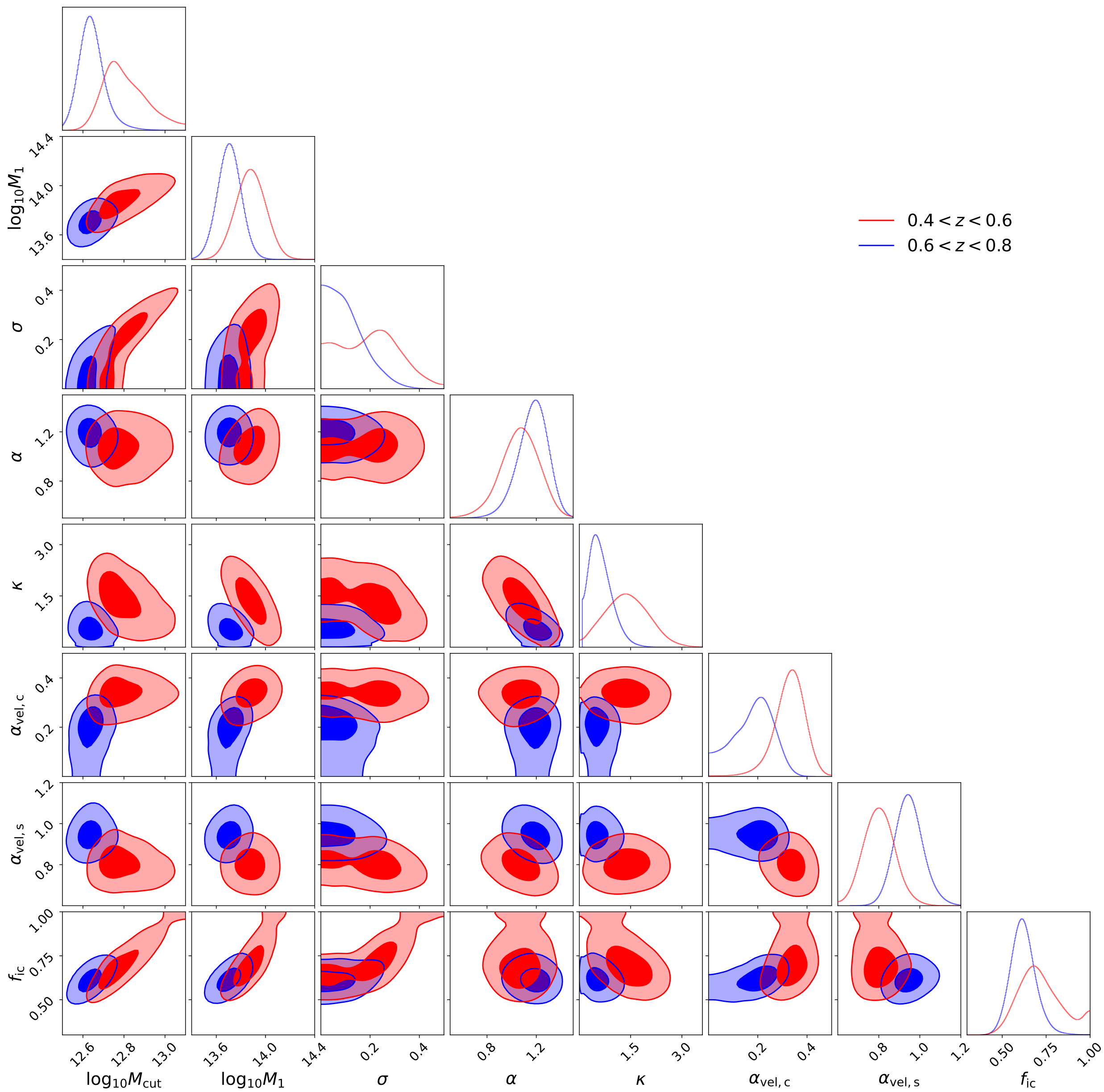


Figure 3.6.2: The DESI One Percent Survey LRG HOD posterior. The red and blue contours correspond to the marginalized posteriors for LRGs in the $0.4 < z < 0.6$ and $0.6 < z < 0.8$ bins, respectively. We are showing only the 1 and 2 σ contours for clarity.

less massive halos. Our results also compare well with earlier results from CMASS DR10, which reported a 9-10% satellite fraction for a red luminosity-limited sample with half the DESI density[126]. The difference in average halo mass between the two redshift bins can simply be attributed to halo growth at fixed density.

The linear bias factor is also calculated in our study by comparing the real-space clustering amplitudes of galaxies with predictions made by the linear theory, specifically

$$b_{\rm lin} = (\xi_{\rm gals}/\xi_{\rm lin})^{1/2}, \tag{3.13}$$

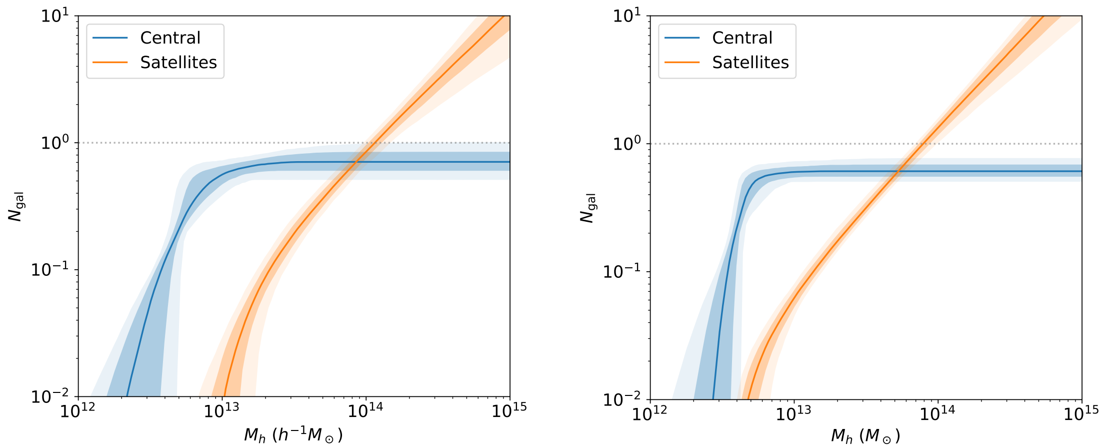
where $\xi_{\rm gal}$ denotes the real-space two-point correlation functions of galaxies, while $\xi_{\rm lin}$ denotes the theoretical linear matter correlation function at the mean redshift in the respective redshift bin. This correlation function is measured with `CLASS`[181]. For each sample of the MCMC chain, $b_{\rm lin}$ is calculated for a uniform subsample of the full posterior, which we summarise with mean and standard deviation. The LRG linear bias is $1.93^{+0.06}_{-0.04}$ in the redshift range of $0.4 < z < 0.6$ and 2.08 ± 0.03 for LRGs in $0.6 < z < 0.8$. We present a more detailed description of redshift evolution in section 3.6.3.

The velocity bias constraints are also mostly consistent with BOSS and eBOSS studies. For the $0.4 < z < 0.6$ bin, we find significant central velocity bias at $33^{+5}_{-7}\%$, indicating that the peculiar velocity of the central galaxies relative to the central subhalos is approximately 30% of the halo velocity dispersion. This is qualitatively consistent with previous studies that also found significant central velocity bias, but somewhat larger in amplitude than the CMASS constraints at $22 \pm 2\%$[55;176]. We also find negative satellite velocity bias $80 \pm 7\%$, indicating that the velocity dispersion of satellite galaxies within halos is 20% less than that of the halo particles at the same radii. This is consistent with[55], who found 86% in CMASS, whereas[176] found less significant velocity bias at 98%. In the higher redshift bin $0.6 < z < 0.8$, we find a central velocity bias of $19^{+6}_{-9}\%$ and a satellite velocity bias $95^{+7}_{-6}\%$. We do not have eBOSS constraints to compare against, but we can check with simulated DESI samples presented in[176], where we applied DESI photometric selection to IllustrisTNG galaxies[182–184]. There, we found that the mock DESI LRG sample at $z = 0.8$ has a velocity bias of $\alpha_c = 0.14$ and $\alpha_s = 0.92$, consistent with our constraints here.

We also run the equivalent analyses on the projected 2PCF w_p, where we follow the same procedure as for $\xi(r_\mathrm{p}, r_\pi)$, but we exclude the two velocity bias parameters from the HOD model. We also employ a tabulation scheme to accelerate the HOD forward model calculation for these w_p fits (please find details in Chap. 1.3.2 and Chap. 2), note that we only employ the tabulation method to quickly evaluate the HOD posterior when using the Z07+f_ic model to fit the $w_\mathrm{p} + n_z$ data vector, for all other cases, we use the standard AbacusHOD approach. We include the w_p marginalized constraints in Table 3.6.3. However, for brevity, we skip their visualization. Comparing the marginalized posteriors in Table 3.6.3, the w_p results are consistent with the $\xi(r_\mathrm{p}, r_\pi)$ results. In terms of inferred quantities, the two data vectors also yield mostly consistent results. The only minor discrepancy between the two fits is that the w_p fits favor larger M_cut and M_1 parameters, and as a result a larger f_ic. However, this is compensated by differences in the σ constraints, resulting in essentially identical mean halo mass constraints. In other words, both data vectors place strong constraints on the mean halo mass of the galaxies, but neither breaks the M_cut vs. σ degeneracy and favors slightly different loci along this degeneracy.

Comparing the posterior means in the two redshift bins, we see several interesting trends. The average halo mass of the LRG sample increases over time. It is expected given that halos accrete mass over time, and a fixed density sample at lower redshift would occupy the more massive halos. The satellite fraction also decreases over time. This trend is not as significant but might be interpreted as a result of galaxy mergers. We discuss redshift evolution in more detail in section 3.6.3, and also in a dedicated paper[185].

Figure 3.6.4 showcases the predicted distribution of the 2PCF from the $\xi(r_\mathrm{p}, r_\pi)$posteriors. For this visualization, we choose the $w_p + \xi_0 + \xi_2$ projections of the redshift-space 2PCF because it is difficult to visualize comparisons of the 2D $\xi(r_\mathrm{p}, r_\pi)$function. w_p is the projected 2PCF, whereas $\xi_{0,2}$ are the monopole and quadruple moments of the redshift-space 2PCF. The blue curves represent the One Percent Survey measurements, with jackknife error bars. The orange-shaded regions denote the 1 and 2σ posterior constraints. The solid orange line showcases the posterior mean. Again, we see that the best-fit models are consistent with



(a) LRG $0.4 < z < 0.6$

(b) LRG $0.6 < z < 0.8$

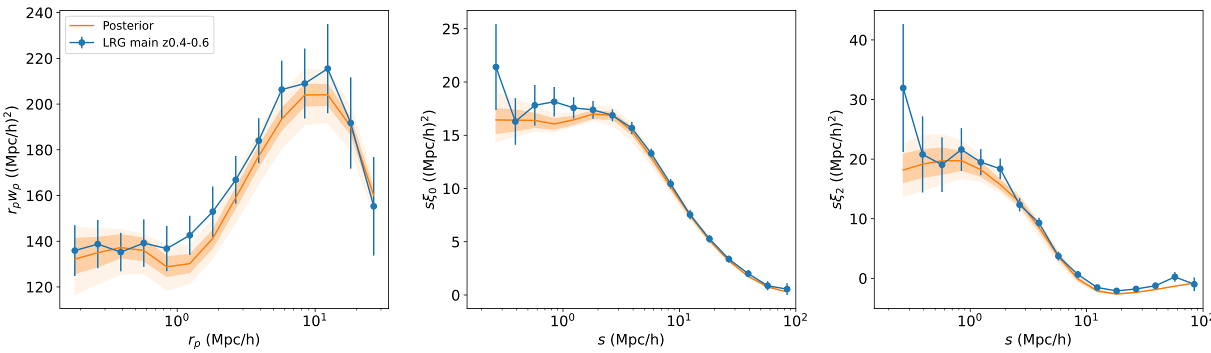
Figure 3.6.3: The LRG HOD best-fit the posterior. The shaded regions correspond to 1 and 2σ posteriors (68% and 95% intervals centered around the median prediction). The horizontal dotted line denotes $N_{\rm gal} = 1$.

the observed clustering well in both redshift bins. However, in the w_p comparisons, the model predicts a larger amplitude than the data at the 1-halo and 2-halo transition regime ($r_p \sim 1h^{-1}$Mpc) in both redshift bins. This is not a significant discrepancy with the current sample size but potentially points to limitations of halo boundary definitions and possibly environment-dependent galaxy occupation.

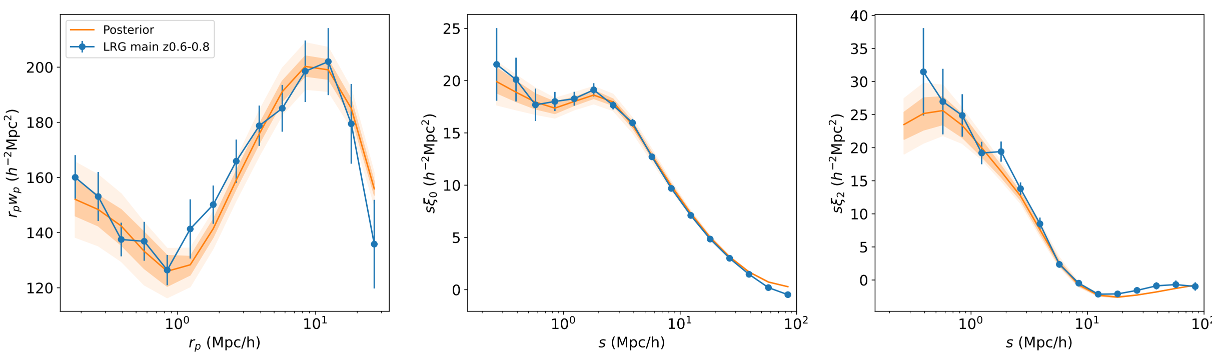
3.6.2 LRG at $z > 0.8$

As shown in Figure 3.2.1, the number density of the LRG main sample experiences a significant decline beyond a redshift of 0.8, indicating a marked change in the astrophysical properties of this population. To further investigate the HOD and shed light on the evolution of the LRG sample in this redshift range, we subdivide the high-redshift LRGs into two redshift intervals, $0.8 < z < 0.95$ and $0.95 < z < 1.1$. We use the $z = 0.8$ and $z = 1.1$ snapshots and employ the Z07+$f_{\rm ic}$ model to fit $w_{\rm p} + n_z$ data vector using the tabulation method for this aspect of the analysis.

Figure 3.6.5 shows the 1 and 2 σ confidence level contours for the HOD parameters. Due to the higher number density, the fit from the lower redshift interval displays a much tighter constraint as compared to the higher redshift interval, as anticipated. We find similar



(a) LRG $0.4 < z < 0.6$



(b) LRG $0.6 < z < 0.8$

Figure 3.6.4: The LRG w_p and multipoles posterior predicatives compared to the data. The blue lines correspond to the One Percent Survey measurement with jackknife error bars. The solid orange line denotes the posterior mean. The orange-shaded regions correspond to the 1 and 2σ full posterior using the full mock covariance matrix.

Params	LRG $0.8 < z < 1.1$	
	$0.8 < z < 0.95$	$0.95 < z < 1.1$
$\log M_{\rm cut}$	$12.89^{+0.12}_{-0.13}$	$12.68^{+0.38}_{-0.26}$
$\log M_1$	$13.96^{+0.15}_{-0.14}$	$13.60^{+0.47}_{-0.29}$
σ	$0.37^{+0.13}_{-0.18}$	$0.53^{+0.25}_{-0.29}$
α	$0.91^{+0.18}_{-0.22}$	$0.72^{+0.31}_{-0.34}$
κ	$0.74^{+0.46}_{-0.42}$	$0.51^{+0.43}_{-0.33}$
$f_{\rm ic}$	$0.92^{+0.08}_{-0.18}$	$0.19^{+0.14}_{-0.07}$
$f_{\rm sat}$	$0.110^{+0.016}_{-0.012}$	$0.151^{+0.048}_{-0.041}$
$\log_{10}\overline{M}_{\rm h}$	$13.29^{+0.02}_{-0.02}$	$13.00^{+0.03}_{-0.03}$
$b_{\rm lin}$	$2.31^{+0.04}_{-0.04}$	$2.13^{+0.05}_{-0.05}$
$\overline{n_z}$	4.56	1.84

Table 3.6.4: The results for the fits to high-z LRG sample with two redshift bin: $0.8 < z < 0.95$ and $0.95 < z < 1.1$. We show the mean$\pm1\sigma$ error for HOD and derived parameters. We also list the average comoving number density in units of $10^{-4}(h^{-1}\mathrm{Mpc})^{-3}$. Masses are in units of $h^{-1}M_\odot$.

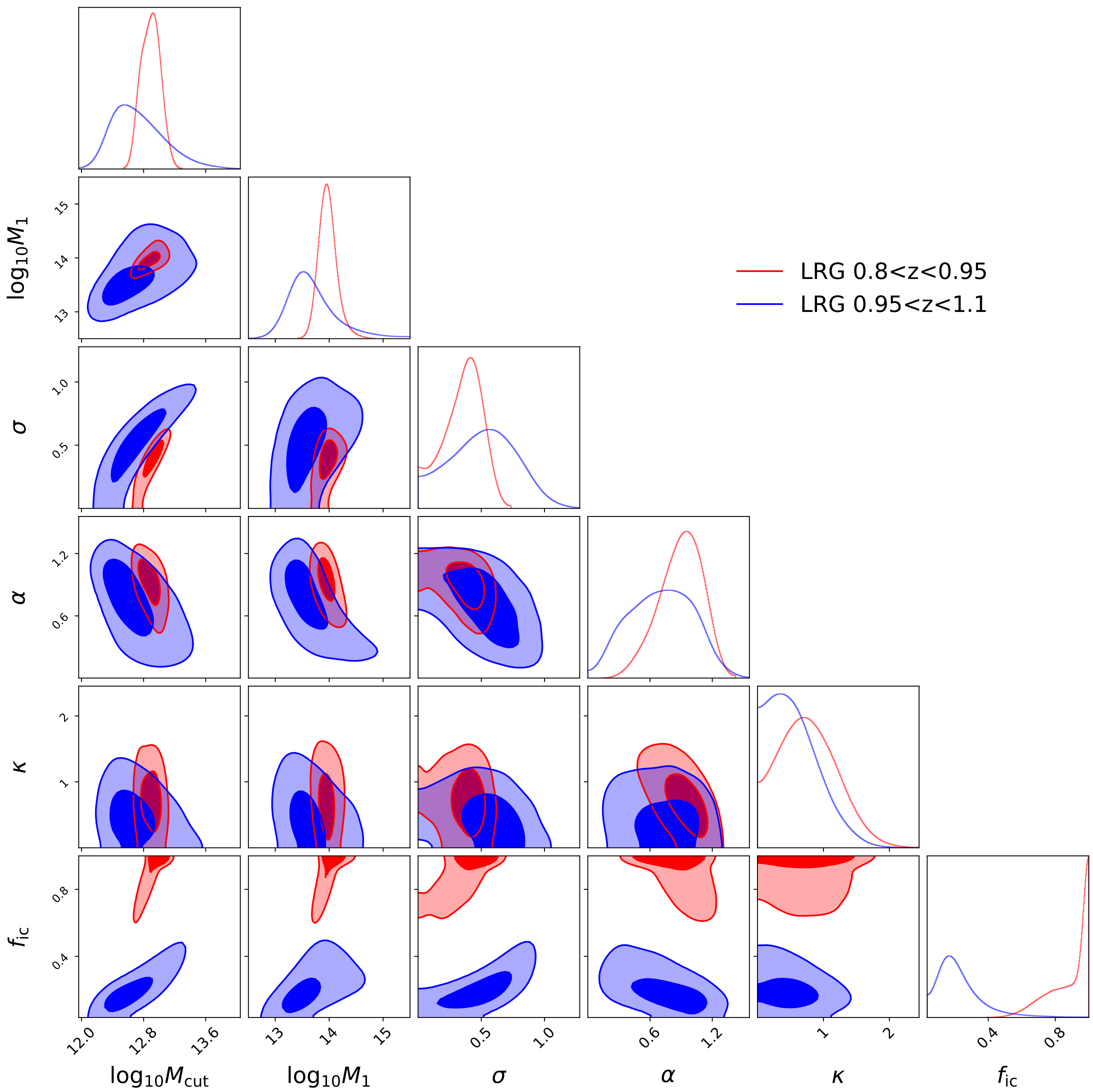


Figure 3.6.5: The DESI One-Percent survey LRG at $z > 0.8$ HOD posterior. The results from $0.8 < z < 0.95$ and $0.95 < z < 1.1$ are shown in red and blue respectively. The contours represent 68 and 95% confidence levels. 1D marginalized distribution for each parameter is shown at the top of each column.

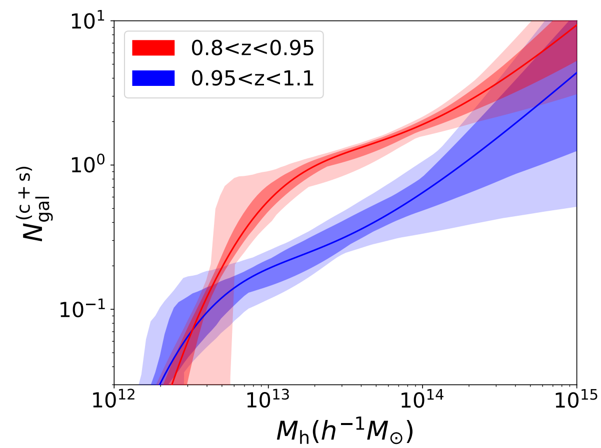


Figure 3.6.6: The HOD posterior band (central+satellite) of LRG sample at $z > 0.8$. The results from $0.8 < z < 0.95$ and $0.95 < z < 1.1$ are shown in red and blue respectively. The shaded regions correspond to 1 and 2 σ posteriors.

degeneracies between $\log M_{\rm cut}$ and σ, $\log M_{\rm cut}$ and $\log M_1$ as LRGs at $z < 0.8$. The 1D distribution depicts a change in the mean value of each parameter in the Z07 model as the redshift increases. This trend is consistent with the comparison of the $0.4 < z < 0.6$ and $0.6 < z < 0.8$ bins in Figure 3.6.2. The most notable difference is observed in the incompleteness parameter $f_{\rm ic}$. As $f_{\rm ic}$ can effectively control the number density of the mock galaxies, it decreases significantly as the number density of the sample decreases, indicating an increase in incompleteness.

The results of our analysis provide a compelling reason for conducting a detailed study of the HOD of high-redshift LRGs. As shown in Figure 3.6.6, the 1 and 2 σ uncertainty bands of the HOD function for the high-redshift LRG sample have a minimal overlap for the range $12.8 < \log M_{\rm halo} < 14.2$, indicating a significant difference in the HOD between the two redshift intervals. Furthermore, Table 3.6.4 summarizes the marginalized statistics for the high-redshift LRG main sample, which show differences in the mean completeness $f_{\rm ic}$ (from 92% to 19%), the satellite fraction $f_{\rm sat}$ (from 11.0% to 15.1%), mean halo mass $\log_{10} \overline{M}_{\rm h}$ (from 13.29 to 13.00), and linear bias $b_{\rm lin}$ (from 2.31 to 2.13). These results indicate that the HOD of the high-redshift LRG main sample evolves with redshift, which we will discuss in detail in the following section.

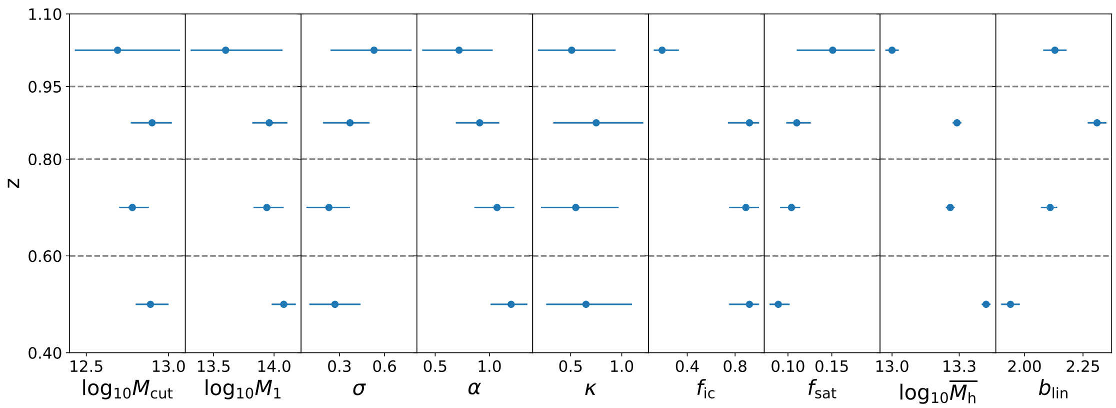


Figure 3.6.7: The marginalized results of HOD and derived parameters of LRG main sample for each redshift bin. Markers and error bars show the mean of the fits and 1-σ error. The numerical results of this plot are also listed in Table 3.6.3 and 3.6.4.

3.6.3 Redshift evolution of DESI LRG HOD

The evolution of HOD and derived parameters of LRGs across all redshift bins are shown in Fig. 3.6.7. We only use data vector $w_{\rm p} + n(z)$ in this part of the analysis for consistency. The central galaxy host halo mass threshold ($M_{\rm cut}$) does not exhibit any significant trend with redshift within the uncertainties. The scatter in the halo mass threshold (σ) tends to increase with increasing redshift, indicating a smoother transition in the central HOD on high redshift. In terms of satellite parameters, both M_1 and α show a declining trend with redshift, while κ shows little variation with redshift and is weakly constrained by the data.

The satellite fraction ($f_{\rm sat}$) shows a mild trend of increasing with redshift at $z < 0.95$, rising from $f_{\rm sat} \simeq 9\%$ at $z \simeq 0.5$ to $f_{\rm sat} \simeq 11\%$ at $z \simeq 0.9$. This increase in satellite fraction can be interpreted as a result of the merging of galaxies over time. At $z > 0.95$, the satellite fraction increases substantially to 15%, but the significance is low.

The mean halo mass does not exhibit a clear trend except at $z > 0.95$, where the mean halo mass drops off substantially. A similar behavior is seen for the galaxy bias, which increases with redshift at $z < 0.95$ (consistent with studies of DESI-like photometric LRGs in[40]) but drops off at $z > 0.95$. Furthermore, the highest redshift bin also exhibits a drastic drop in completeness. These comparisons serve as strong evidence that the LRG sample at $z > 0.95$ is physically different from the LRGs at $z < 0.95$.

[141] showed that the DESI LRG selection should yield a highly stellar mass complete sample at $\log_{10} M_* > 11.5$ between redshift $0.4 < z < 0.9$. At $z > 0.9$ the completeness at the high mass end starts to deviate from 1 and drops further at $z > 1$. This point is confirmed by explicitly computing the stellar mass functions of DESI SV3 LRGs via SED fitting in Gao et al. in prep, where a substantial incompleteness at $\log_{10} M_* > 11.5$ is observed at $z > 1$. This incompleteness at the high mass end can partially explain the drop in the inferred bias and halo mass in the highest redshift bin.

Another potential contributing effect is that the DESI LRG selection employs a sliding color-magnitude cut in $r - W1$ vs $W1$ that turns over at $W1 =\sim -19$ to include more galaxies at the high redshift end (see Figure 3 of[40]). However, this turnover also includes red galaxies with very faint $W1$ magnitudes into the LRG sample, thus possibly decreasing the mean halo mass and mean galaxy bias.

Separately,[186] inferred star formation histories from DESI SV1 LRG SEDs and found evidence that recently quenched (post-starburst) galaxies constitute a growing fraction of the massive galaxy population with increasing redshift. The study showed that these galaxies are significantly brighter than the parent LRG sample at fixed stellar mass. Thus, at fixed brightness, these galaxies likely have lower halo masses and lower biases compared to the parent LRG sample. If these galaxies are indeed a significant fraction of high redshift LRGs, then they would contribute to the drop in the mean halo mass and bias. This could also explain the relatively high satellite fraction as recently accreted satellites are also likely to have been recently quenched.

An important caveat to consider in the interpretation of the mean halo masses relates to the use of fixed redshift snapshots in this analysis. To obtain an unbiased HOD parameter estimation, the mean redshift of the redshift bin must remain close to the snapshot redshift. However, the choice of snapshots is limited to several primary redshifts provided by the AbacusSummit simulation. For instance, we use the snapshot at $z = 0.8$ for both $0.6 < z < 0.8$ and $0.8 < z < 0.95$ redshift bins, resulting in mean sample redshifts that are lower and higher than the snapshot redshift, respectively.

As a result, the mean halo mass has been overestimated for $0.8 < z < 0.95$ and underestimated for $0.6 < z < 0.8$. This effect is particularly pronounced for the $0.95 < z < 1.1$ redshift bin, as the drop in number density results in a more significant bias of the mean sample redshift from the snapshot redshift of $z = 1.1$. Mitigating this effect could, to some extent, resolve the non-monotonic shape in the mean halo mass evolution. Nevertheless, we maintain that this effect does not fully explains the drop in halo mass in the highest redshift bin because the drop in linear bias is agnostic to simulation snapshots as the underlying clustering amplitude is computed at the mean redshift of the sample instead of the simulation snapshot.

This systematic highlights the need for future redshift evolution studies to use more accurate redshifts. The shift in the mean value of HOD and derived parameters also provide a strong scientific incentive to conduct a detailed study of LRGs using the halo light-cone catalogs in combination with a redshift-evolved HOD model.

3.7 QSO HOD results

We analyze the QSO sample following the same procedure. First, we construct the mock-based covariance matrix by fitting a baseline HOD on the QSO $\xi(r_\mathrm{p}, r_\pi)$measurement. However, we find that directly fitting $\xi(r_\mathrm{p}, r_\pi)$returns poor fits because the QSO $\xi(r_\mathrm{p}, r_\pi)$measurement below $r_p \sim 1h^{-1}$Mpc have particularly low signal-to-noise, and the corresponding jackknife errors do not behave properly.

Instead, we fit just the projected w_p with the 5-parameter+incompleteness model, where we do achieve a good fit. Then we independently tune the two velocity bias parameters to match the $\xi(r_\mathrm{p}, r_\pi)$signal at $r_p > 1h^{-1}$Mpc. We again obtain a good fit. With that, we populate the 1800 small boxes and generate a mock-based covariance matrix for QSO $\xi(r_\mathrm{p}, r_\pi)$.

Again, we sample the baseline HOD parameter to obtain marginalized posteriors for the QSO HOD. The results are visualized in Figure 3.7.1 and summarized towards the bottom of Table 3.6.3. Figure 3.7.3 shows the corresponding HOD posterior. In general, the QSO HOD parameters are much less constrained than the LRG parameters due to the limited sample size. For the same reason, we also do not test additional model extensions as we already achieve an excellent best-fit χ^2 with the baseline model. We will conduct such tests when a significantly larger sample of QSOs becomes available. We also showcase the w_p-only constraints in Table 3.6.3, and we find them to be consistent with the $\xi(r_\mathrm{p}, r_\pi)$constraints.

The HOD constraints compare well with those inferred for eBOSS QSOs as presented in Table 1 of Alam et al.[38]. Specifically, they found $\log M_\mathrm{cut} = 12.2$ and $\log M_1 = 14.1$, consistent with our findings. Perhaps the most unexpected parameter constraint is the central velocity bias, where we find $\alpha_c = 1.54^{+0.17}_{-0.08}$, meaning that the central galaxies exhibit large peculiar velocities relative to the halo center. There are several potential explanations for this. While one possibility is that it points towards energetic processes within the AGN, we speculate that this may also be due to the rather large redshift uncertainties with the QSO sample[187]. Specifically, QSO primary spectral lines such as Mg II line to C IV suffer from large systematic velocity shifts caused by astrophysical effects e.g.,[188–190]. Another potential explanation is that QSOs preferentially occupy recently merged halos, resulting in large velocity dispersion relative to the halo center of mass. In the following paragraph, we also show that uncertainties in satellite fraction can also be degenerate with velocity bias. Thus, the large velocity bias in the QSO sample could be due to redshift uncertainties, AGN physics and mergers, and/or uncertainties in satellite fraction.

In terms of derived quantities, we infer a mean halo mass of $\log_{10} \overline{M}_h = 12.65^{+0.09}_{-0.04}$ for the DESI QSO sample, consistent with the $\log_{10} \overline{M}_h = 12.7$ found for the eBOSS QSO sample in both[191] and[38]. We report a linear bias of $b_\mathrm{lin} = 2.63^{+0.37}_{-0.26}$ at $z \simeq 1.5$ for the DESI QSO sample, which is slightly higher than the first-year eBOSS quasar sample with $b_\mathrm{lin} = 2.45 \pm 0.05$ at $z = 1.55$, as found by Laurent et al.[192]. However, considering the error bar, the results are consistent. We infer a satellite fraction of $3^{+8}_{-2}\%$ for the QSO sample, consistent with $f_\mathrm{sat} = 5\%$ found in[191] using a subhalo abundance matching model but significantly lower than the 30% inferred with a multi-tracer HOD fit in[38]. The large satellite fraction inferred in[38] is directly a result of a small inferred $\alpha = 0.4$, which results in a large number of satellites in low-mass halos. However, we speculate that this result might in fact be degenerate with our finding as most of these satellites in low-mass halos do not have companion centrals

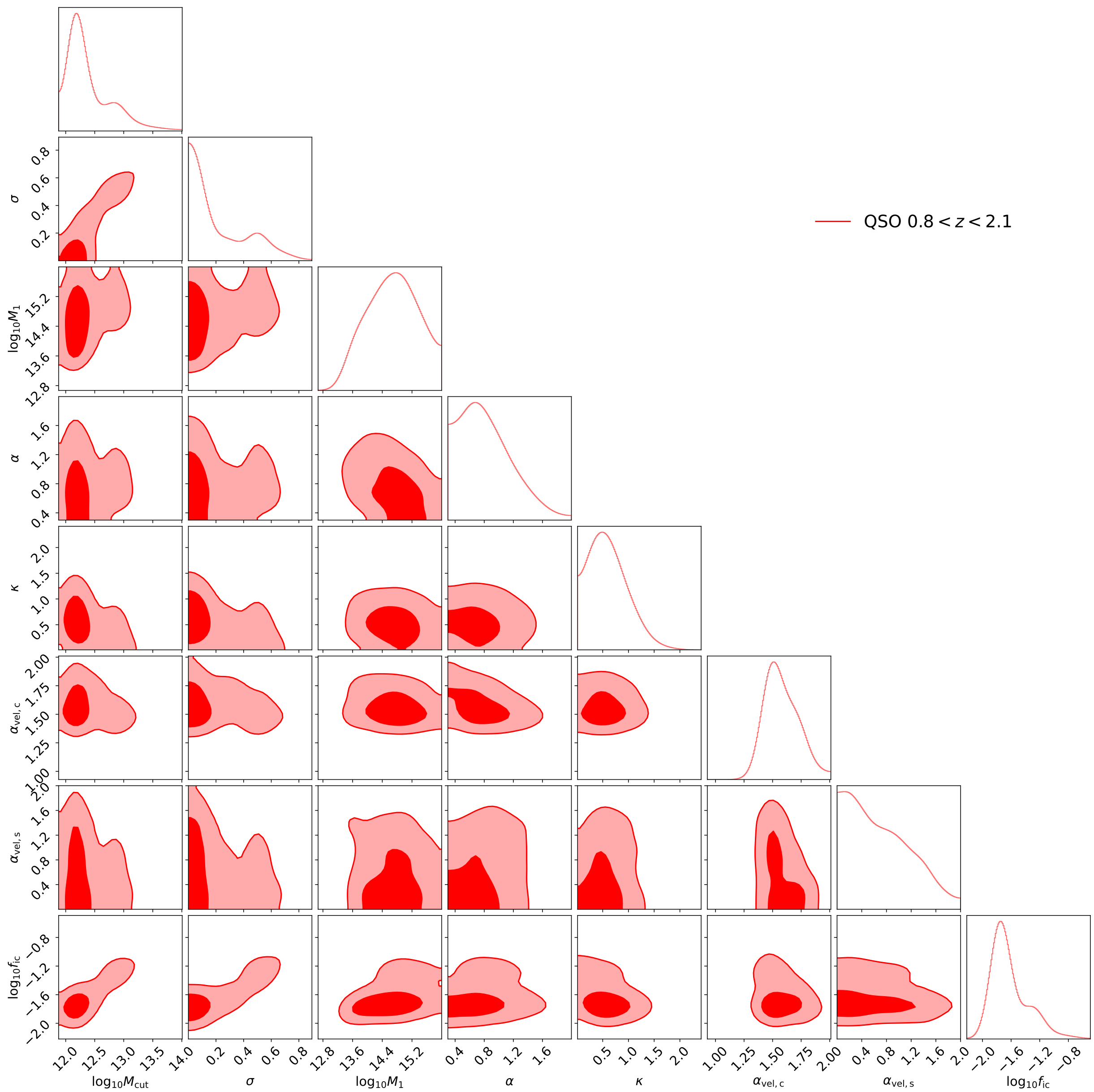


Figure 3.7.1: The DESI One Percent Survey QSO HOD posterior. The red contours correspond to the 1 and 2 σ marginalized posteriors for QSOs in the $0.8 < z < 2.1$ bin.

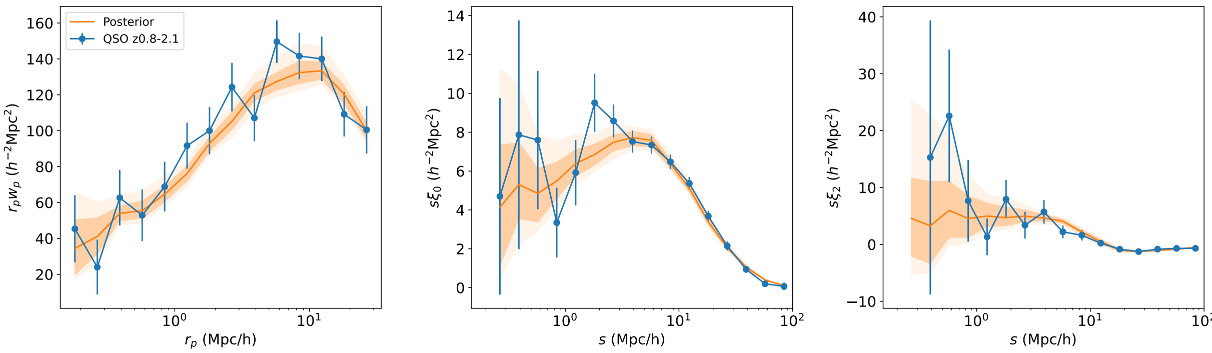


Figure 3.7.2: The QSO w_p and multipoles posterior predictives compared to the data. The blue lines correspond to the One Percent Survey measurement with jackknife error bars. The solid orange line denote the prediction corresponding to the posterior mean. The orange-shaded regions correspond to the 1 and 2σ full posterior using the full mock covariance matrix.

due to the low completeness. These "rogue" satellites are ill-defined in the vanilla HOD context, and can simply be re-classified as centrals. This could also be connected to the large central velocity bias we found, which could disappear if we re-classify some centrals as satellites, which would naturally have larger velocity dispersion. Our results compare well with earlier SDSS Quasar HOD fits.[193] obtained a satellite fraction of 7-10% and found the inferred satellite fraction to be dependent on the assumed HOD model.[42] obtained a lower satellite fraction of $\sim 0.1\%$ for $z \sim 1.4$ quasars.

These types of questions also show that the QSO–halo connection physics is poorly understood. In fact, we use the standard 5-parameter model in this analysis simply because we have no evidence that a different model is favored. We reserve a more comprehensive analysis of QSO HOD for a future paper when a much larger sample of QSOs becomes available.

3.8 Mock products

We apply the best-fit HODs obtained for the LRG and QSO samples to all 25 base boxes available in AbacusSummit at Planck cosmology to create high-fidelity mocks. For each tracer at each redshift snapshot, the total sample volume is $200h^{-3}$Gpc^3 comoving. The volume provided by AbacusSummit is an order of magnitude larger than other simulations

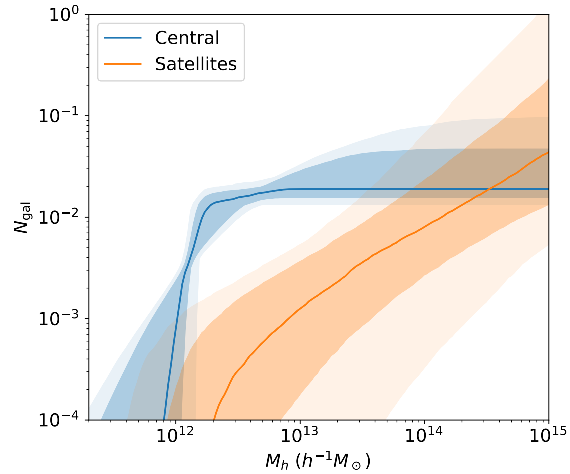


Figure 3.7.3: The HOD posterior for the QSO sample. The shaded regions correspond to 1 and 2σ posteriors.

of the comparable resolution, reaching 5-10 times the volume expected to be observed by DESI. Given the volume and resolution, these mocks are critical for testing and calibrating DESI cosmology pipelines at the necessary precision.

In addition to using just the best-fit HODs, we also create additional mocks where we perturb the HOD parameters around the best-fit to generate mocks that share the same cosmology but differ in bias prescriptions. The perturbations are sampled from the 3σ region of the parameter space around the best-fit values. We repeat this procedure for both the baseline HOD model and an extended HOD model that also includes environment-based assembly bias B and satellite radial profile parameter s, resulting in a set of mocks that encompass a diverse range of possible HODs.

These variety mocks enable key robustness tests of large-scale cosmology inference pipelines against galaxy–halo connection systematics. Specifically, large-scale cosmology pipelines that utilize the BAO/RSD features or full-shape information often assume a much simpler bias model. As a result, complexities in the galaxy–halo connection modeling could become degenerate with cosmology and thus result in systematic bias in the inferred cosmology. Thus, it is essential to test cosmology inference pipelines against a range of mocks with varying bias models and demonstrate that cosmology inference remains unbiased. We defer a detailed discussion of these tests to a dedicated paper[194].

In addition to constructing cubic mocks, we utilize our best fits to construct redshift-dependent mocks on the AbacusSummit lightcones[168]. The benefit of having mocks on the lightcone is that they provide an accurate synthetic map of the sky, which is crucial for testing out systematic and observational effects such as fiber collisions. Lightcones also enable explicit modeling of redshift evolution, such as in the forward modeling pipelines being developed for novel summary statistics e.g.,[147;195]. Our procedure for generating lightcone mocks is as follows:

1. Adopting the AbacusHOD algorithm, we first subsample the halo lightcone catalogs assuming the same envelope as the cubic boxes and pre-compute various assembly bias parameters and decorations to the HOD model.

2. We read in the best-fit parameters as we found in section 3.6. We linearly interpolate the HOD parameters as a function of redshift, pivoting on the two best-fits in the two redshift bins.

3. Finally, we generate the galaxy mocks for all tracers/the LRGs at all available redshifts of the 25 fiducial cosmologies, i.e. `AbacusSummit_c000_ph000-024`. We note that due to the geometry, each of our mocks provides an octant of the sky until $z \approx 0.8$, decreasing gradually as we go to higher redshifts.

All said mocks will be made publicly available at a future date as a part of DESI EDR.

3.9 Conclusions

In this paper, we present a comprehensive analysis of the halo occupation distribution of the DESI One Percent Survey LRG and QSO samples using the AbacusSummit cubic boxes.

For LRGs, we study the sample in two fiducial redshift bins $0.4 < z < 0.6$, $0.6 < z < 0.8$, and also in a third high redshift bin $0.8 < z < 1.1$. In the fiducial bins, we compare the baseline HOD model with extended models with galaxy assembly bias and satellite profile bias. We find no evidence for model extensions at current precisions and the baseline model is favored by the data. For both redshift bins, we constrain the baseline parameter posteriors with the $\xi(r_\mathrm{p}, r_\pi)$data vector and mock-based covariance matrix. The resulting model posteriors produce the correct redshift-space clustering. We find strong constraints on inferred properties such as the average halo mass and the satellite fraction, which are broadly consistent with results from eBOSS and BOSS. We also find consistency between $\xi(r_\mathrm{p}, r_\pi)$fit and w_p-only fit. The marginalized posterior constraints are summarized in Table 3.6.3. To highlight a few key constraints: the LRG sample in $0.4 < z < 0.6$ yields a satellite fraction of $11 \pm 1\%$ and a mean halo mass of $\log_{10} \overline{M}_h = 13.40^{+0.02}_{-0.02}$, whereas the $0.6 < z < 0.8$ sample results in a satellite fraction of $14 \pm 1\%$ and a mean halo mass of $\log_{10} \overline{M}_h = 13.24^{+0.02}_{-0.02}$ in $0.6 < z < 0.8$.

Combining the fiducial analysis at $z < 0.8$ and high redshift analysis at $z > 0.8$, we find clear trends of evolution, especially in physical parameters like satellite fraction and mean halo mass. Specifically the mean halo mass decreases with redshift whereas the satellite fraction increases with redshift. This motivates future redshift evolution studies that should shed light on the physics of the evolution of massive galaxies. The marginalized posterior constraints are summarized in Table 3.6.4.

The QSO sample is limited by sample size, and we are not able to conduct meaningful comparisons between different HOD models. Regardless, we derive good fits on the data and present posterior constraints. The marginalized posterior constraints are summarized in Table 3.6.3. We infer a satellite fraction of $3^{+8}_{-2}\%$ and a mean halo mass of $\log_{10} \overline{M}_h = 12.65^{+0.09}_{-0.04}$ in redshift range $0.8 < z < 2.1$. The inferred mean halo mass is consistent with previous results, but there is some discrepancy in the satellite fraction. We speculate that such discrepancy is model dependent and we intend to revisit of this issue when a significantly larger sample of QSOs becomes available.

Finally, we leverage our HOD fits to generate a large suite of DESI-like mocks. We highlight mocks with varied HODs that test the robustness of large-scale cosmology pipelines, and lightcone-based mocks that are important for building realism and testing observational systematics.

Acknowledgements

This material is based upon work supported by the U.S. Department of Energy (DOE), Office of Science, Office of High-Energy Physics, under Contract No. DE–AC02–05CH11231, and by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract. Additional support for DESI was provided by the U.S. National Science Foundation (NSF), Division of Astronomical Sciences under Contract No. AST-0950945 to the NSF's National Optical-Infrared Astronomy Research Laboratory; the Science and Technologies Facilities Council of the United Kingdom; the Gordon and Betty Moore Foundation; the Heising-Simons Foundation; the French Alternative Energies and Atomic Energy Commission (CEA); the National Council of Science and Technology of Mexico (CONACYT); the Ministry of Science and Innovation of Spain (MICINN), and by the DESI Member Institutions: https://www.desi.lbl.gov/collaborating-institutions. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U. S. National Science Foundation, the U. S. Department of Energy, or any of the listed funding agencies.

The authors are honored to be permitted to conduct scientific research on Iolkam Du'ag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation.

Data Availability

The simulation data are available at https://abacussummit.readthedocs.io/en/latest/. The AbacusHOD code package is publicly available as a part of the abacusutils package at http://https://github.com/abacusorg/abacusutils. An example usage can be found at https://abacusutils.readthedocs.io/en/latest/hod.html. All mock products will be made available at https://data.desi.lbl.gov.

The MCMC chains generated, along with the clustering measurements used in this study - encompassing correlation functions and covariance matrices - are available in a machine-readable format at https://doi.org/10.5281/zenodo.7972386.

Appendix

3.A Choice of prior

The selection of priors for the HOD parameters can potentially impact the inferred constraints. In order to eliminate extreme values of the HOD parameters and gain a more comprehensive understanding of the Galaxy-Halo connection model, Gaussian priors shown in Tab. 3.6.2 were applied to the HOD parameters in the primary analysis. In order to assess the potential impact of these priors on the results, additional MCMC analyses were conducted using only flat priors.

In this test, we employ the Z07+$f_{\rm ic}$ model to fit $w_{\rm p}+n_z$ of LRG at redshift $0.6 < z < 0.8$. Figure 3.A.1 illustrates the 1 and 2 σ confidence level contours for the HOD parameters when different priors are applied. It is clear that the fit with a Gaussian prior displays a tighter contour, particularly for the satellite parameters $\log_{10} M_1$, α, and κ. The Gaussian prior helps to eliminate isolated regions in the contour, where the data have limited power to constrain the parameters. The 1D distribution of each parameter in Figure 3.A.1 indicates that the Gaussian prior has a smaller impact on the central parameters as compared to the satellite parameters and hardly any impact on incompleteness.

Figure 3.A.2 shows the 2 σ uncertainty band of HOD posterior and best fit for both cases. The two bands overlap heavily with each other. The Gaussian prior fit has a slightly narrower band for $\log_{10} M_{\rm cut} > 14.2$, where the satellite parameters have a larger impact on the HOD. Additionally, the Gaussian prior presents a smoother best fit than the flat prior, which exhibits a distinct step on the lower mass end. As a result, our choice of Gaussian prior help removes nonphysical HOD without altering any of our main conclusions.

Finally, we expect the difference from choices of priors to be further reduced as we achieve tighter constraints with more precise measurements using larger samples in the future.

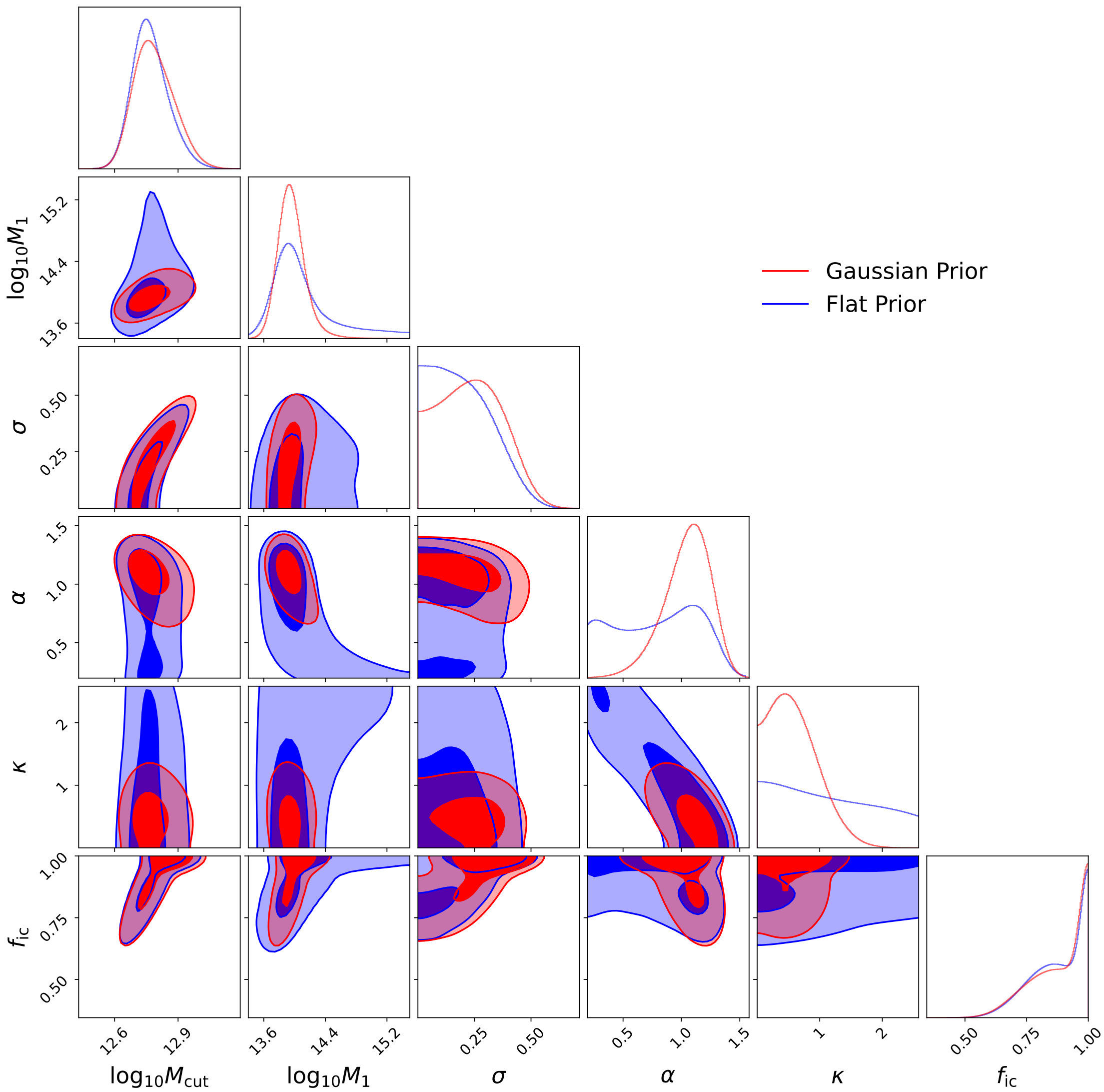


Figure 3.A.1: Marginalized probability distribution of HOD parameters for LRG sample at $0.6 < z < 0.8$ with and without Gaussian priors listed in Tab. 3.6.2. The results using Gaussian prior and flat prior are shown in red and blue respectively. The contours represent 68 and 95% confidence levels. 1D marginalized distribution for each parameter is shown at the top of each column.

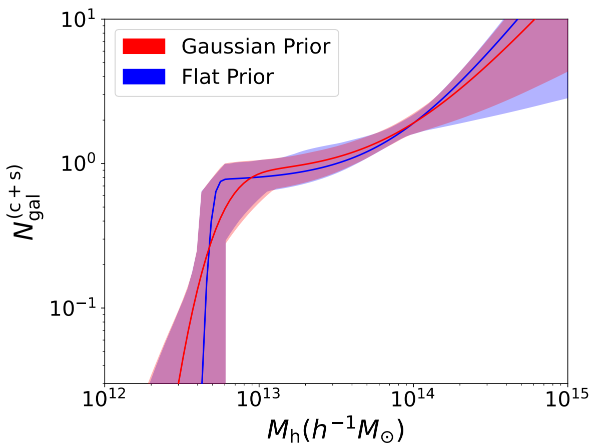


Figure 3.A.2: 2-σ band of LRG sample HOD at $0.6 < z < 0.8$ with and without Gaussian prior. The red is the 95% CL uncertainty from fit using Gaussian priors, the blue is the 95% CL uncertainty from fit using flat priors. Lines are the corresponding best fit.

Chapter 4

Extracting cosmological information from the small-scale clustering

4.1 Introduction

Extracting cosmological information from clustering is a crucial task in modern cosmology. One of the primary cosmological probes being used to provide constraining power is the redshift-space distortions (RSD). RSD arises because the observed position of a galaxy in redshift space is distorted along the line of sight due to its peculiar velocity, which itself is a direct consequence of the gravitational growth of structures in the universe. The major information derived from the analysis of redshift-space clustering is the cosmological parameter pairing $f\sigma_8$. Here, f represents the growth rate of cosmic structures as determined from linear perturbation theory, while σ_8 is a measure of the amplitude of the matter power spectrum on a scale of 8 h^{-1}Mpc. Together, this combination, $f\sigma_8$, allows us to probe the rate at which structures in the universe grow and the overall scale of these structures.

Perturbation theory has been a valuable tool for modeling redshift space clustering[196–206]. It provided a theoretical framework for understanding the distribution of galaxies on large scales, where the dynamics are predominantly linear. However, as we delve into smaller scales, the dynamics of the universe become increasingly non-linear, and the simplicity of perturbation theory falls short. On these scales, effects such as non-linear gravitational clustering and non-linear bias, as well as the Finger of God effect, where the virial motions of galaxies within clusters elongate the structures along the line of sight in redshift space become significant[207;208]. These complexities introduce higher-order terms that the perturbative approach struggles to accurately account for. These non-linear factors can convolute the cosmological signals we are interested in, making it difficult to accurately model and interpret the observed clustering.

Yet, as demonstrated by studies conducted by Reid et al.[209] and Zhai et al.[210], the ability to constrain parameters through redshift-space distortions (RSD) analysis significantly enhances when we incorporate information from the non-linear regime. This highlights the potential value and importance of tapping into the rich details embedded in these non-linear scales. Therefore, it becomes necessary to directly model the relationship between galaxies and the dark matter halos that host them. HOD framework, which successfully models galaxy-halo connection, is an essential tool for interpreting small-scale clustering measurements, as it encompasses the complex, non-linear processes of galaxy formation and evolution within dark matter halos.

Leveraging high-resolution dark matter simulation and the HOD framework, numerous studies have aimed to develop simulation-based forward modeling tools to facilitate precise Redshift-Space Distortion (RSD) analysis on scales as small as sub $1h^{-1}$Mpc.

For instance, Lange et al.[133] conducted an analysis on anisotropic redshift-space clustering in the scale range from 0.4 to 63 h^{-1}Mpc based on the Cosmological Evidence Modeling (CEM) framework[211]. In another study, Kobayashi et al.[134] combined an emulator of halo clustering with the HOD model and performed a full-shape analysis of redshift-space power spectra of the BOSS DR12 galaxy catalog, resulting in a 5% determination of σ_8. Yuan et al.[136] applied a hybrid MCMC + emulator approach to the BOSS CMASS data, inferring σ_8 constraints from the small scale. Zhai et al.[137] used an emulator based on the Aemulus Project[212] to obtain cosmological constraints using correlation function multipoles at scales of 0.1 to 60 h^{-1}Mpc.

With their unprecedented volume and wealth of data, large-scale surveys such as DESI,

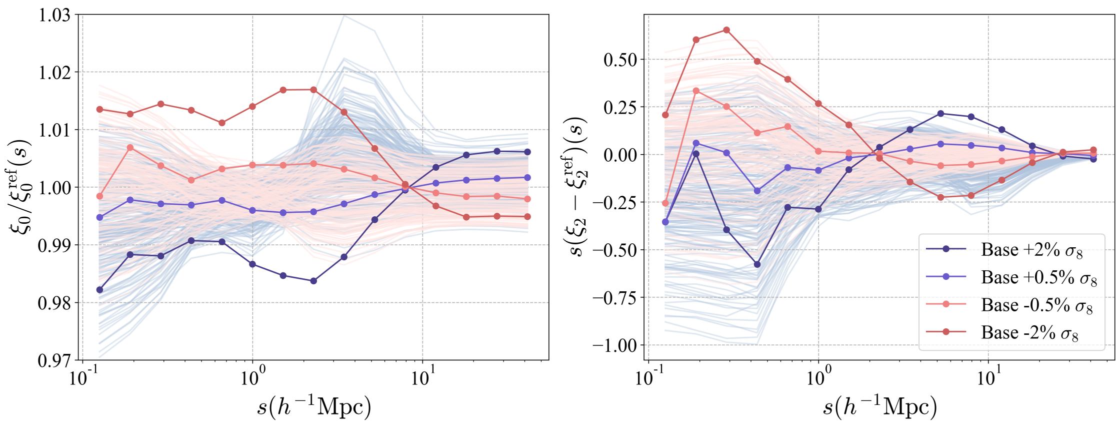


Figure 4.2.1: Ratio of small scale correlation function monopole (left) to the reference value (computed in fiducial cosmology and HOD), and difference of correlation function quadrupole (right). Blue and red bend corresponding to 1-σ HOD variance on fiducial cosmology based on 2PCF only and both 2PCF and 3PCF.

PSF[213], and Euclid[214] offer unparalleled opportunities to leverage these small-scale signals for cosmological insights.

4.2 Testing degeneracy between cosmological and HOD parameters

Extracting cosmological information from small-scale clustering is one of the research avenues I would like to explore in the next few years. I plan to build a tool for predicting the small-scale clustering based on the `AbacusSummit` simulations and reduce the uncertainty in the HOD parameters by combining higher-order statistics and multi-tracer techniques[215;216]. DESI has four distinct target classes that partially overlap in redshift, creating an excellent opportunity for multi-tracer analysis. Furthermore, the number densities of low redshift tracers are high, making higher-order measurements more effective.

Fig. 4.2.1 presents some preliminary tests based on the mock LRG results discussed in Chap. 2. The blue band indicates a 1-σ variance in the small-scale 2PCF multipole of mock LRGs due to changes in HOD under the fiducial cosmology. The introduction of higher-order information, represented by the red band, reduces HOD variance. Mocks were also populated using fiducial HOD parameters, as shown in Tab. 2.1, for cosmology with differing σ_8 values. The lines correspond to these different σ_8 values.

For the monopole, it's evident that σ_8 is completely degenerate with HOD parameters on a scale larger than $10h^{-1}$Mpc. However, on a scale smaller than $10h^{-1}$Mpc, σ_8 demonstrates different patterns within the HOD band. Regarding the quadrupole, 2% σ_8 differences overlap with the 2PCF-only HOD band, making it difficult to break the HOD degeneracy. However, a combination of projected 2PCF and 3PCF could narrow the HOD band and show no overlap with 2% difference lines on a scale of $2 < s < 10h^{-1}$Mpc.

This plot suggests that, with the extra constraining power provided by the projected 3PCF on HOD parameters, σ_8 could be better constrained using the 2PCF multipole. I plan to perform this analysis on DESI Year 1 data, once the effects of fiber assignment are thoroughly understood.

Chapter 5

Jackknife co-variance matrix for various two-point statistics

This chapter originally appeared in the literature as

Hanyu Zhang et al. DESI Mock Challenge: Jackknife co-variance matrix for various two-point statistics.

Chapter Abstract

The covariance of large-scale structure clustering measurements is typically computed by taking a sample variance of a large number of simulations. An alternative internal resampling method, the Jackknife technique, has the potential to reduce computational resources and does not require an assumption of the fiducial cosmology. We investigate the usage of the Jackknife method in two-point statistics measurements of the correlation function and the power spectrum multipoles in the context of the Dark Energy Survey Instrument experiment. We made methodological improvements for the application of the Jackknife method to power spectrum multipoles that reduce potential biases resulting from the window function. We check the robustness of Jackknife variance with respect to a number of choices that are difficult to make based on first principles, such as the number of Jackknife regions and the choice between two-dimensional and three-dimensional regions. We also study the robustness of different

clustering measurements. Our findings show that the Jackknife method successfully recovers the variance of the monopole for the correlation function but overestimates the variance on large scales for the quadrupole and hexadecapole by up to 20%. The Jackknife variance for the power spectrum needs to be re-scaled by a simulation-independent factor to agree with the reference value. We also compare covariance matrices obtained with the Jackknife method and simulations by investigating the ability of cosmological parameter constraints.

5.1 Introduction

Large spectroscopic galaxy surveys are essential for understanding the expansion history and structure growth of the universe. The Dark Energy Spectroscopic Instrument [104;217] (DESI) is a prime example of such a survey. Over the next five years, it will collect approximately 35 million galaxy and quasar spectra, covering more than 14,000 deg^2 of the sky, and achieve percent-level measurements of key cosmological parameters. Two-point statistics from DESI will provide exquisite cosmological information, specifically the two-point correlation function (CF) and its Fourier space counterpart, the power spectrum ($P(k)$). These measurements are crucial for studying cosmology and understanding the large-scale structure of the universe.

The two-point correlation function (CF) is typically measured by binning the galaxy pairs as the function of separations, while the power spectrum ($P(k)$) is binned by wavenumber. However, these measurements are inherently stochastic due to the finite volume and sparse sampling of galaxies in the survey. In addition, the measurements are correlated due to gravitational evolution and the effects of the survey window function. To accurately conduct cosmological analyses based on these measurements, it is essential to have a robust and trustworthy estimation of the covariance matrices. These matrices capture the correlations between different measurements and are crucial for properly interpreting the data and extracting cosmological information.

The covariances of the CF and $P(k)$ are usually estimated using simulations. The standard approach involves generating a large number of simulations that are well-matched with the data and using them to compute the CF and $P(k)$. The sample variance of these measurements is then used as an estimator of the data covariance, indicating how much the measurements could vary across different random realizations of the data. However, this method has several potential drawbacks. The sample variance of an ensemble of measurements is itself a stochastic variable that can be a noisy realization of the true covariance. To make this noise negligible, a large number of high-fidelity simulations are required, which can be computationally expensive. Moreover, these simulations need to be run in a fixed cosmological model that is usually chosen to ensure the resulting simulations are similar to the data. However, this choice can be difficult to make in a consistent manner and may introduce bias into the analysis[218;219].

Recent advances in approximate mock generation methods, such as the log-normal method[220], pinpointing orbit-crossing collapsed hierarchical objects[221;222], Patchy[223;224], PThalos[225–227], COLA[228], QPM[229],EZmocks[230], and FastPM[231], have significantly reduced the computational cost associated with simulations. However, even with these techniques, generating an appropriate number of simulations for a large survey like DESI can be challenging.

To address this challenge many alternative methods of estimating covariance have been proposed in the literature. They include analytic estimates[232;233], shrinkage estimation[234], the covariance tapering method[235], and fitting functions[236] In this paper we study the Jackknife resampling method for estimating covariance matrices.

The Jackknife method is a well-known resampling technique used for estimating biases and covariances in various fields of physics and applied statistics. In cosmology, it has been widely used to estimate CF covariances, especially for small-scale. Studies by Loh[237] have demonstrated that resampling techniques can work effectively for estimating n-point statistics measurements of point distributions. More specifically, Norberg et al.[114] have shown that the Jackknife method can be used to estimate CF errors, although it tends to overestimate them on smaller scales. Other studies, such as the work by Friedrich et al.[238] in the context of weak lensing shear measurements and Favole et al.[239] for the SDSS-III BOSS CMASS galaxy sample, have also investigated the effectiveness of resampling methods for estimating CF covariances. Recently, Mohammad and Percival[166] have made suggestions for improving the standard Jackknife estimators for CF measurements.

When using the Jackknife method to estimate covariance, several methodological decisions must be made, and these are not always easy to justify from the first principles. For example, in the case of CF measurements, the size and geometry of the region excluded from each iteration can be selected in various seemingly equivalent ways. Natural choices include a small sub-cube in the 3-dimensional distribution of galaxies or a 2-dimensional square column along one of the directions. However, for $P(k)$ measurements, removing part of the volume induces window effects that can introduce bias in the estimates. In this paper, we investigate how the performance of the Jackknife resampling method varies with these choices and propose a method for minimizing the effect of the induced window for $P(k)$ Jackknife covariance.

Rephrasing above paragraph: One of the fundamental requirement for validating approximate models is to have access to ideal measurement with all the non-linearity included. The only possible way to include non-linearity is using simulation which has finite volume and hence have associated cosmic variance. This often sets the limit about how precisely we can test our models. For many of the systematic studies we might be interested in the shift of parameter values as oppose to absolute value due to some change in the simulated universe. One example where such technique is used is to test the accuracy of N-body codes where we initiate each code with exactly the same initial condition but run them through different N-body code[240]. In such a scenario the error on any measurements are large but the error in the difference is much smaller as most of the comic variance cancels out. Estimating errors in such scenario through several realizations will be very expensive and no analytical prescription have been shown to work. The methods developed in this paper gives an easy way to obtain accurate covariance matrices which was applied in the Grove et al.[240] and will continue to be very useful for ongoing and future analysis preparations.

This paper is structured as follows. In Sec. 5.2, we introduce how the sample covariance matrix is estimated from a set of simulations and describe the Jackknife resampling technique. In Sec. 5.3, we introduce the simulation suite, clustering measurements, and the combination of configurations we studied. In Sec. 5.4, we present the performance of the Jackknife method on both CF and $P(k)$ compared to the reference method and study the impact of parameter constraint. Finally, we summarize our findings and conclusions in Sec. 5.5.

5.2 Estimating Co-variances

5.2.1 Reference

Let's say we have N_{mocks} number of mocks from which we make several measurements X_i and Y_i. X and Y can be a CF, a $P(k)$, or a difference of CF or $P(k)$ from a pair of simulations in which case the subscript i runs over different bins. Once we measure X_i and Y_i from all the mocks the sample variance can be computed as,

$$C_{ij}^{\text{ref}} = \frac{1}{N_{\text{mocks}} - 1} \sum_{k=1}^{N_{\text{mocks}}} (X_i^k - \bar{X}_i)(Y_j^k - \bar{Y}_j), \tag{5.1}$$

where,

$$\bar{X}_i = \frac{1}{N_{\text{mocks}}} \sum_{k=1}^{N_{\text{mocks}}} X_i^k, \quad \bar{Y}_i = \frac{1}{N_{\text{mocks}}} \sum_{k=1}^{N_{\text{mocks}}} Y_i^k \tag{5.2}$$

and the superscript runs over simulations.

This estimate of the covariance matrix is unbiased (as long as the fiducial cosmology of the mocks is reasonably close to the real cosmology of the data) but is noisy. The sample covariance matrix follows a Wishart distribution[241], and the variance in each element of the estimated covariance matrix is

$$\sigma^2(C_{ij}) = \frac{1}{N_{\text{mocks}} - 1}(C_{ij}^2 + C_{ii}C_{jj}) \tag{5.3}$$

which goes to zero as N_{mocks} increases, as expected. The label ref highlights the fact that our Jackknife estimated co-variances will be compared to the results derived with this formula, which effectively serves as a reference point.

5.2.2 Jackknife

In our paper, we utilize the "delete-1" version of the Jackknife method. Specifically, we split the sample into $N_{\rm sub}$ sub-samples and for each Jackknife realization, we omit one of the sub-samples so that it represents a fraction of $(N_{\rm sub}-1)/N_{\rm sub}$ of the original volume. We denote each Jackknife realization with a superscript, such that the i-th measurement from the k-th Jackknife realization is denoted as X_i^k. The Jackknife estimator of the covariance is then calculated as follows:

$$C_{ij}^{\rm jk} = \frac{(N_{\rm sub}-1)}{N_{\rm sub}} \sum_{k=1}^{N_{\rm sub}} (X_i^k - \bar{X}_i)(Y_j^k - \bar{Y}_j), \tag{5.4}$$

where

$$\overline{X}_i = \frac{1}{N_{\rm sub}} \sum_{k=1}^{N_{\rm sub}} X_i^k, \quad \bar{Y}_i = \frac{1}{N_{\rm sub}} \sum_{k=1}^{N_{\rm sub}} Y_i^k \tag{5.5}$$

One potential shortcoming of the Jackknife method above is that, due to the finite size of the data, the resulting estimates can be noisy. To address this issue and remove the fluctuation realization vice,[242;243] proposed a hybrid approach in which the Jackknife resampling is applied to an ensemble of independent simulation realizations. Specifically, if we have $N_{\rm jk}$ mocks, the covariance is estimated from each one of them using eqn. 5.4 and then averaged as follows

$$\bar{C}_{ij}^{\rm jk} = \frac{1}{N_{\rm jk}} \sum_{m=1}^{N_{\rm jk}} C_{ij}^m. \tag{5.6}$$

In contrast to the standard Jackknife method, the hybrid approach requires a small number of simulations. However, the total number of realizations needed is smaller than that required for the sample covariance method to obtain reliable estimates.

clustering	bias	$N_{\rm sub}$
$X = Y = \xi_\ell(s)$	HOD galaxies	125(3D)
$X = Y = \delta\xi_\ell(s)$		
$X = \xi_\ell^{\rm S}(s), Y = \xi_\ell^{\rm F}(s)$	Halos, $\log(M_{\rm cut}) = 11.0$	64(3D)
$X = Y = P_\ell(k)$	Halos, $\log(M_{\rm cut}) = 11.5$	400(2D)
$X = Y = \delta P_\ell(k)$		
$X = P_\ell^{\rm S}(k), Y = P_\ell^{\rm F}(k)$	Halos, $\log(M_{\rm cut}) = 12.0$	100(2D)

Table 5.3.1: All combinations of Jackknife co-variance studied in this work. The left column shows different clustering measurements apply to eqn. 5.1 and 5.4. The superscripts S and F correspond to SLICS and FastPM respectively. The column in the middle shows the catalogs used, we study HOD galaxies and halos with 3 different cut-off masses. The right column shows 4 different Jackknife sub-volume configurations.

5.3 Method

5.3.1 Simulations

We use the Scinet Light Cone Simulations (SLICS) at redshift $z = 1.041$ to create halo and galaxy catalogs. SLICS have a volume of side length $L_{\rm box} = 505\ h^{-1}$Mpc, and 1536^3 dark matter particles with the mass of $2.88 \times 10^9\ h^{-1}M_\odot$ per particle. The SLICS provide enough independent N-body simulations to make a solid external co-variance matrix estimation as a reference compared to the Jackknife method.

To estimate the covariance of difference in a pair of simulations and covariance of a pair of simulations, we use 139 FastPM realizations matching the initial condition of the SLICS. Two simulation share the same volume, resolution, cosmological parameters, and redshift, the only difference is the N-body simulation code.

5.3.2 Clustering statistics

For CF, we measure the redshift-space two-dimensional correlation function $\xi(s, \mu)$, using the Landy and Szalay[62] estimator:

$$\xi(s, \mu) = \frac{DD(s, \mu) - 2DR(s, \mu) + RR(s, \mu)}{RR(s, \mu)} \tag{5.7}$$

Where DD is the number of pairs of data points with certain redshift-space separation s, and cosine of the angle of the pair to the line-of-sight μ. DR is the cross-pair counts between the data and uniform random distribution, and RR is the number of pairs from a uniform random distribution.

We then integrate over μ to obtain the multipoles of the CF,

$$\xi_\ell(s) = \frac{2\ell+1}{2}\int_{-1}^{1}\xi(s,\mu)\mathcal{L}_\ell(\mu)d\mu \tag{5.8}$$

where $L_l(\mu)$ is the Legendre polynomial of order l. We measure correlation function monopole $\xi_0(s)$, quadrupole $\xi_2(s)$ and hexadecapole $\xi_4(s)$ in 20 bins equally spaced in s between 0.1 h^{-1}Mpc and 50 h^{-1}Mpc using CorrFunc package[116;117] setting number of mu bins as 100.

Similarly, for $P(k)$, we measure power spectrum multipoles using the FFT-based method.

$$P_\ell(k) = \frac{2\ell+1}{2}\int_{-1}^{1}P(k,\mu)\mathcal{L}_\ell(\mu)d\mu \tag{5.9}$$

We refer the reader to Bianchi et al.[244] for a detailed description of this method. We measure power spectrum monopole $P_0(k)$, quadrupole $P_2(k)$ and hexadecapole $P_4(k)$ in 30 bins equally spaced in k between 0 Mpc^{-1}h and 0.3 Mpc^{-1}h using NBODYKIT package[245]. We employ a Triangular Shaped Cloud (TSC) method to assign galaxies to the 3D Cartesian grid. We set the number of cells per side on the mesh to 256, this grid configuration implies a Nyquist frequency of $k_{\rm Ny} > 1\,{\rm Mpc}^{-1}h$.

5.3.3 Comparing covariances

We compare the covariances computed using the sample variance (reference method) as described in Section 5.2.1 to co-variances computed using the hybrid Jackknife method as described in Section 5.2.2.

We used a total of 139 independent simulations for the reference method and randomly selected 20 of them for use in the hybrid Jackknife method. These simulations were placed in a cube with periodic boundary conditions. To create Jackknife sub-samples, we removed some portion of this volume using two different methods. In the first method (hereafter referred to as 3D), we divided the cube into a certain number of sub-cubes, using values of 64 and 125 for different settings. The Jackknife sub-samples were then created by removing one of the sub-cubes from the total volume. In the second method (hereafter referred to as 2D), we chose two dimensions of the cube and divided it into a certain number of sub-squares, using values of 100 and 400 in different settings. To create Jackknife sub-samples, we removed all the data that were projected into one of the sub-squares, effectively removing a cubic prism based on that sub-square.

Table 5.3.1 shows all combinations of Jackknife co-variances that we study in our work. The first column lists the clustering measurements. By setting different X and Y in equations 5.1 and 5.4, we look at the covariance of the CF ($\xi(r)$) and the $P(k)$, where we set $X = Y = \xi_\ell^{\mathrm{S}}(s)$ or $X = Y = P_\ell^{\mathrm{S}}(k)$, the covariance of difference of CF and Pk from two simulations, where we set $X = Y = \xi_\ell^{\mathrm{S}}(s) - \xi_\ell^{\mathrm{F}}(s)$ or $X = Y = P_\ell^{\mathrm{S}}(k) - P_\ell^{\mathrm{F}}(k)$, and the covariance of a pair of simulation, where we set $X = \xi_\ell^{\mathrm{S}}(s)$, $Y = \xi_\ell^{\mathrm{F}}(s)$ or $X = P_\ell^{\mathrm{S}}(k)$, $Y = P_\ell^{\mathrm{F}}(k)$. The superscripts S and F represent SLICS and FastPM simulation respectively. The second column lists different choices for the data point. We look at three different lower cut-off masses corresponding to $\log(M_{\mathrm{cut}})$ of 11, 11.5, and 12. We also look at a case where halos are not selected by their mass but instead are populated by galaxies according to the halo occupation distribution (HOD) prescription described in Chuang et al.[194]. The third column lists choices for the Jackknife regions introduced above.

5.4 Results

5.4.1 Correlation Function

Auto correlation function

We start with the covariance for the auto-correlation function using equations. 5.1 and 5.4 with $X = Y = \xi_\ell(s)$.

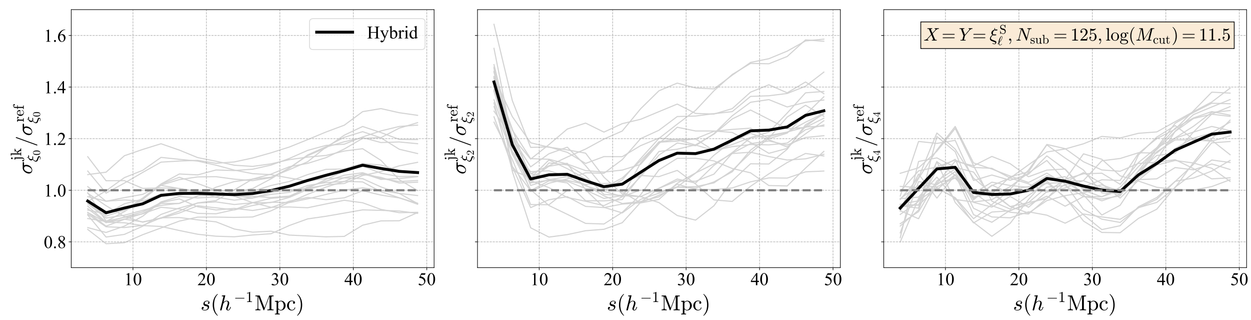


Figure 5.4.1: The ratios of the errors from the Jackknife and the reference method for the auto CF monopole, quadrupole, and hexadecapole. We use the SLICS halo catalog with a cut-off mass $\log M(\mathrm{cut}) = 11.5$. Jackknife estimates slice the data into $N_{\mathrm{sub}} = 125$ sub-samples. The light gray lines show the ratios from the Jackknife using only one simulation realization and the black line shows the averaged error ratio from the hybrid Jackknife.

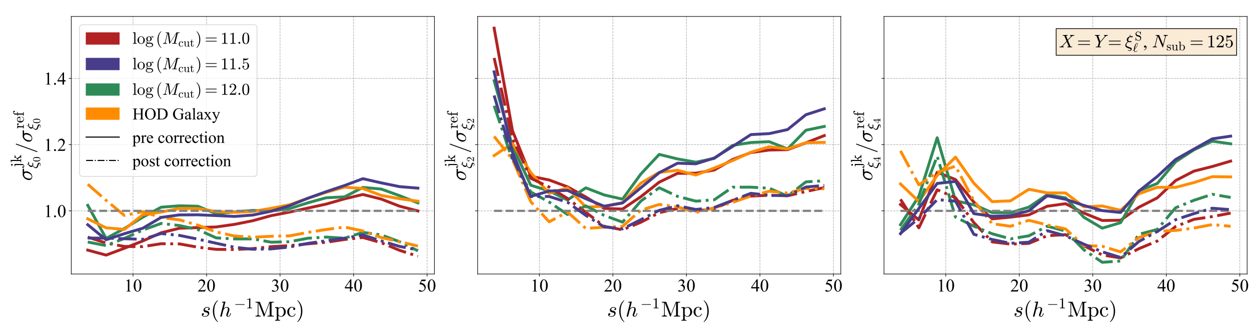


Figure 5.4.2: This plot shows the influence of the cut-off mass of simulation (galaxy bias) on the accuracy of Jackknife estimation. The ratios of error from Jackknife and reference estimation for the auto CF multipole from different SLICS catalogs are shown. We show results before (solid lines) and after Mohammad-Percival correction (dot-dash lines). The Jackknife estimates slice the data into $N_{\mathrm{sub}} = 125$ sub-samples.

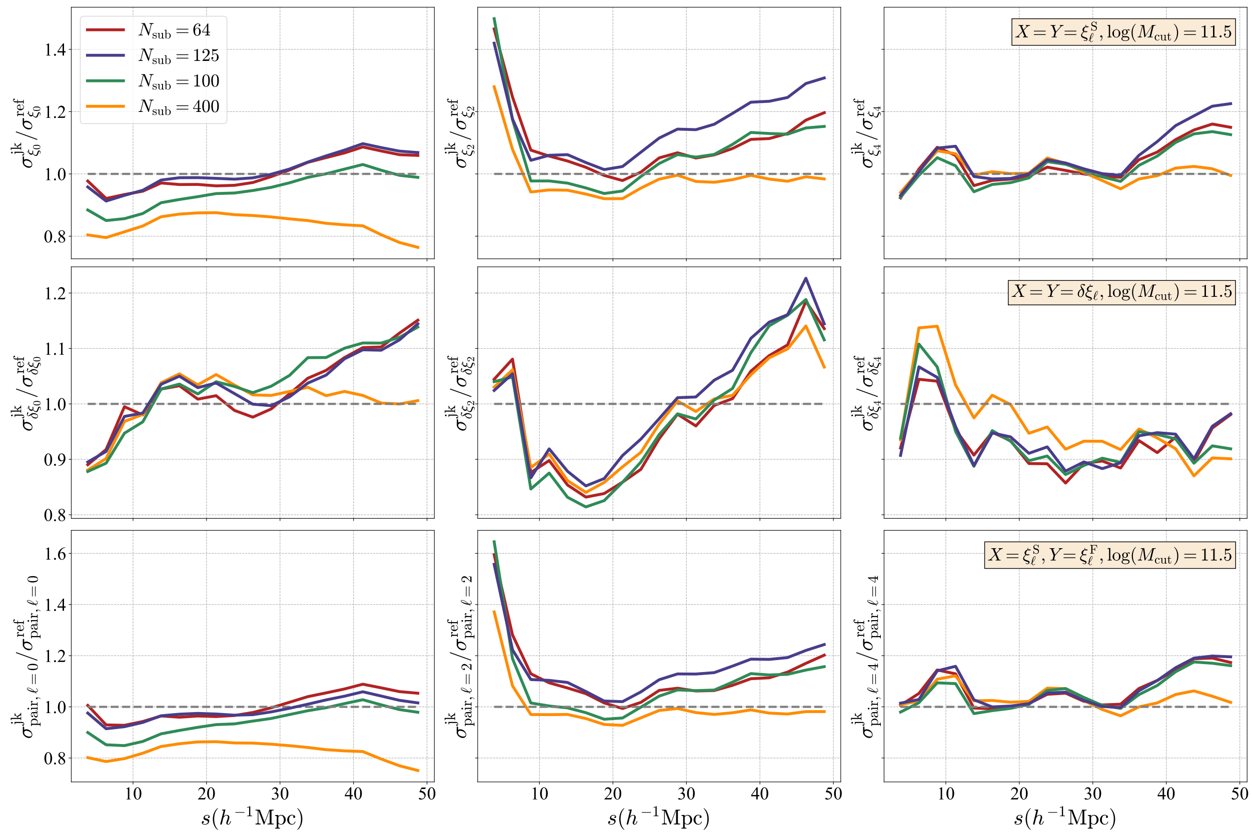


Figure 5.4.3: This plot shows the influence of subsample numbers on the accuracy of Jackknife estimation. The ratios of the errors from the Jackknife and the reference estimation for the errors of auto CF (first row), difference in CFs (second row), and covariance of CFs of a simulation pair (third row) are shown. We use the SLICS halo catalog with $\log(M_{\rm cut}) = 11.5$. Jackknife estimates are average (hybrid) from 20 boxes to remove the fluctuation realization vice.

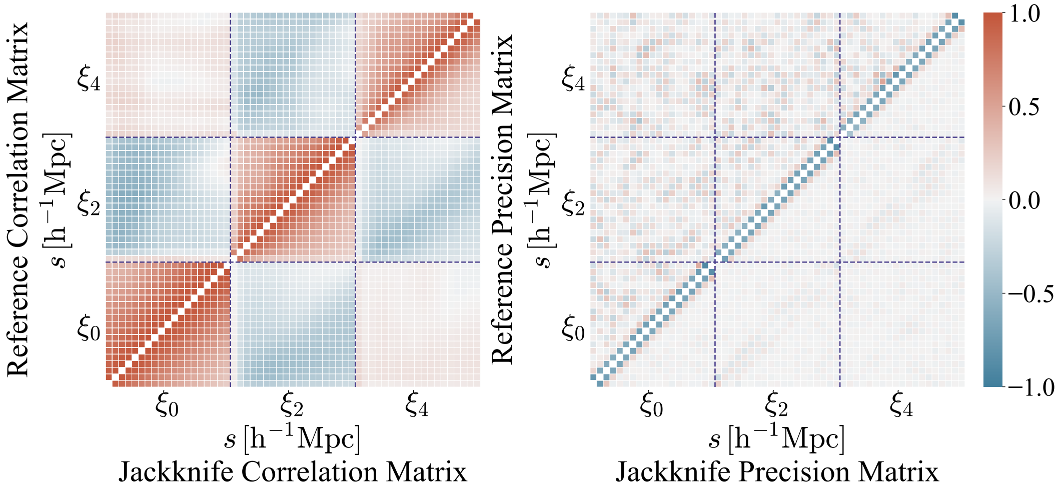


Figure 5.4.4: Comparison of correlation matrices and precision matrices obtained from the reference method (upper left triangle) and the hybrid Jackknife method(lower right triangle). The left panel shows the correlation matrices for the auto CF from the SLICS halo catalog with $\log(M_{\rm cut}) = 11.5$, the panel on the right show the precision matrices The reference matrices are obtained from 139 independent simulation boxes, while the Jackknife matrices are obtained using the hybrid Jackknife method averaged over 20 realizations with $N_{\rm sub} = 125$.

The results from SLICS halo catalogs with $\log(M_{\rm cut}) = 11.5$ and $N_{\rm sub} = 125$ are displayed on Fig. 5.4.1. For the hybrid Jackknife error, we randomly select 20 realizations, while the reference covariance is computed from all 139 SLICS realizations. The ratio of errors between the hybrid Jackknife and reference methods is shown in the black line. Light gray lines show the ratios obtained using only one realization to perform the Jackknife estimation. We find fluctuations of up to 40% among realizations, but the hybrid Jackknife estimates of the errors agree reasonably well with the reference errors. For the monopole, the deviation is less than 10% for all scales, and the Jackknife slightly overestimates the error for scales above $30h^{-1}$Mpc. For quadrupole, the Jackknife estimates significantly overestimate the error for scales below $8h^{-1}$Mpc and start to deviate from the reference measurements at $20h^{-1}$Mpc, overestimating by 30% at $50h^{-1}$Mpc. The hexadecapole shows good agreement until around $34h^{-1}$Mpc, after which the Jackknife overestimates the errors.

We find the Jackknife method begins to overestimate the errors at around $30h^{-1}$Mpc, particularly for the quadrupole, which is consistent with the results reported by Mohammad and Percival[166]. We then further test the Jackknife with the correction weights applied to

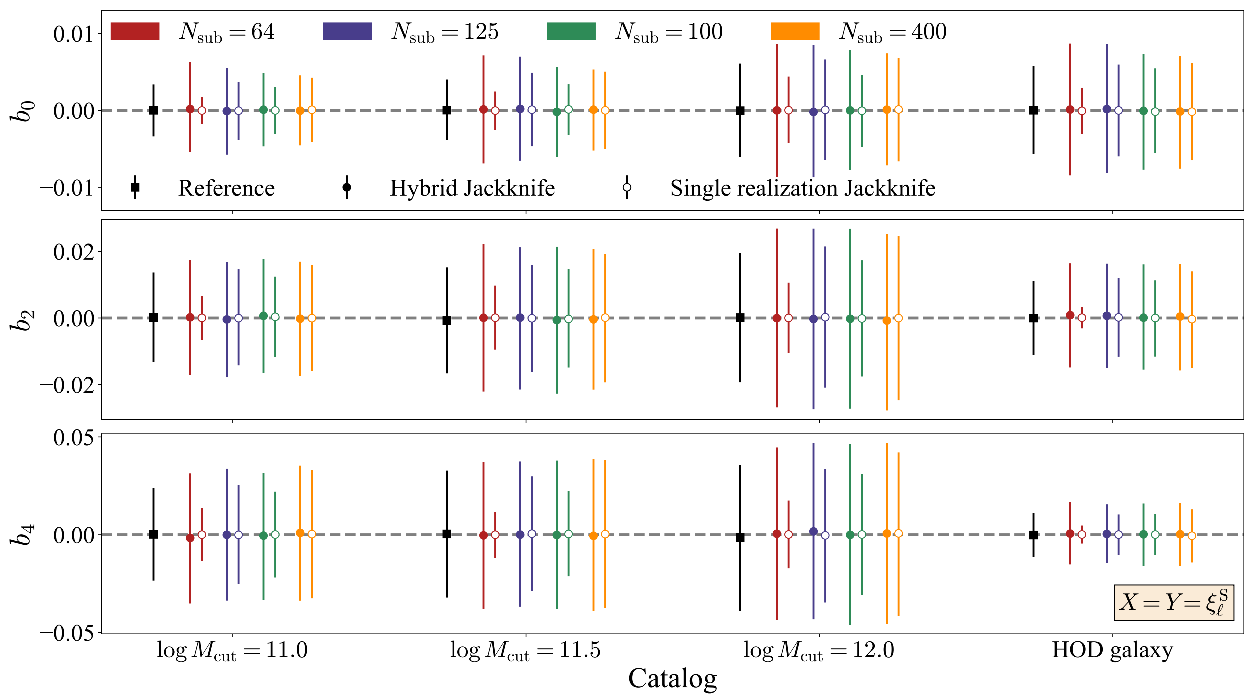


Figure 5.4.5: Constraints on b_ℓ using different covariance matrices. The mean $\pm 1\sigma$ errors are shown. The filled black square markers correspond to the constraint using reference covariance matrices. The circle markers correspond to the constraint using Jackknife covariance matrices, with filled circles representing the hybrid Jackknife covariance averaged from 20 realizations, and open circles representing Jackknife from a specific realization. The different colors correspond to different choices of $N_{\rm sub}$. The x-axes represent the catalog used.

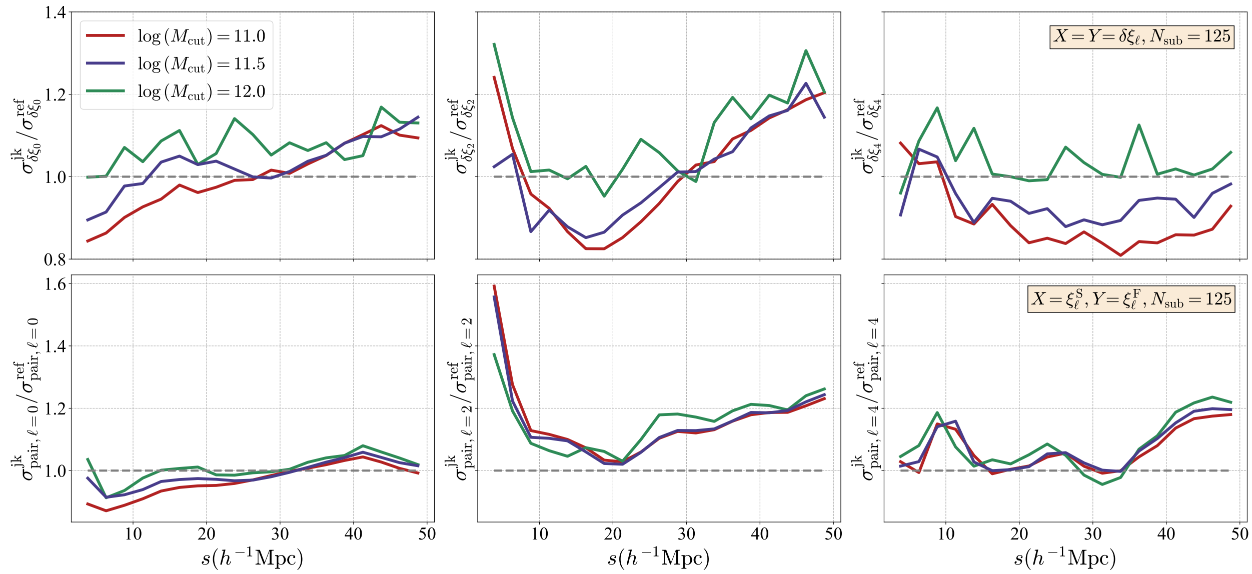


Figure 5.4.6: This plot shows the influence of the cut-off mass of the halo catalog on the accuracy of Jackknife estimation. The ratios of error from the Jackknife and the reference estimation for the difference of CFs (first row) and covariance of CFs for the simulation pair (second row) are shown. The Jackknife estimates slice the data into $N_{\rm sub} = 125$ sub-samples and are average (hybrid) from 20 boxes to remove the fluctuation realization vice.

cross-pairs, as proposed by Mohammad and Percival[166]. Fig. 5.4.2 shows relative differences between the Jackknife and reference estimates of auto CF errors before (solid lines) and after correction (dash-dot lines). The corrected Jackknife method suppresses the ratio for scales above $10h^{-1}$Mpc. While the post-correction Jackknife underestimates the error by around 10% for the monopole, the correction weights for cross-pair help reconcile the offset on the quadrupole and hexadecapole for scales above $20h^{-1}$Mpc. Overall, the Jackknife with correction weights leads to a better recovery of the true covariance.

Rephrasing this paragraph: In the structure formation scenario the correlation structure and noise for the halos with different masses and galaxies could be substantially different. Therefore, we compare the error estimates from Jacknife with mock based error for different halos mass cutt-off and HOD based galaxy catalogues. The results for these cases are shown in Fig. 5.4.2. We also apply the correction suggested in Mohammad and Percival[166] which is shown with dashed line. We find that overall behavious is similar to previous figure and hence independent of halo mass or HOD galaxies. For monopole and hexadecapole pre-correction jackknife error estimates is consistent with mock based error estimates independent of halo mass cutt-off or HOD galaxies. But for the quadrupole we find that raw jackknife overestimates the error above $20h^{-1}$Mpc but correction helps.

The first row of Fig. 5.4.3 shows the dependence of the Jackknife method on the choice of the Jackknife region for a fixed halo mass cut off $\log(M_{\rm cut}) = 11.5$. For monopole, $N_{\rm sub} = 400$(2D) case appears to be the most biased, underestimating the errors by around 20% for all scales. The $N_{\rm sub} = 64$(3D) and $N_{\rm sub} = 125$(3D) choices lead to very similar errors across all scales and are generally preferable to 2D region choices. However, for the quadrupole and the hexadecapole, the $N_{\rm sub} = 400$ choice appears to be the best. The ratios for scales above $20h^{-1}$Mpc do not increase as much as in other cases, and the deviation is smaller than 10% for almost all scales.

Fig. 5.4.4 displays the correlation and precision matrices (inverse of the covariance matrix) for the auto-correlation function of the SLICS halo catalog with $\log(M_{\rm cut}) = 11.5$ from the reference method and the Jackknife method with $N_{\rm sub} = 125$. The correlation patterns between the two methods are similar. We observe more obvious differences from precision matrices, reference precision matrices are noisier compared to the one from the Jackknife method. Based on the number of realizations we used for the reference method (139 realizations), this is expected.

To investigate the difference between the two methods in terms of parameter constraints, we perform the same Markov Chain Monte Carlo (MCMC) analyses as the leading paper[194] using the reference and Jackknife covariance matrices under different configurations. We use the average of 139 SLICS clustering measurements as our observed data, and the theoretical model is constructed as follows

$$X_\ell^{\rm the} = (b_\ell + 1)X_\ell^{\rm obs} \tag{5.10}$$

Where X could be ξ or $P(k)$, $\ell = 0, 2, 4$ corresponding to the monopole, quadrupole, and hexadecapole respectively. By construction, the expectation value of b_0, b_2 and b_4 are 0. The fitting results are presented in Fig. 5.4.5. The x-axis displays the different catalogs we used for fitting, while the y-axis shows the parameters for monopole, quadrupole, and hexadecapole, respectively. The black square markers represent constraints using sample covariance matrices as reference. The colored circle markers represent constraints using Jackknife covariance matrices for comparison, with filled circles indicating fitting using Hybrid Jackknife averaged from 20 realizations, and open circles indicating Jackknife from a single realization. The different colors correspond to different choices of $N_{\rm sub}$.

Although the mean values of b0, b2, and b4 are consistent with zero, as expected, the magnitudes of their errors differ significantly. Hybrid Jackknife covariance matrices always have larger error bars compared to the reference method, and the errors of parameters are overestimated by 20% to 77% depending on the choices of $N_{\rm sub}$ and catalogs. Take the monopole as an example, according to the top left panel of Fig. 5.4.3, the diagonal errors from the hybrid Jackknife and reference method are similar for $N_{\rm sub} = 64$ and $N_{\rm sub} = 125$. For $N_{\rm sub} = 100$, the error is underestimated on small scales by approximately 10%, whereas for $N_{\rm sub} = 400$, the error is underestimated on all scales by around 20%. The deviations from the reference also affect the constraint on b_0, as shown in the first row of Fig. 5.4.5, the hybrid Jackknife method with 2D region choice has smaller error bars compared to 3D, and hybrid Jackknife with $N_{\rm sub} = 400$ have the tightest constraints. However, for all types of catalogs, the hybrid Jackknife with $N_{\rm sub}$=400 has larger error bars compared to the reference method. This is because the hybrid Jackknife covariance matrices have less noisy off-diagonal terms compared to the reference method, as visually evident in the left panel of Fig. 5.4.4. The constraint on b_2 and b_4 show a similar trend.

The constraints obtained from Jackknife covariance matrices of single realizations are tighter compared to the averaged hybrid Jackknife covariance matrices. The difference in parameter errors between the hybrid and single realization Jackknife methods is highly correlated with $N_{\rm sub}$. For example, the single realization Jackknife with $N_{\rm sub} = 64$ could underestimate parameter errors by up to 80% compared to the hybrid Jackknife method, while the difference for $N_{\rm sub} = 400$ is smaller than 10% for most cases. The covariance matrix computed from the Jackknife of a single realization with a small number of sub-volumes is similar to the sample covariance matrix computed from a small number of simulation realizations. Although the diagonal errors are accurate, the off-diagonal terms could be noisy and could impact the parameter constraints. Therefore, we recommend using the hybrid Jackknife method, at least for small $N_{\rm sub}$ configurations, to obtain more reliable parameter constraints.

As expected, the errors of the parameters increase with the cut-off masses of the halo due to the decrease in number density. Furthermore, the errors of b_2 and b_4 from the HOD galaxy catalogs are smaller than those from the halo catalogs. This is also expected based on our fitting range ($s < 50\,\mathrm{Mpc}^{-1}h$). HOD prescription generates satellite galaxies, which provide additional information on small scale and bring extra signal-to-noise to the redshift-space quadrupole and hexadecapole clustering.

Pair Simulations: difference and cross covariance of CF

We will next show the results for variance estimation of the difference of the CFs between SLICS and FastPM simulations and the covariance of CFs of SLICS and FastPM. We use the same formulas from eqs. 5.1 and 5.4 but now with $X = Y = \delta\xi_\ell = \xi_\ell^{\rm S}(s) - \xi_\ell^{\rm F}(s)$ for the difference in CFs and $X = \xi_\ell^{\rm S}(s), Y = \xi_\ell^{\rm F}(s)$ for pair covariance, where superscripts S and F denote SLICS and FastPM simulations respectively. These results are presented on Fig. 5.4.3 and 5.4.6.

Regarding the difference of correlation functions of two simulations, the first row of Fig. 5.4.6 shows that the errors from the Jackknife method overall have good agreement with the reference method, with most deviations being within 20%. However, in detail, the Jackknife method underestimates the errors of the difference for the monopole and quadrupole on small scales for halo catalogs with lower cut-off mass and overestimates errors on scales above $30h^{-1}$Mpc for all halo catalogs. As for the hexadecapole, the halo catalog with $\log(M_{\rm cut}) = 12$ shows the best agreement, while the errors of halo catalogs with lower cut-off mass are underestimated. Moreover, the results do not seem to depend much on the choice of the Jackknife sub-volume number, as shown in the second row of Fig. 5.4.3.

For the covariance of CFs of a simulation pair, the performance is almost the same as for auto CFs. The third row of Fig. 5.4.3 compares the Jackknife method with different $N_{\rm sub}$

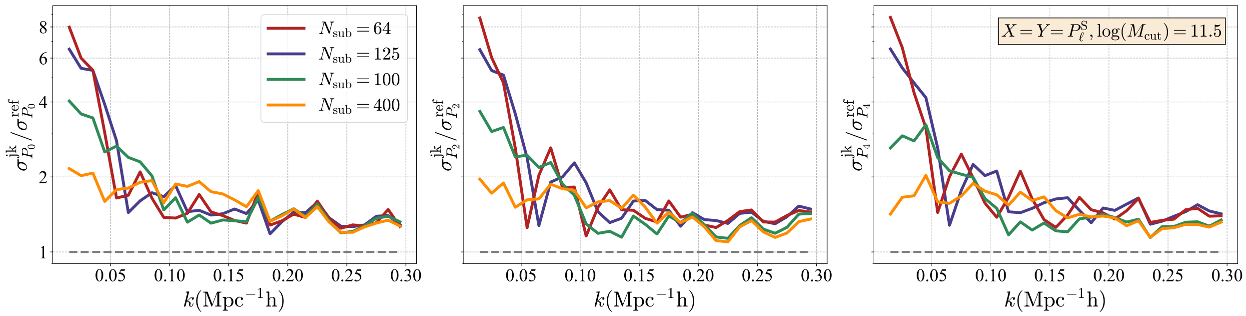


Figure 5.4.7: A similar plot as Fig. 5.4.3 shows the influence of subsample numbers on the accuracy of Jackknife estimation for auto power spectrum.

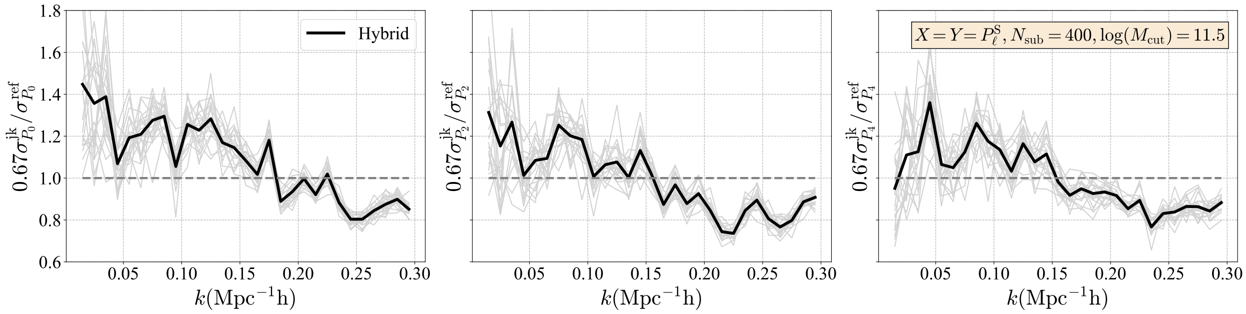


Figure 5.4.8: A similar plot as Fig .5.4.1 shows ratios of errors from the scaled Jackknife and the reference estimations for power spectrum.

configurations, and the results look just like the auto CF case (first row). As shown in the second row of Fig. 5.4.6, the Jackknife estimations are consistent with the reference method ($< 20\%$ deviation for most scales) and do not show a dependence on the cut-off mass of halo catalogs.

In general, the Jackknife method provides a fast and accurate way of estimating the variance of the difference of CF of a simulation pair and covariance of CF of a simulation pair, which are hard to measure from the reference method but essential for fast simulation calibration.

5.4.2 Power Spectrum

Auto power spectrum

We first study the covariance for auto power spectrum using Eqn. 5.1 and 5.4 with $X = Y = P_\ell(k)$. When applying the Jackknife method for the power spectrum, we remove the sub-

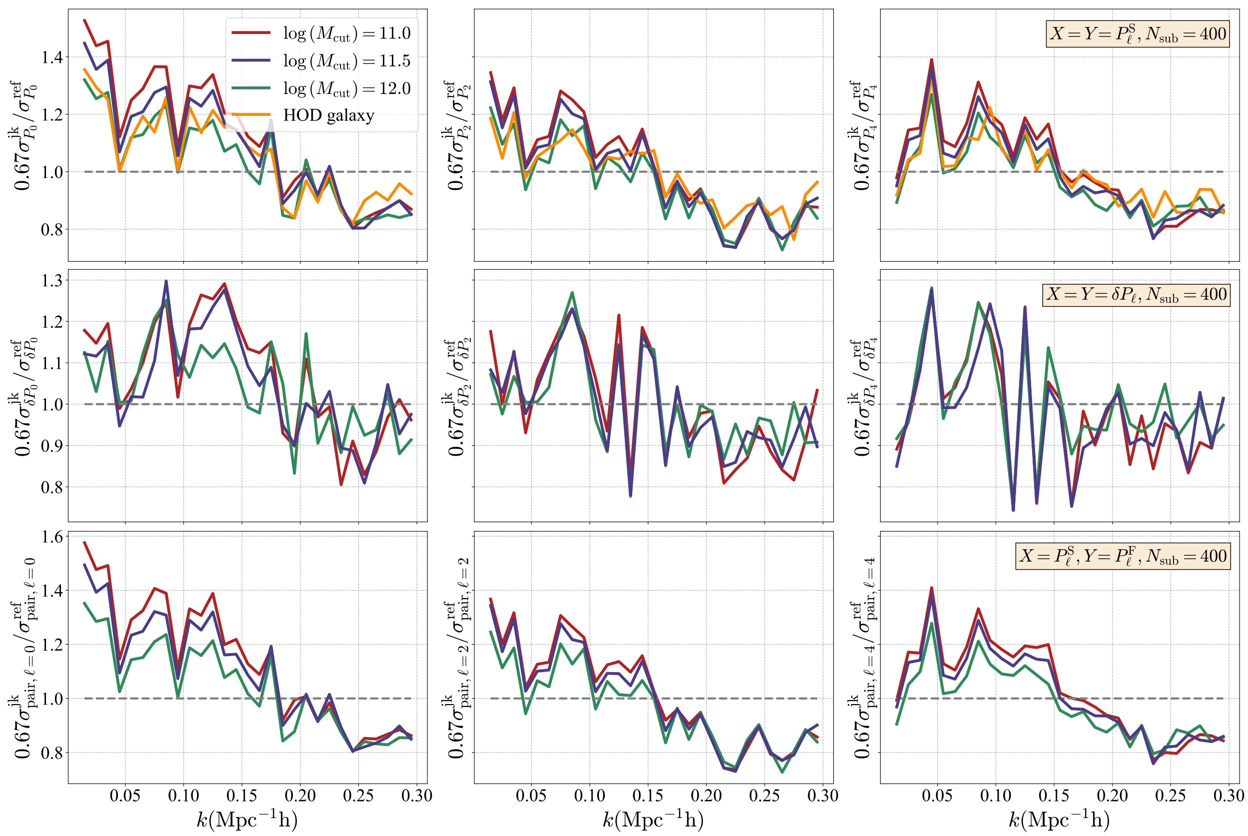


Figure 5.4.9: A similar plot as Fig .5.4.6 shows the influence of the catalogs on the accuracy of the Jackknife estimation for power spectrum case with an extra row on top for auto power spectrum.

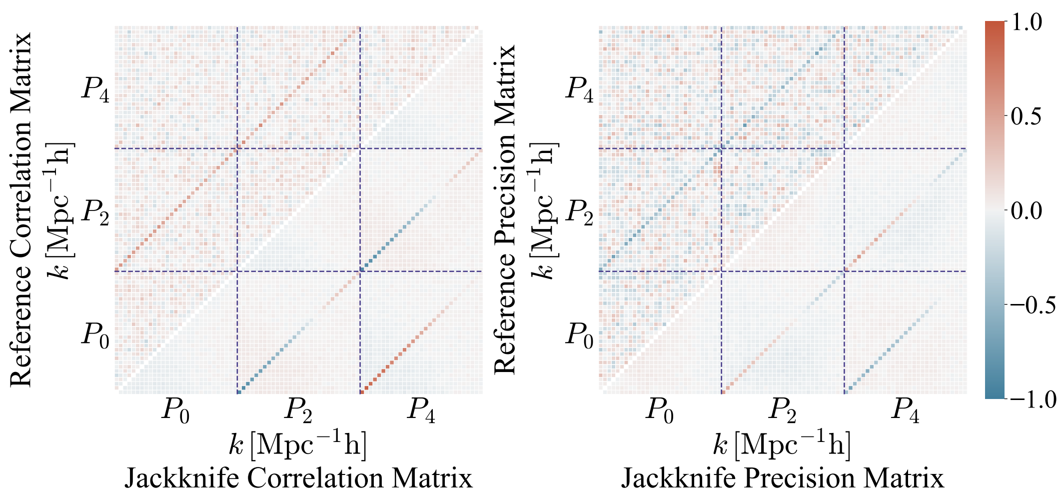


Figure 5.4.10: A similar plot as Fig. 5.4.4 for auto power spectrum multipole. All arrangements are same except we set $N_{\rm sub} = 400$ for $P(k)$ Jackknife matrices.

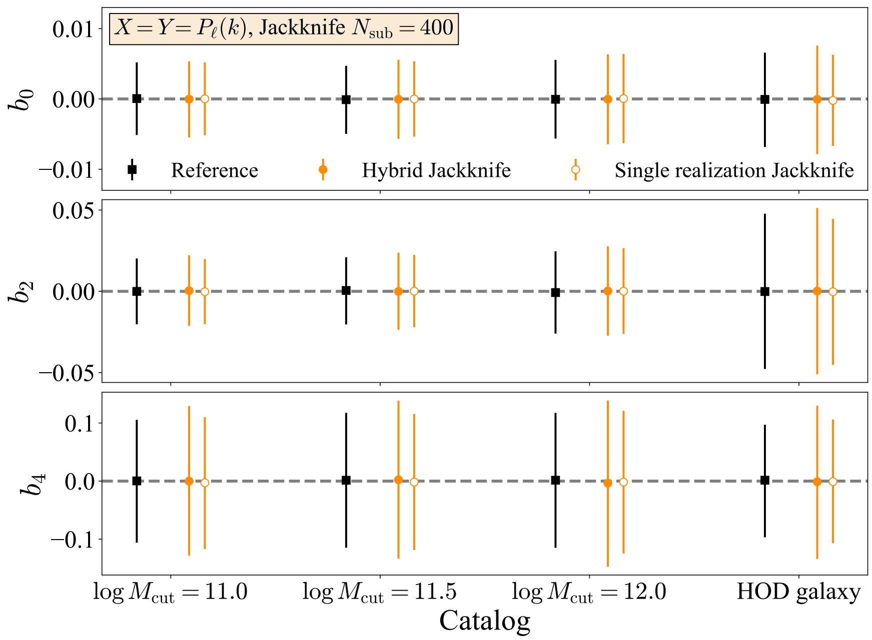


Figure 5.4.11: A similar plot as Fig. 5.4.5 for auto power spectrum

volume in configuration space and perform the Fourier transformation afterward. The way we slice the volume plays an important role since the window effect can significantly influence the jackknife estimation of error, especially for large scales. We first test the influence of $N_{\rm sub}$ for SLICS halo catalog with $\log(M_{\rm cut}) = 11.5$. Fig. 5.4.7 shows the ratio of error from the Jackknife to the reference method for different numbers of sub-volumes. As expected, a larger $N_{\rm sub}$ leads to smaller deviations on large scale with $k < 0.1$, and the behaviors are similar for bins with $k > 0.15$. However, the Jackknife method overestimates the errors for all scales compared to the reference method. In order to match the error numerically, we take the configuration with less window function effect ($N_{\rm sub} = 400$) and apply a scaling factor f as described below,

$$f = \frac{1}{3N_{\rm bin}} \sum_{l=0,2,4} \sum_{i=1}^{N_{\rm bin}} \frac{\sigma_{P_l}^{\rm ref}}{\sigma_{P_l}^{\rm jk}}, \tag{5.11}$$

We calculate the scaling factor f for both SLICS and FastPM catalogs and find $f \approx 0.67$ for both cases. We consider f as a simulation-independent scaling factor that can be applied to the Jackknife covariance to numerically make it comparable to the reference method, at least for the diagonal terms. The scaled jackknife covariance is then $C_{ij}^{\rm jk,scaled} = f^2 C_{ij}^{\rm jk}$. Note that the scaling factor f depends on the $N_{\rm sub}$ and the range of k that we chose, and $f = 0.67$ is specific to the case $N_{\rm sub} = 400$ and $0.01 < k < 0.3$ Mpc^{-1}h.

Figure. 5.4.8 shows the comparison of the reference and jackknife method after applying the scaling factor. We find a clear trend that scaled Jackknife tends to overestimate errors on large scale and underestimate errors on small scale. However, most of the deviations are within 20%. Box to box fluctuation is around 40% on large scale (k¡0.05), and less than 10% on a smaller scale (k¿0.2). The shape of ratios looks similar for monopole, quadrupole, and hexadecapole.

We further compare the influence of catalog, the first row of Fig. 5.4.9 shows scaled jackknife errors have a better agreement for HOD galaxy catalog and halo catalog with larger cut-off mass, we find scaling factor $f = 0.67$ is independent of catalogs and works well for all cases.

Fig. 5.4.10 shows the correlation matrix and precision matrix comparison using the halo catalog with $\log(M_{\rm cut}) = 11.5$. The matrices from the reference method are noisier and have different sub-diagonal patterns (P_0, P_2 and P_0, P_4 correlation) compared to matrices from the Jackknife method. We apply the same analysis as in Sec. 5.4.1 to the covariance matrices for the $P(k)$, by fitting to the average of 139 SLICS power spectrum multipole using a simplified theoretical model suggested by Eqn. 5.10. The marginalized posterior of the fittings are shown in Fig. 5.4.11, with the same arrangement as Fig. 5.4.5, but displaying only the results from the reference covariance matrices and the Jackknife covariance matrices with $N_{\rm sub} = 400$.

Although we observe different patterns in the correlation and precision matrices, the parameter constraints from the reference method and the hybrid Jackknife method appear to be consistent with each other. The mean values of b_0, b_2 and b_4 are around zero as expected, and the magnitudes of errors are also similar. The hybrid Jackknife covariance matrices have slightly larger parameter errors, up to 10% depending on the catalog we fit to. This could be explained by the fact that the hybrid Jackknife method shows less noisy off-diagonal terms (see Fig. 5.4.10). Also, the parameter errors from the Jackknife covariance matrices of a single realization have less than a 10% difference compared to the hybrid Jackknife method, given that we are using a relatively large $N_{\mathrm{sub}} = 400$.

Difference in $P(k)$ and covariance of $P(k)$ of simulation pair

One of our motivations for estimating the covariance of the power spectrum using the Jackknife method is to estimate the errors of the difference in a pair of simulations ($X = Y = \delta P_\ell(k) = P_\ell^{\mathrm{S}}(k) - P_\ell^{\mathrm{F}}(k)$) and covariance of a pair of simulations ($X = P_\ell^{\mathrm{S}}(k), Y = P_\ell^{\mathrm{F}}(k)$).

For the difference of $P(k)$, the scaling factor works even better compare to auto $P(k)$, as shown in the second row of Fig. 5.4.9. The scale (k) dependence is not as clear as the auto $P(k)$ cases, and most of the deviations are within 20 %. The Jackknife estimation performs equally well for all halo catalogs, and it does not depend on the cut-off mass. Regarding the covariance of a simulation pair, the shapes of the ratios look identical to auto $P(k)$ as shown in the third row of Fig. 5.4.9. The Jackknife method overestimates the errors on large scales and underestimates the errors on small scales, and it works better on halo catalogs with a larger cut-off mass.

5.5 Conclusions

We systematically study the performance of the Jackknife re-sampling technique in estimating the covariance of auto, the difference in pair of simulations, and the covariance of a pair of simulations for redshift-space CF and $P(k)$ multipoles in the context of DESI. We use 139 SLICS and FastPM simulations with the same initial conditions to prepare halo catalogs with different cut-off masses and HOD galaxy catalogs to estimate reference and jackknife covariance. We test the performance of the Jackknife method for four different catalogs: the halo catalog with cut-off mass $M_{\mathrm{cut}} = 10^{11.0}, 10^{11.5}, 10^{12.0} h^{-1} M_\odot$ and galaxy catalog generated using HOD prescription. Additionally, we test the influence of different jackknife sub-volumes by slicing volume to $N_{\mathrm{sub}} = 64, 125$ (3D) or $N_{\mathrm{sub}} = 100, 400$ (2D) sub-samples.

We find the following result for correlation functions:

(a) The Jackknife method gives consistent error estimates as a mock-based method within 20%.

(b) The number of jackknife region do not affect strongly the error estimate unless we start making very small jackknife region such as for $N_{\rm sub} = 400$, which corresponds to $\approx 25h^{-1}$Mpc on two sides.

(c) Error on the difference of correlation function for pair simulations is estimated within 20% which is a fairly good estimate and extremely efficient. This will be the most valuable use case for jackknife covariance estimates.

(d) The cross-covariance between a pair of simulations is also estimated very well except when very small jackknife regions are used (e.g. $N_{\rm sub} = 400$).

(e) We also tested a recently proposed correction scheme on jackknife and found consistent results as[166].

We study the error estimates on the power spectrum extensively using the jackknife method for the first time. We find the following result for the case of the power spectrum:

(a) We find that the jackknife error for the power spectrum requires a constant correction factor when compared to the mock-based error. This correction factor is independent of details of halos or galaxies. But depending on the window function and hence the geometry of the jackknife region.

(b) The Jackknife error's scale dependence is slightly different compared to mock based error estimate. This leads to overestimation of error on large scales and underestimation at small scales.

(c) Error on the difference of power spectrum for pair simulations is estimated within 20% which is a fairly good estimate and extremely efficient.

(d) The cross-covariance between a pair of simulations is consistent with the mock-based estimate with a similar deviation as described for the auto-power spectrum.

We also investigate the performance of the Jackknife covariance in terms of cosmological constraints. For auto CF, the hybrid Jackknife covariance matrices show larger parameter errors compared to the reference covariance matrices. However, the estimations of parameter errors from the hybrid Jackknife could be more accurate since the reference covariance matrices use only 139 realizations and have noisy off-diagonal terms. We also show that the Jackknife covariance matrices from single realizations could significantly bias the parameter error estimation. For auto $P(k)$, the parameter errors from the hybrid Jackknife method show good agreement with the reference method.

This paper is part of a series produced by the DESI Covariance Matrix Mock Challenge team, which focuses on the verification of Jackknife covariance performance. Chuang et al.[194] provides an extensive overview of the DESI mock challenge project, validating the data analysis pipeline by testing covariance using various methods. Other accompanying papers, including Variu et al.[246]; Balaguera-Antolínez et al.[247]; EZ et al.[248]; Chen et al.[249] delve into the specifics of covariance matrices derived from the FastPM, BAM, EZ mock, and analytical prediction, respectively.

Acknowledgements

We would like to thank Joachim for help with SLICS simulations.

SA acknowledge support of the Department of Atomic Energy, Government of india, under project no. 12-R&D-TFR-5.02-0200. SA was partially supported by the European Research Council through the COSFORM Research Grant (#670193) and STFC consolidated grant no. RA5496.

This material is based upon work supported by the U.S. Department of Energy (DOE), Office of Science, Office of High-Energy Physics, under Contract No. DE–AC02–05CH11231, and by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract. Additional support for DESI was provided by the U.S. National Science Foundation (NSF), Division of Astronomical Sciences under Contract No. AST-0950945 to the NSF's National Optical-Infrared Astronomy Research Laboratory;

the Science and Technologies Facilities Council of the United Kingdom; the Gordon and Betty Moore Foundation; the Heising-Simons Foundation; the French Alternative Energies and Atomic Energy Commission (CEA); the National Council of Science and Technology of Mexico (CONACYT); the Ministry of Science and Innovation of Spain (MICINN), and by the DESI Member Institutions: https://www.desi.lbl.gov/collaborating-institutions. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U. S. National Science Foundation, the U. S. Department of Energy, or any of the listed funding agencies.

The authors are honored to be permitted to conduct scientific research on Iolkam Du'ag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation.

We acknowledge the use of the NASA astrophysics data system https://ui.adsabs.harvard.edu and the arXiv open-access repository https://arxiv.org. The software was hosted on the GitHub platform https://github.com. The manuscript was typeset using the overleaf cloud-based LaTeXeditor https://www.overleaf.com.

Data Availability

The clustering statistics used in this study, including correlation functions, power spectra, and covariance matrices are available at https://doi.org/10.5281/zenodo.8021712.

Chapter 6

Conclusion

Throughout this thesis, we have embarked on a journey to navigate the complex landscape of the galaxy-halo connection and cosmology, utilizing the wealth of data provided by the Dark Energy Spectroscopic Instrument (DESI) survey. The findings underscore the intertwined nature of these fields and their relevance to our understanding of the universe.

We have demonstrated the power of higher-order statistics in unraveling the complex relationship between galaxies and their dark matter halos. By employing two-point and three-point projected correlation functions, we have examined their capability to constrain HOD parameters for key DESI tracers using mock catalogs. This study has highlighted the additional constraining power of the three-point correlation function, owing to its sensitivity to non-linear scale clustering. Moreover, the implementation of a fast HOD fitting pipeline represents a significant advancement in the efficiency of such analyses.

Data from the DESI One-Percent Survey was meticulously analyzed, demonstrating that the baseline HOD model effectively captures small-scale clustering. This analysis also brought to light clear redshift evolution trends in the HOD and physical parameters for the LRG sample, paving the way for further investigation of LRGs using halo light-cone catalogs in tandem with a redshift-evolved HOD model.

We present HOD's utility in extracting cosmological information from small-scale observations. The exploration of the degeneracy between HOD and cosmological parameters on small-scale two-point correlation function multipoles provided vital insights into the source of small-scale constraining power.

An in-depth exploration of the jackknife resampling method evaluated its performance in estimating covariance matrices for various clustering measurements. The study highlighted the method's robustness, which depends on several factors such as the number of jackknife regions and the dimensionality of these regions.

In conclusion, this thesis has illuminated the critical role of understanding the galaxy-halo connection in interpreting cosmological data from large galaxy surveys like DESI. This work lays a solid foundation for future explorations in this field and invites further investigations to refine these models and methods.

Bibliography

[1] Risa H. Wechsler and Jeremy L. Tinker. The Connection Between Galaxies and Their Dark Matter Halos. ARA&A, 56:435–487, September 2018. doi: 10.1146/annurev-astro-081817-051756.

[2] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J. M. Delouis, E. Di Valentino, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, M. Farhang, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, P. Lemos, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, D. Molinari, L. Montier, G. Morgante,

A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A. S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Zacchei, and A. Zonca. Planck 2018 results. VI. Cosmological parameters. A&A, 641:A6, September 2020. doi: 10.1051/0004-6361/201833910.

[3] DESI Collaboration, B. Abareshi, J. Aguilar, S. Ahlen, Shadab Alam, David M. Alexander, R. Alfarsy, L. Allen, C. Allende Prieto, O. Alves, J. Ameel, E. Armengaud, J. Asorey, Alejandro Aviles, S. Bailey, A. Balaguera-Antolínez, O. Ballester, C. Baltay, A. Bault, S. F. Beltran, B. Benavides, S. BenZvi, A. Berti, R. Besuner, Florian Beutler, D. Bianchi, C. Blake, P. Blanc, R. Blum, A. Bolton, S. Bose, D. Bramall, S. Brieden, A. Brodzeller, D. Brooks, C. Brownewell, E. Buckley-Geer, R. N. Cahn, Z. Cai, R. Canning, R. Capasso, A. Carnero Rosell, P. Carton, R. Casas, F. J. Castander, J. L. Cervantes-Cota, S. Chabanier, E. Chaussidon, C. Chuang, C. Circosta, S. Cole, A. P. Cooper, L. da Costa, M. C. Cousinou, A. Cuceu, T. M. Davis, K. Dawson, R. de la Cruz-Noriega, A. de la Macorra, A. de Mattia, J. Della Costa, P. Demmer, M. Derwent, A. Dey, B. Dey, G. Dhungana, Z. Ding, C. Dobson, P. Doel, J. Donald-McCann, J. Donaldson, K. Douglass, Y. Duan, P. Dunlop, J. Edelstein, S. Eftekharzadeh, D. J. Eisenstein, M. Enriquez-Vargas, S. Escoffier, M. Evatt, P. Fagrelius, X. Fan, K. Fanning, V. A. Fawcett, S. Ferraro, J. Ereza, B. Flaugher, A. Font-Ribera, J. E. Forero-Romero, C. S. Frenk, S. Fromenteau, B. T. Gänsicke, C. Garcia-Quintero, L. Garrison, E. Gaztañaga, F. Gerardi, H. Gil-Marín, S. Gontcho a Gontcho, Alma X. Gonzalez-Morales, G. Gonzalez-de-Rivera, V. Gonzalez-Perez, C. Gordon, O. Graur, D. Green, C. Grove, D. Gruen, G. Gutierrez, J. Guy, C. Hahn, S. Harris, D. Herrera, Hiram K.

Herrera-Alcantar, K. Honscheid, C. Howlett, D. Huterer, V. Iršič, M. Ishak, P. Jelinsky, L. Jiang, J. Jimenez, Y. P. Jing, R. Joyce, E. Jullo, S. Juneau, N. G. Karaçaylı, M. Karamanis, A. Karcher, T. Karim, R. Kehoe, S. Kent, D. Kirkby, T. Kisner, F. Kitaura, S. E. Koposov, A. Kovács, A. Kremin, Alex Krolewski, B. L'Huillier, O. Lahav, A. Lambert, C. Lamman, Ting-Wen Lan, M. Landriau, S. Lane, D. Lang, J. U. Lange, J. Lasker, L. Le Guillou, A. Leauthaud, A. Le Van Suu, Michael E. Levi, T. S. Li, C. Magneville, M. Manera, Christopher J. Manser, B. Marshall, Paul Martini, W. McCollam, P. McDonald, Aaron M. Meisner, J. Mena-Fernández, J. Meneses-Rizo, M. Mezcua, T. Miller, R. Miquel, P. Montero-Camacho, J. Moon, J. Moustakas, E. Mueller, Andrea Muñoz-Gutiérrez, Adam D. Myers, S. Nadathur, J. Najita, L. Napolitano, E. Neilsen, Jeffrey A. Newman, J. D. Nie, Y. Ning, G. Niz, P. Norberg, Hernán E. Noriega, T. O'Brien, A. Obuljen, N. Palanque-Delabrouille, A. Palmese, P. Zhiwei, D. Pappalardo, X. PENG, W. J. Percival, S. Perruchot, R. Pogge, C. Poppett, A. Porredon, F. Prada, J. Prochaska, R. Pucha, A. Pérez-Fernández, I. Pérez-Ràfols, D. Rabinowitz, A. Raichoor, S. Ramirez-Solano, César Ramírez-Pérez, C. Ravoux, K. Reil, M. Rezaie, A. Rocher, C. Rockosi, N. A. Roe, A. Roodman, A. J. Ross, G. Rossi, R. Ruggeri, V. Ruhlmann-Kleider, C. G. Sabiu, S. Safonova, K. Said, A. Saintonge, Javier Salas Catonga, L. Samushia, E. Sanchez, C. Saulder, E. Schaan, E. Schlafly, D. Schlegel, J. Schmoll, D. Scholte, M. Schubnell, A. Secroun, H. Seo, S. Serrano, Ray M. Sharples, Michael J. Sholl, Joseph Harry Silber, D. R. Silva, M. Sirk, M. Siudek, A. Smith, D. Sprayberry, R. Staten, B. Stupak, T. Tan, Gregory Tarlé, Suk Sien Tie, R. Tojeiro, L. A. Ureña-López, F. Valdes, O. Valenzuela, M. Valluri, M. Vargas-Magaña, L. Verde, M. Walther, B. Wang, M. S. Wang, B. A. Weaver, C. Weaverdyck, R. Wechsler, Michael J. Wilson, J. Yang, Y. Yu, S. Yuan, Christophe Yèche, H. Zhang, K. Zhang, Cheng Zhao, Rongpu Zhou, Zhimin Zhou, H. Zou, J. Zou, S. Zou, Y. Zu, and DESI Collaboration. Overview of the Instrumentation for the Dark Energy Spectroscopic Instrument. AJ, 164(5):207, November 2022. doi: 10.3847/1538-3881/ac882b.

[4] Donald G. York, J. Adelman, Jr. Anderson, John E., Scott F. Anderson, James Annis, Neta A. Bahcall, J. A. Bakken, Robert Barkhouser, Steven Bastian, Eileen Berman, William N. Boroski, Steve Bracker, Charlie Briegel, John W. Briggs, J. Brinkmann, Robert Brunner, Scott Burles, Larry Carey, Michael A. Carr, Francisco J. Castander, Bing Chen, Patrick L. Colestock, A. J. Connolly, J. H. Crocker, István Csabai, Paul C. Czarapata, John Eric Davis, Mamoru Doi, Tom Dombeck, Daniel Eisenstein, Nancy Ellman, Brian R. Elms, Michael L. Evans, Xiaohui Fan, Glenn R. Federwitz, Larry Fiscelli, Scott Friedman, Joshua A. Frieman, Masataka Fukugita, Bruce Gillespie, James E. Gunn, Vijay K. Gurbani, Ernst de Haas, Merle Haldeman, Frederick H. Harris, J. Hayes, Timothy M. Heckman, G. S. Hennessy, Robert B. Hindsley, Scott Holm, Donald J. Holmgren, Chi-hao Huang, Charles Hull, Don Husby, Shin-Ichi Ichikawa, Takashi Ichikawa, Željko Ivezić, Stephen Kent, Rita S. J. Kim, E. Kinney, Mark Klaene, A. N. Kleinman, S. Kleinman, G. R. Knapp, John Korienek, Richard G. Kron, Peter Z. Kunszt, D. Q. Lamb, B. Lee, R. French Leger, Siriluk Limmongkol, Carl Lindenmeyer, Daniel C. Long, Craig Loomis, Jon Loveday, Rich Lucinio, Robert H. Lupton, Bryan MacKinnon, Edward J. Mannery, P. M. Mantsch, Bruce Margon, Peregrine McGehee, Timothy A. McKay, Avery Meiksin, Aronne Merelli, David G. Monet, Jeffrey A. Munn, Vijay K. Narayanan, Thomas Nash, Eric Neilsen, Rich Neswold, Heidi Jo Newberg, R. C. Nichol, Tom Nicinski, Mario Nonino, Norio Okada, Sadanori Okamura, Jeremiah P. Ostriker, Russell Owen, A. George Pauls, John Peoples, R. L. Peterson, Donald Petravick, Jeffrey R. Pier, Adrian Pope, Ruth Pordes, Angela Prosapio, Ron Rechenmacher, Thomas R. Quinn, Gordon T. Richards, Michael W. Richmond, Claudio H. Rivetta, Constance M. Rockosi, Kurt Ruthmansdorfer, Dale Sandford, David J. Schlegel, Donald P. Schneider, Maki Sekiguchi, Gary Sergey, Kazuhiro Shimasaku, Walter A. Siegmund, Stephen Smee, J. Allyn Smith, S. Snedden, R. Stone, Chris Stoughton, Michael A. Strauss, Christopher Stubbs, Mark SubbaRao, Alexander S. Szalay, Istvan Szapudi, Gyula P. Szokoly, Anirudda R. Thakar, Christy Tremonti, Douglas L. Tucker, Alan Uomoto, Dan Vanden Berk, Michael S. Vogeley, Patrick Waddell, Shu-i. Wang, Masaru Watanabe, David H. Weinberg, Brian Yanny, Naoki

Yasuda, and SDSS Collaboration. The Sloan Digital Sky Survey: Technical Summary. AJ, 120(3):1579–1587, September 2000. doi: 10.1086/301513.

[5] Martin White, Lars Hernquist, and Volker Springel. The Halo Model and Numerical Simulations. ApJ, 550(2):L129–L132, April 2001. doi: 10.1086/319644.

[6] F. R. Pearce, A. Jenkins, C. S. Frenk, S. D. M. White, P. A. Thomas, H. M. P. Couchman, J. A. Peacock, and G. Efstathiou. Simulations of galaxy formation in a cosmological volume. MNRAS, 326(2):649–666, September 2001. doi: 10.1046/j.1365-8711.2001.04616.x.

[7] Andreas A. Berlind, David H. Weinberg, Andrew J. Benson, Carlton M. Baugh, Shaun Cole, Romeel Davé, Carlos S. Frenk, Adrian Jenkins, Neal Katz, and Cedric G. Lacey. The Halo Occupation Distribution and the Physics of Galaxy Formation. ApJ, 593(1): 1–25, August 2003. doi: 10.1086/376517.

[8] Sownak Bose, Daniel J. Eisenstein, Lars Hernquist, Annalisa Pillepich, Dylan Nelson, Federico Marinacci, Volker Springel, and Mark Vogelsberger. Revealing the galaxy-halo connection in IllustrisTNG. MNRAS, 490(4):5693–5711, December 2019. doi: 10.1093/mnras/stz2546.

[9] Boryana Hadzhiyska, Sownak Bose, Daniel Eisenstein, Lars Hernquist, and David N. Spergel. Limitations to the 'basic' HOD model and beyond. MNRAS, 493(4):5506–5519, April 2020. doi: 10.1093/mnras/staa623.

[10] Boryana Hadzhiyska, Sownak Bose, Daniel Eisenstein, and Lars Hernquist. Extensions to models of the galaxy-halo connection. MNRAS, 501(2):1603–1620, February 2021. doi: 10.1093/mnras/staa3776.

[11] Boryana Hadzhiyska, Sandro Tacchella, Sownak Bose, and Daniel J. Eisenstein. The galaxy-halo connection of emission-line galaxies in IllustrisTNG. MNRAS, 502(3): 3599–3617, April 2021. doi: 10.1093/mnras/stab243.

[12] A. J. Benson, F. R. Pearce, C. S. Frenk, C. M. Baugh, and A. Jenkins. A comparison of semi-analytic and smoothed particle hydrodynamics galaxy formation. MNRAS, 320(2):261–280, January 2001. doi: 10.1046/j.1365-8711.2001.03966.x.

[13] Vimal Simha, David H. Weinberg, Romeel Davé, Mark Fardal, Neal Katz, and Benjamin D. Oppenheimer. Testing subhalo abundance matching in cosmological smoothed particle hydrodynamics simulations. MNRAS, 423(4):3458–3473, July 2012. doi: 10.1111/j.1365-2966.2012.21142.x.

[14] Jonás Chaves-Montero, Raul E. Angulo, Joop Schaye, Matthieu Schaller, Robert A. Crain, Michelle Furlong, and Tom Theuns. Subhalo abundance matching and assembly bias in the EAGLE simulation. MNRAS, 460(3):3100–3118, August 2016. doi: 10.1093/mnras/stw1225.

[15] Simon D. M. White and Carlos S. Frenk. Galaxy Formation through Hierarchical Clustering. ApJ, 379:52, September 1991. doi: 10.1086/170483.

[16] G. Kauffmann, S. D. M. White, and B. Guiderdoni. The formation and evolution of galaxies within merging dark matter haloes. MNRAS, 264:201–218, September 1993. doi: 10.1093/mnras/264.1.201.

[17] Rachel S. Somerville and Joel R. Primack. Semi-analytic modelling of galaxy formation: the local Universe. MNRAS, 310(4):1087–1110, December 1999. doi: 10.1046/j.1365-8711.1999.03032.x.

[18] Shaun Cole, Cedric G. Lacey, Carlton M. Baugh, and Carlos S. Frenk. Hierarchical galaxy formation. MNRAS, 319(1):168–204, November 2000. doi: 10.1046/j.1365-8711.2000.03879.x.

[19] R. G. Bower, A. J. Benson, R. Malbon, J. C. Helly, C. S. Frenk, C. M. Baugh, S. Cole, and C. G. Lacey. Breaking the hierarchy of galaxy formation. MNRAS, 370(2):645–655, August 2006. doi: 10.1111/j.1365-2966.2006.10519.x.

[20] Quan Guo, Shaun Cole, Vincent Eke, Carlos Frenk, and John Helly. Spatial and luminosity distributions of galactic satellites. MNRAS, 434(3):1838–1848, September 2013. doi: 10.1093/mnras/stt903.

[21] William H. Press and Paul Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. ApJ, 187:425–438, February 1974. doi: 10.1086/152650.

[22] Andrey V. Kravtsov, Andreas A. Berlind, Risa H. Wechsler, Anatoly A. Klypin, Stefan Gottlöber, Brandon Allgood, and Joel R. Primack. The Dark Side of the Halo Occupation Distribution. ApJ, 609(1):35–49, July 2004. doi: 10.1086/420959.

[23] Argyro Tasitsiomi, Andrey V. Kravtsov, Risa H. Wechsler, and Joel R. Primack. Modeling Galaxy-Mass Correlations in Dissipationless Simulations. ApJ, 614(2):533–546, October 2004. doi: 10.1086/423784.

[24] Charlie Conroy, Risa H. Wechsler, and Andrey V. Kravtsov. Modeling Luminosity-dependent Galaxy Clustering through Cosmic Time. ApJ, 647(1):201–214, August 2006. doi: 10.1086/503602.

[25] Peter S. Behroozi, Charlie Conroy, and Risa H. Wechsler. A Comprehensive Analysis of Uncertainties Affecting the Stellar Mass-Halo Mass Relation for 0 ¡ z ¡ 4. ApJ, 717 (1):379–403, July 2010. doi: 10.1088/0004-637X/717/1/379.

[26] Y. P. Jing, H. J. Mo, and G. Börner. Spatial Correlation Function and Pairwise Velocity Dispersion of Galaxies: Cold Dark Matter Models versus the Las Campanas Survey. ApJ, 494(1):1–12, February 1998. doi: 10.1086/305209.

[27] Uroš Seljak. Analytic model for galaxy and dark matter clustering. MNRAS, 318(1): 203–213, October 2000. doi: 10.1046/j.1365-8711.2000.03715.x.

[28] J. A. Peacock and R. E. Smith. Halo occupation numbers and galaxy bias. MNRAS, 318(4):1144–1156, November 2000. doi: 10.1046/j.1365-8711.2000.03779.x.

[29] A. J. Benson, C. S. Frenk, C. M. Baugh, S. Cole, and C. G. Lacey. The clustering evolution of the galaxy distribution. MNRAS, 327(4):1041–1056, November 2001. doi: 10.1046/j.1365-8711.2001.04824.x.

[30] Román Scoccimarro, Ravi K. Sheth, Lam Hui, and Bhuvnesh Jain. How Many Galaxies Fit in a Halo? Constraints on Galaxy Formation Efficiency from Spatial Clustering. ApJ, 546(1):20–34, January 2001. doi: 10.1086/318261.

[31] Andreas A. Berlind and David H. Weinberg. The Halo Occupation Distribution: Toward an Empirical Determination of the Relation between Galaxies and Mass. ApJ, 575(2):587–616, August 2002. doi: 10.1086/341469.

[32] Asantha Cooray and Ravi Sheth. Halo models of large scale structure. Phys. Rep., 372(1):1–129, December 2002. doi: 10.1016/S0370-1573(02)00276-4.

[33] Zheng Zheng, Andreas A. Berlind, David H. Weinberg, Andrew J. Benson, Carlton M. Baugh, Shaun Cole, Romeel Davé, Carlos S. Frenk, Neal Katz, and Cedric G. Lacey. Theoretical Models of the Halo Occupation Distribution: Separating Central and Satellite Galaxies. ApJ, 633(2):791–809, November 2005. doi: 10.1086/466510.

[34] Zheng Zheng, Alison L. Coil, and Idit Zehavi. Galaxy Evolution from Halo Occupation Distribution Modeling of DEEP2 and SDSS Galaxy Clustering. ApJ, 667(2):760–779, October 2007. doi: 10.1086/521074.

[35] Zheng Zheng, Idit Zehavi, Daniel J. Eisenstein, David H. Weinberg, and Y. P. Jing. Halo Occupation Distribution Modeling of Clustering of Luminous Red Galaxies. ApJ, 707(1):554–572, December 2009. doi: 10.1088/0004-637X/707/1/554.

[36] Martin White, M. Blanton, A. Bolton, D. Schlegel, J. Tinker, A. Berlind, L. da Costa, E. Kazin, Y. T. Lin, M. Maia, C. K. McBride, N. Padmanabhan, J. Parejko, W. Percival, F. Prada, B. Ramos, E. Sheldon, F. de Simoni, R. Skibba, D. Thomas, D. Wake, I. Zehavi, Z. Zheng, R. Nichol, Donald P. Schneider, Michael A. Strauss, B. A. Weaver, and David H. Weinberg. The Clustering of Massive Galaxies at z

~0.5 from the First Semester of BOSS Data. ApJ, 728(2):126, February 2011. doi: 10.1088/0004-637X/728/2/126.

[37] Zhongxu Zhai, Jeremy L. Tinker, ChangHoon Hahn, Hee-Jong Seo, Michael R. Blanton, Rita Tojeiro, Hugo O. Camacho, Marcos Lima, Aurelio Carnero Rosell, Flavia Sobreira, Luiz N. da Costa, Julian E. Bautista, Joel R. Brownstein, Johan Comparat, Kyle Dawson, Jeffrey A. Newman, Abhishek Prakash, Alexandre Roman-Lopes, and Donald P. Schneider. The Clustering of Luminous Red Galaxies at z $\sim$ 0.7 from EBOSS and BOSS Data. ApJ, 848(2):76, October 2017. doi: 10.3847/1538-4357/aa8eee.

[38] Shadab Alam, John A. Peacock, Katarina Kraljic, Ashley J. Ross, and Johan Comparat. Multitracer extension of the halo model: probing quenching and conformity in eBOSS. MNRAS, 497(1):581–595, September 2020. doi: 10.1093/mnras/staa1956.

[39] Graziano Rossi, Peter D. Choi, Jeongin Moon, Julian E. Bautista, Hector Gil-Marín, Romain Paviot, Mariana Vargas-Magaña, Sylvain de la Torre, Sebastien Fromenteau, Ashley J. Ross, Santiago Ávila, Etienne Burtin, Kyle S. Dawson, Stéphanie Escoffier, Salman Habib, Katrin Heitmann, Jiamin Hou, Eva-Maria Mueller, Will J. Percival, Alex Smith, Cheng Zhao, and Gong-Bo Zhao. The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: N-body mock challenge for galaxy clustering measurements. MNRAS, 505(1):377–407, July 2021. doi: 10.1093/mnras/staa3955.

[40] Rongpu Zhou, Jeffrey A. Newman, Yao-Yuan Mao, Aaron Meisner, John Moustakas, Adam D. Myers, Abhishek Prakash, Andrew R. Zentner, David Brooks, Yutong Duan, Martin Landriau, Michael E. Levi, Francisco Prada, and Gregory Tarle. The clustering of DESI-like luminous red galaxies using photometric redshifts. MNRAS, 501(3):3309–3331, March 2021. doi: 10.1093/mnras/staa3764.

[41] S. Avila, V. Gonzalez-Perez, F. G. Mohammad, A. de Mattia, C. Zhao, A. Raichoor, A. Tamone, S. Alam, J. Bautista, D. Bianchi, E. Burtin, M. J. Chapman, C. H. Chuang, J. Comparat, K. Dawson, T. Divers, H. du Mas des Bourboux, H. Gil-Marin, E. M. Mueller, S. Habib, K. Heitmann, V. Ruhlmann-Kleider, N. Padilla, W. J. Percival,

A. J. Ross, H. J. Seo, D. P. Schneider, and G. Zhao. The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: exploring the halo occupation distribution model for emission line galaxies. MNRAS, 499(4):5486–5507, December 2020. doi: 10.1093/mnras/staa2951.

[42] Jonathan Richardson, Zheng Zheng, Suchetana Chatterjee, Daisuke Nagai, and Yue Shen. The Halo Occupation Distribution of SDSS Quasars. ApJ, 755(1):30, August 2012. doi: 10.1088/0004-637X/755/1/30.

[43] Alex Smith, Etienne Burtin, Jiamin Hou, Richard Neveux, Ashley J. Ross, Shadab Alam, Jonathan Brinkmann, Kyle S. Dawson, Salman Habib, Katrin Heitmann, Jean-Paul Kneib, Brad W. Lyke, Hélion du Mas des Bourboux, Eva-Maria Mueller, Adam D. Myers, Will J. Percival, Graziano Rossi, Donald P. Schneider, Pauline Zarrouk, and Gong-Bo Zhao. The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: N-body mock challenge for the quasar sample. MNRAS, 499(1):269–291, November 2020. doi: 10.1093/mnras/staa2825.

[44] Darren J. Croton, Liang Gao, and Simon D. M. White. Halo assembly bias and its effects on galaxy clustering. MNRAS, 374(4):1303–1309, February 2007. doi: 10.1111/j.1365-2966.2006.11230.x.

[45] Liang Gao, Volker Springel, and Simon D. M. White. The age dependence of halo clustering. MNRAS, 363(1):L66–L70, October 2005. doi: 10.1111/j.1745-3933.2005.00084.x.

[46] Arnau Pujol, Kai Hoffmann, Noelia Jiménez, and Enrique Gaztañaga. What determines large scale galaxy clustering: halo mass or local density? A&A, 598:A103, February 2017. doi: 10.1051/0004-6361/201629121.

[47] M. Celeste Artale, Idit Zehavi, Sergio Contreras, and Peder Norberg. The impact of assembly bias on the halo occupation in hydrodynamical simulations. MNRAS, 480(3):3978–3992, November 2018. doi: 10.1093/mnras/sty2110.

[48] Idit Zehavi, Sergio Contreras, Nelson Padilla, Nicholas J. Smith, Carlton M. Baugh, and Peder Norberg. The Impact of Assembly Bias on the Galaxy Content of Dark Matter Halos. ApJ, 853(1):84, January 2018. doi: 10.3847/1538-4357/aaa54a.

[49] V. Gonzalez-Perez, J. Comparat, P. Norberg, C. M. Baugh, S. Contreras, C. Lacey, N. McCullagh, A. Orsi, J. Helly, and J. Humphries. The host dark matter haloes of [O II] emitters at 0.5 ¡ z ¡ 1.5. MNRAS, 474(3):4024–4038, March 2018. doi: 10.1093/mnras/stx2807.

[50] S. Contreras, I. Zehavi, N. Padilla, C. M. Baugh, E. Jiménez, and I. Lacerna. The evolution of assembly bias. MNRAS, 484(1):1133–1148, March 2019. doi: 10.1093/mnras/stz018.

[51] G. Favole, V. Gonzalez-Perez, D. Stoppacher, Á. Orsi, J. Comparat, S. A. Cora, C. A. Vega-Martínez, A. R. H. Stevens, C. Maraston, D. Croton, A. Knebe, A. J. Benson, A. D. Montero-Dorta, N. Padilla, F. Prada, and D. Thomas. [O II] emitters in MultiDark-Galaxies and DEEP2. MNRAS, 497(4):5432–5453, October 2020. doi: 10.1093/mnras/staa2292.

[52] V. Gonzalez-Perez, W. Cui, S. Contreras, C. M. Baugh, J. Comparat, A. J. Griffin, J. Helly, A. Knebe, C. Lacey, and P. Norberg. Do model emission line galaxies live in filaments at z $\sim$ 1? MNRAS, 498(2):1852–1870, October 2020. doi: 10.1093/mnras/staa2504.

[53] Boryana Hadzhiyska, Lars Hernquist, Daniel Eisenstein, Ana Maria Delgado, Sownak Bose, Rahul Kannan, Rüdiger Pakmor, Volker Springel, Sergio Contreras, Monica Barrera, Fulvio Ferlito, César Hernández-Aguayo, Simon D. M. White, and Carlos Frenk. The MillenniumTNG Project: Refining the one-halo model of red and blue galaxies at different redshifts. *arXiv e-prints*, art. arXiv:2210.10068, October 2022. doi: 10.48550/arXiv.2210.10068.

[54] Simone M. Weinmann, Frank C. van den Bosch, Xiaohu Yang, and H. J. Mo. Properties

of galaxy groups in the Sloan Digital Sky Survey - I. The dependence of colour, star formation and morphology on halo mass. MNRAS, 366(1):2–28, February 2006. doi: 10.1111/j.1365-2966.2005.09865.x.

[55] Hong Guo, Zheng Zheng, Idit Zehavi, Kyle Dawson, Ramin A. Skibba, Jeremy L. Tinker, David H. Weinberg, Martin White, and Donald P. Schneider. Velocity bias from the small-scale clustering of SDSS-III BOSS galaxies. MNRAS, 446(1):578–594, January 2015. doi: 10.1093/mnras/stu2120.

[56] Hong Guo, Zheng Zheng, Idit Zehavi, Peter S. Behroozi, Chia-Hsun Chuang, Johan Comparat, Ginevra Favole, Stefan Gottloeber, Anatoly Klypin, Francisco Prada, David H. Weinberg, and Gustavo Yepes. Redshift-space clustering of SDSS galaxies - luminosity dependence, halo occupation distribution, and velocity bias. MNRAS, 453 (4):4368–4383, November 2015. doi: 10.1093/mnras/stv1966.

[57] Jia-Ni Ye, Hong Guo, Zheng Zheng, and Idit Zehavi. Properties and Origin of Galaxy Velocity Bias in the Illustris Simulation. ApJ, 841(1):45, May 2017. doi: 10.3847/1538-4357/aa70e7.

[58] Sihan Yuan, Daniel J. Eisenstein, and Lehman H. Garrison. Exploring the squeezed three-point galaxy correlation function with generalized halo occupation distribution models. MNRAS, 478(2):2019–2033, August 2018. doi: 10.1093/mnras/sty1089.

[59] P. J. E. Peebles and M. G. Hauser. Statistical Analysis of Catalogs of Extragalactic Objects. III. The Shane-Wirtanen and Zwicky Catalogs. ApJS, 28:19, November 1974. doi: 10.1086/190308.

[60] M. Davis and P. J. E. Peebles. A survey of galaxy redshifts. V. The two-point position and velocity correlations. ApJ, 267:465–482, April 1983. doi: 10.1086/160884.

[61] A. J. S. Hamilton. Toward Better Ways to Measure the Galaxy Correlation Function. ApJ, 417:19, November 1993. doi: 10.1086/173288.

[62] Stephen D. Landy and Alexander S. Szalay. Bias and Variance of Angular Correlation Functions. ApJ, 412:64, July 1993. doi: 10.1086/172900.

[63] Sihan Yuan, Lehman H. Garrison, Boryana Hadzhiyska, Sownak Bose, and Daniel J. Eisenstein. ABACUSHOD: a highly efficient extended multitracer HOD framework and its application to BOSS and eBOSS data. MNRAS, 510(3):3301–3320, March 2022. doi: 10.1093/mnras/stab3355.

[64] Yu Wang, Xiaohu Yang, H. J. Mo, Frank C. van den Bosch, and YaoQuan Chu. The three-point correlation function of galaxies: comparing halo occupation models with observations. MNRAS, 353(1):287–300, September 2004. doi: 10.1111/j.1365-2966.2004.08141.x.

[65] E. Gaztañaga and R. Scoccimarro. The three-point function in large-scale structure: redshift distortions and galaxy bias. MNRAS, 361(3):824–836, August 2005. doi: 10.1111/j.1365-2966.2005.09234.x.

[66] Pablo Fosalba, Jun Pan, and István Szapudi. Cosmological Three-Point Function: Testing the Halo Model against Simulations. ApJ, 632(1):29–48, October 2005. doi: 10.1086/432906.

[67] Felipe A. Marín, Risa H. Wechsler, Joshua A. Frieman, and Robert C. Nichol. Modeling the Galaxy Three-Point Correlation Function. ApJ, 672(2):849–860, January 2008. doi: 10.1086/523628.

[68] Gauri V. Kulkarni, Robert C. Nichol, Ravi K. Sheth, Hee-Jong Seo, Daniel J. Eisenstein, and Alexander Gray. The three-point correlation function of luminous red galaxies in the Sloan Digital Sky Survey. MNRAS, 378(3):1196–1206, July 2007. doi: 10.1111/j.1365-2966.2007.11872.x.

[69] Felipe Marín. The Large-scale Three-point Correlation Function of Sloan Digital Sky Survey Luminous Red Galaxies. ApJ, 737(2):97, August 2011. doi: 10.1088/0004-637X/737/2/97.

[70] H. Guo, Z. Zheng, Y. P. Jing, I. Zehavi, C. Li, D. H. Weinberg, R. A. Skibba, R. C. Nichol, G. Rossi, C. G. Sabiu, D. P. Schneider, and C. K. McBride. Modelling the redshift-space three-point correlation function in SDSS-III. MNRAS, 449:L95–L99, April 2015. doi: 10.1093/mnrasl/slv020.

[71] Beth A. Reid and David N. Spergel. Constraining the Luminous Red Galaxy Halo Occupation Distribution Using Counts-In-Cylinders. ApJ, 698(1):143–154, June 2009. doi: 10.1088/0004-637X/698/1/143.

[72] Heather D. Berrier, Elizabeth J. Barton, Joel C. Berrier, James S. Bullock, Andrew R. Zentner, and Risa H. Wechsler. Counts-in-cylinders in the Sloan Digital Sky Survey with Comparisons to N-body Simulations. ApJ, 726(1):1, January 2011. doi: 10.1088/0004-637X/726/1/1.

[73] Masamune Oguri and Yen-Ting Lin. Inferring Host Dark Matter Halo Masses of Individual Galaxies from Neighboring Galaxy Counts. ApJ, 801(2):94, March 2015. doi: 10.1088/0004-637X/801/2/94.

[74] D. Gruen, O. Friedrich, E. Krause, J. DeRose, R. Cawthon, C. Davis, J. Elvin-Poole, E. S. Rykoff, R. H. Wechsler, A. Alarcon, G. M. Bernstein, J. Blazek, C. Chang, J. Clampitt, M. Crocce, J. De Vicente, M. Gatti, M. S. S. Gill, W. G. Hartley, S. Hilbert, B. Hoyle, B. Jain, M. Jarvis, O. Lahav, N. MacCrann, T. McClintock, J. Prat, R. P. Rollins, A. J. Ross, E. Rozo, S. Samuroff, C. Sánchez, E. Sheldon, M. A. Troxel, J. Zuntz, T. M. C. Abbott, F. B. Abdalla, S. Allam, J. Annis, K. Bechtol, A. Benoit-Lévy, E. Bertin, S. L. Bridle, D. Brooks, E. Buckley-Geer, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, C. E. Cunha, C. B. D'Andrea, L. N. da Costa, S. Desai, H. T. Diehl, J. P. Dietrich, P. Doel, A. Drlica-Wagner, E. Fernandez, B. Flaugher, P. Fosalba, J. Frieman, J. García-Bellido, E. Gaztanaga, T. Giannantonio, R. A. Gruendl, J. Gschwend, G. Gutierrez, K. Honscheid, D. J. James, T. Jeltema, K. Kuehn, N. Kuropatkin, M. Lima, M. March, J. L. Marshall, P. Martini, P. Melchior, F. Menanteau, R. Miquel, J. J. Mohr, A. A. Plazas, A. Roodman, E. Sanchez, V. Scarpine,

M. Schubnell, I. Sevilla-Noarbe, M. Smith, R. C. Smith, M. Soares-Santos, F. Sobreira, M. E. C. Swanson, G. Tarle, D. Thomas, V. Vikram, A. R. Walker, J. Weller, Y. Zhang, and DES Collaboration. Density split statistics: Cosmological constraints from counts and lensing in cells in DES Y1 and SDSS data. Phys. Rev. D, 98(2): 023507, July 2018. doi: 10.1103/PhysRevD.98.023507.

[75] Kuan Wang, Yao-Yuan Mao, Andrew R. Zentner, Frank C. van den Bosch, Johannes U. Lange, Chad M. Schafer, Antonia Sierra Villarreal, Andrew P. Hearin, and Duncan Campbell. How to optimally constrain galaxy assembly bias: supplement projected correlation functions with count-in-cells statistics. MNRAS, 488(3):3541–3567, September 2019. doi: 10.1093/mnras/stz1733.

[76] Frank C. van den Bosch, Peder Norberg, H. J. Mo, and Xiaohu Yang. Probing dark matter haloes with satellite kinematics. MNRAS, 352(4):1302–1314, August 2004. doi: 10.1111/j.1365-2966.2004.08021.x.

[77] Surhud More, Frank C. van den Bosch, Marcello Cacciato, Ramin Skibba, H. J. Mo, and Xiaohu Yang. Satellite kinematics - III. Halo masses of central galaxies in SDSS. MNRAS, 410(1):210–226, January 2011. doi: 10.1111/j.1365-2966.2010.17436.x.

[78] Ravi K. Sheth and Rien van de Weygaert. A hierarchy of voids: much ado about nothing. MNRAS, 350(2):517–538, May 2004. doi: 10.1111/j.1365-2966.2004.07661.x.

[79] Steven R. Furlanetto and Tsvi Piran. The evidence of absence: galaxy voids in the excursion set formalism. MNRAS, 366(2):467–479, February 2006. doi: 10.1111/j.1365-2966.2005.09862.x.

[80] Jeremy L. Tinker, Charlie Conroy, Peder Norberg, Santiago G. Patiri, David H. Weinberg, and Michael S. Warren. Void Statistics in Large Galaxy Redshift Surveys: Does Halo Occupation of Field Galaxies Depend on Environment? ApJ, 686(1):53–71, October 2008. doi: 10.1086/589983.

[81] Francisco-Shu Kitaura, Chia-Hsun Chuang, Yu Liang, Cheng Zhao, Charling Tao, Sergio Rodríguez-Torres, Daniel J. Eisenstein, Héctor Gil-Marín, Jean-Paul Kneib, Cameron McBride, Will J. Percival, Ashley J. Ross, Ariel G. Sánchez, Jeremy Tinker, Rita Tojeiro, Mariana Vargas-Magana, and Gong-Bo Zhao. Signatures of the Primordial Universe from Its Emptiness: Measurement of Baryon Acoustic Oscillations from Minima of the Density Field. Phys. Rev. Lett., 116(17):171301, April 2016. doi: 10.1103/PhysRevLett.116.171301.

[82] Eyal Neistein, Cheng Li, Sadegh Khochfar, Simone M. Weinmann, Francesco Shankar, and Michael Boylan-Kolchin. A tale of two populations: the stellar mass of central and satellite galaxies. MNRAS, 416(2):1486–1499, September 2011. doi: 10.1111/j.1365-2966.2011.19145.x.

[83] Zheng Zheng and Hong Guo. Accurate and efficient halo-based galaxy clustering modelling with simulations. MNRAS, 458(4):4015–4024, June 2016. doi: 10.1093/mnras/stw523.

[84] Hanyu Zhang, Lado Samushia, David Brooks, Axel de la Macorra, Peter Doel, Enrique Gaztañaga, Satya Gontcho A Gontcho, Klaus Honscheid, Robert Kehoe, Theodore Kisner, Aaron Meisner, Claire Poppett, Michael Schubnell, Gregory Tarle, Kai Zhang, and Hu Zou. Constraining galaxy-halo connection with high-order statistics. MNRAS, 515(4):6133–6150, October 2022. doi: 10.1093/mnras/stac2147.

[85] Edmund Bertschinger. Simulations of Structure Formation in the Universe. ARA&A, 36:599–654, January 1998. doi: 10.1146/annurev.astro.36.1.599.

[86] Mark Vogelsberger, Federico Marinacci, Paul Torrey, and Ewald Puchwein. Cosmological simulations of galaxy formation. *Nature Reviews Physics*, 2(1):42–66, January 2020. doi: 10.1038/s42254-019-0127-2.

[87] J. S. Bagla. Cosmological N-Body simulation: Techniques, Scope and Status. *Current Science*, 88:1088–1100, April 2005. doi: 10.48550/arXiv.astro-ph/0411043.

[88] W. Dehnen and J. I. Read. N-body simulations of gravitational dynamics. *European Physical Journal Plus*, 126:55, May 2011. doi: 10.1140/epjp/i2011-11055-3.

[89] Mark Vogelsberger, Shy Genel, Volker Springel, Paul Torrey, Debora Sijacki, Dandan Xu, Greg Snyder, Dylan Nelson, and Lars Hernquist. Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe. MNRAS, 444 (2):1518–1547, October 2014. doi: 10.1093/mnras/stu1536.

[90] Joop Schaye, Robert A. Crain, Richard G. Bower, Michelle Furlong, Matthieu Schaller, Tom Theuns, Claudio Dalla Vecchia, Carlos S. Frenk, I. G. McCarthy, John C. Helly, Adrian Jenkins, Y. M. Rosas-Guevara, Simon D. M. White, Maarten Baes, C. M. Booth, Peter Camps, Julio F. Navarro, Yan Qu, Alireza Rahmati, Till Sawala, Peter A. Thomas, and James Trayford. The EAGLE project: simulating the evolution and assembly of galaxies and their environments. MNRAS, 446(1):521–554, January 2015. doi: 10.1093/mnras/stu2058.

[91] A. Vale and J. P. Ostriker. Linking halo mass to galaxy luminosity. MNRAS, 353(1): 189–200, September 2004. doi: 10.1111/j.1365-2966.2004.08059.x.

[92] A. Vale and J. P. Ostriker. The non-parametric model for linking galaxy luminosity with halo/subhalo mass. MNRAS, 371(3):1173–1187, September 2006. doi: 10.1111/j.1365-2966.2006.10605.x.

[93] Hong Guo, Zheng Zheng, Peter S. Behroozi, Idit Zehavi, Chia-Hsun Chuang, Johan Comparat, Ginevra Favole, Stefan Gottloeber, Anatoly Klypin, Francisco Prada, Sergio A. Rodríguez-Torres, David H. Weinberg, and Gustavo Yepes. Modelling galaxy clustering: halo occupation distribution versus subhalo matching. MNRAS, 459(3): 3040–3058, July 2016. doi: 10.1093/mnras/stw845.

[94] Xiaohu Yang, H. J. Mo, and Frank C. van den Bosch. Constraining galaxy formation and cosmology with the conditional luminosity function of galaxies. MNRAS, 339(4): 1057–1080, March 2003. doi: 10.1046/j.1365-8711.2003.06254.x.

[95] Asantha Cooray. Halo model at its best: constraints on conditional luminosity functions from measured galaxy statistics. MNRAS, 365(3):842–866, January 2006. doi: 10.1111/j.1365-2966.2005.09747.x.

[96] Asantha Cooray and Masami Ouchi. Where are z = 4 Lyman Break Galaxies? Results from conditional luminosity function models of luminosity-dependent clustering. MNRAS, 369(4):1869–1879, July 2006. doi: 10.1111/j.1365-2966.2006.10437.x.

[97] Lingyu Wang, Asantha Cooray, and Seb Oliver. A Halo Model of Local IRAS Galaxies Selected at 60 Micron Using Conditional Luminosity Functions. *arXiv e-prints*, art. arXiv:1010.5484, October 2010. doi: 10.48550/arXiv.1010.5484.

[98] K. Hoffmann, E. Gaztañaga, R. Scoccimarro, and M. Crocce. Testing the consistency of three-point halo clustering in Fourier and configuration space. MNRAS, 476(1): 814–829, May 2018. doi: 10.1093/mnras/sty187.

[99] K. Hoffmann, J. Bel, and E. Gaztañaga. Linear and non-linear bias: predictions versus measurements. MNRAS, 465(2):2225–2235, February 2017. doi: 10.1093/mnras/stw2876.

[100] Sihan Yuan, Daniel J. Eisenstein, and Lehman H. Garrison. Using galaxy pairs to investigate the three-point correlation function in the squeezed limit. MNRAS, 472(1): 577–590, November 2017. doi: 10.1093/mnras/stx2032.

[101] Zachary Slepian and Daniel J. Eisenstein. Computing the three-point correlation function of galaxies in O(N$\hat{2}$) time. MNRAS, 454(4):4142–4158, December 2015. doi: 10.1093/mnras/stv2119.

[102] Zachary Slepian and Daniel J. Eisenstein. Accelerating the two-point and three-point galaxy correlation functions using Fourier transforms. MNRAS, 455(1):L31–L35, January 2016. doi: 10.1093/mnrasl/slv133.

[103] Oliver H. E. Philcox, Zachary Slepian, Jiamin Hou, Craig Warner, Robert N. Cahn, and

Daniel J. Eisenstein. ENCORE: an O $(\mathrm{N}_g{}^2)$ estimator for galaxy N-point correlation functions. MNRAS, 509(2):2457–2481, January 2022. doi: 10.1093/mnras/stab3025.

[104] DESI Collaboration, Amir Aghamousa, Jessica Aguilar, Steve Ahlen, Shadab Alam, Lori E. Allen, Carlos Allende Prieto, James Annis, Stephen Bailey, Christophe Balland, Otger Ballester, Charles Baltay, Lucas Beaufore, Chris Bebek, Timothy C. Beers, Eric F. Bell, José Luis Bernal, Robert Besuner, Florian Beutler, Chris Blake, Hannes Bleuler, Michael Blomqvist, Robert Blum, Adam S. Bolton, Cesar Briceno, David Brooks, Joel R. Brownstein, Elizabeth Buckley-Geer, Angela Burden, Etienne Burtin, Nicolas G. Busca, Robert N. Cahn, Yan-Chuan Cai, Laia Cardiel-Sas, Raymond G. Carlberg, Pierre-Henri Carton, Ricard Casas, Francisco J. Castander, Jorge L. Cervantes-Cota, Todd M. Claybaugh, Madeline Close, Carl T. Coker, Shaun Cole, Johan Comparat, Andrew P. Cooper, M. C. Cousinou, Martin Crocce, Jean-Gabriel Cuby, Daniel P. Cunningham, Tamara M. Davis, Kyle S. Dawson, Axel de la Macorra, Juan De Vicente, Timothée Delubac, Mark Derwent, Arjun Dey, Govinda Dhungana, Zhejie Ding, Peter Doel, Yutong T. Duan, Anne Ealet, Jerry Edelstein, Sarah Eftekharzadeh, Daniel J. Eisenstein, Ann Elliott, Stéphanie Escoffier, Matthew Evatt, Parker Fagrelius, Xiaohui Fan, Kevin Fanning, Arya Farahi, Jay Farihi, Ginevra Favole, Yu Feng, Enrique Fernandez, Joseph R. Findlay, Douglas P. Finkbeiner, Michael J. Fitzpatrick, Brenna Flaugher, Samuel Flender, Andreu Font-Ribera, Jaime E. Forero-Romero, Pablo Fosalba, Carlos S. Frenk, Michele Fumagalli, Boris T. Gaensicke, Giuseppe Gallo, Juan Garcia-Bellido, Enrique Gaztanaga, Nicola Pietro Gentile Fusillo, Terry Gerard, Irena Gershkovich, Tommaso Giannantonio, Denis Gillet, Guillermo Gonzalez-de-Rivera, Violeta Gonzalez-Perez, Shelby Gott, Or Graur, Gaston Gutierrez, Julien Guy, Salman Habib, Henry Heetderks, Ian Heetderks, Katrin Heitmann, Wojciech A. Hellwing, David A. Herrera, Shirley Ho, Stephen Holland, Klaus Honscheid, Eric Huff, Timothy A. Hutchinson, Dragan Huterer, Ho Seong Hwang, Joseph Maria Illa Laguna, Yuzo Ishikawa, Dianna Jacobs, Niall Jeffrey, Patrick Jelinsky, Elise Jennings, Linhua Jiang, Jorge Jimenez, Jennifer

Johnson, Richard Joyce, Eric Jullo, Stéphanie Juneau, Sami Kama, Armin Karcher, Sonia Karkar, Robert Kehoe, Noble Kennamer, Stephen Kent, Martin Kilbinger, Alex G. Kim, David Kirkby, Theodore Kisner, Ellie Kitanidis, Jean-Paul Kneib, Sergey Koposov, Eve Kovacs, Kazuya Koyama, Anthony Kremin, Richard Kron, Luzius Kronig, Andrea Kueter-Young, Cedric G. Lacey, Robin Lafever, Ofer Lahav, Andrew Lambert, Michael Lampton, Martin Landriau, Dustin Lang, Tod R. Lauer, Jean-Marc Le Goff, Laurent Le Guillou, Auguste Le Van Suu, Jae Hyeon Lee, Su-Jeong Lee, Daniela Leitner, Michael Lesser, Michael E. Levi, Benjamin L'Huillier, Baojiu Li, Ming Liang, Huan Lin, Eric Linder, Sarah R. Loebman, Zarija Lukić, Jun Ma, Niall MacCrann, Christophe Magneville, Laleh Makarem, Marc Manera, Christopher J. Manser, Robert Marshall, Paul Martini, Richard Massey, Thomas Matheson, Jeremy McCauley, Patrick McDonald, Ian D. McGreer, Aaron Meisner, Nigel Metcalfe, Timothy N. Miller, Ramon Miquel, John Moustakas, Adam Myers, Milind Naik, Jeffrey A. Newman, Robert C. Nichol, Andrina Nicola, Luiz Nicolati da Costa, Jundan Nie, Gustavo Niz, Peder Norberg, Brian Nord, Dara Norman, Peter Nugent, Thomas O'Brien, Minji Oh, Knut A. G. Olsen, Cristobal Padilla, Hamsa Padmanabhan, Nikhil Padmanabhan, Nathalie Palanque-Delabrouille, Antonella Palmese, Daniel Pappalardo, Isabelle Pâris, Changbom Park, Anna Patej, John A. Peacock, Hiranya V. Peiris, Xiyan Peng, Will J. Percival, Sandrine Perruchot, Matthew M. Pieri, Richard Pogge, Jennifer E. Pollack, Claire Poppett, Francisco Prada, Abhishek Prakash, Ronald G. Probst, David Rabinowitz, Anand Raichoor, Chang Hee Ree, Alexandre Refregier, Xavier Regal, Beth Reid, Kevin Reil, Mehdi Rezaie, Constance M. Rockosi, Natalie Roe, Samuel Ronayette, Aaron Roodman, Ashley J. Ross, Nicholas P. Ross, Graziano Rossi, Eduardo Rozo, Vanina Ruhlmann-Kleider, Eli S. Rykoff, Cristiano Sabiu, Lado Samushia, Eusebio Sanchez, Javier Sanchez, David J. Schlegel, Michael Schneider, Michael Schubnell, Aurélia Secroun, Uros Seljak, Hee-Jong Seo, Santiago Serrano, Arman Shafieloo, Huanyuan Shan, Ray Sharples, Michael J. Sholl, William V. Shourt, Joseph H. Silber, David R. Silva, Martin M. Sirk, Anze Slosar, Alex Smith, George F. Smoot, Debopam Som, Yong-Seon Song, David Sprayberry, Ryan Staten, Andy Ste-

fanik, Gregory Tarle, Suk Sien Tie, Jeremy L. Tinker, Rita Tojeiro, Francisco Valdes, Octavio Valenzuela, Monica Valluri, Mariana Vargas-Magana, Licia Verde, Alistair R. Walker, Jiali Wang, Yuting Wang, Benjamin A. Weaver, Curtis Weaverdyck, Risa H. Wechsler, David H. Weinberg, Martin White, Qian Yang, Christophe Yeche, Tianmeng Zhang, Gong-Bo Zhao, Yi Zheng, Xu Zhou, Zhimin Zhou, Yaling Zhu, Hu Zou, and Ying Zu. The DESI Experiment Part I: Science,Targeting, and Survey Design. *arXiv e-prints*, art. arXiv:1611.00036, October 2016. doi: 10.48550/arXiv.1611.00036.

[105] Lehman H. Garrison, Daniel J. Eisenstein, Douglas Ferrer, Nina A. Maksimova, and Philip A. Pinto. The ABACUS cosmological N-body code. MNRAS, 508(1):575–596, November 2021. doi: 10.1093/mnras/stab2482.

[106] Sownak Bose, Daniel J. Eisenstein, Boryana Hadzhiyska, Lehman H. Garrison, and Sihan Yuan. Constructing high-fidelity halo merger trees in ABACUSSUMMIT. MNRAS, 512(1):837–854, May 2022. doi: 10.1093/mnras/stac555.

[107] Lehman H. Garrison, Daniel J. Eisenstein, and Philip A. Pinto. A high-fidelity realization of the Euclid code comparison N-body simulation with ABACUS. MNRAS, 485(3):3370–3377, May 2019. doi: 10.1093/mnras/stz634.

[108] Lehman H. Garrison, Daniel J. Eisenstein, Douglas Ferrer, Jeremy L. Tinker, Philip A. Pinto, and David H. Weinberg. The Abacus Cosmos: A Suite of Cosmological N-body Simulations. ApJS, 236(2):43, June 2018. doi: 10.3847/1538-4365/aabfd3.

[109] Lehman H. Garrison, Daniel J. Eisenstein, Douglas Ferrer, Marc V. Metchnik, and Philip A. Pinto. Improving initial conditions for cosmological N-body simulations. MNRAS, 461(4):4125–4145, October 2016. doi: 10.1093/mnras/stw1594.

[110] Marc Victor Leonard Metchnik. *A fast N-body scheme for computational cosmology*. PhD thesis, University of Arizona, January 2009.

[111] Boryana Hadzhiyska, Daniel Eisenstein, Sownak Bose, Lehman H. Garrison, and Nina

Maksimova. COMPASO: A new halo finder for competitive assignment to spherical overdensities. MNRAS, 509(1):501–521, January 2022. doi: 10.1093/mnras/stab2980.

[112] P. J. E. Peebles. *The large-scale structure of the universe.* 1980.

[113] Zheng Zheng. Projected Three-Point Correlation Functions and Galaxy Bias. ApJ, 614(2):527–532, October 2004. doi: 10.1086/423838.

[114] P. Norberg, C. M. Baugh, E. Gaztañaga, and D. J. Croton. Statistical analysis of galaxy surveys - I. Robust error estimation for two-point clustering statistics. MNRAS, 396 (1):19–38, June 2009. doi: 10.1111/j.1365-2966.2009.14389.x.

[115] Frank C. van den Bosch, Surhud More, Marcello Cacciato, Houjun Mo, and Xiaohu Yang. Cosmological constraints from a combination of galaxy clustering and lensing - I. Theoretical framework. MNRAS, 430(2):725–746, April 2013. doi: 10.1093/mnras/sts006.

[116] Manodeep Sinha and Lehman H. Garrison. CORRFUNC - a suite of blazing fast correlation functions on the CPU. MNRAS, 491(2):3022–3041, January 2020. doi: 10.1093/mnras/stz3157.

[117] Manodeep Sinha and Lehman H. Garrison. Corrfunc: Blazing fast correlation functions with AVX512F SIMD Intrinsics. In *Software Challenges to Exascale Computing. Second Workshop*, pages 3–20, January 2019. doi: 10.1007/978-981-13-7729-7_1.

[118] Will J. Percival, Oliver Friedrich, Elena Sellentin, and Alan Heavens. Matching Bayesian and frequentist coverage probabilities when using an approximate data covariance matrix. MNRAS, 510(3):3207–3221, March 2022. doi: 10.1093/mnras/stab3540.

[119] Antony Lewis and Sarah Bridle. Cosmological parameters from CMB and other data: A Monte Carlo approach. Phys. Rev. D, 66(10):103511, November 2002. doi: 10.1103/PhysRevD.66.103511.

[120] Andrew Gelman and Donald B. Rubin. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7:457–472, January 1992. doi: 10.1214/ss/1177011136.

[121] Stephen P. Brooks and Andrew Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455, 1998. doi: 10.1080/10618600.1998.10474787.

[122] David W. Pearson and Lado Samushia. Computing three-point correlation function randoms counts without the randoms catalogue. MNRAS, 486(1):L105–L109, June 2019. doi: 10.1093/mnrasl/slz067.

[123] S. D. M. White and M. J. Rees. Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering. MNRAS, 183:341–358, May 1978. doi: 10.1093/mnras/183.3.341.

[124] Idit Zehavi, Zheng Zheng, David H. Weinberg, Michael R. Blanton, Neta A. Bahcall, Andreas A. Berlind, Jon Brinkmann, Joshua A. Frieman, James E. Gunn, Robert H. Lupton, Robert C. Nichol, Will J. Percival, Donald P. Schneider, Ramin A. Skibba, Michael A. Strauss, Max Tegmark, and Donald G. York. Galaxy Clustering in the Completed SDSS Redshift Survey: The Dependence on Color and Luminosity. ApJ, 736(1):59, July 2011. doi: 10.1088/0004-637X/736/1/59.

[125] John K. Parejko, Tomomi Sunayama, Nikhil Padmanabhan, David A. Wake, Andreas A. Berlind, Dmitry Bizyaev, Michael Blanton, Adam S. Bolton, Frank van den Bosch, Jon Brinkmann, Joel R. Brownstein, Luiz Alberto Nicolaci da Costa, Daniel J. Eisenstein, Hong Guo, Eyal Kazin, Marcio Maia, Elena Malanushenko, Claudia Maraston, Cameron K. McBride, Robert C. Nichol, Daniel J. Oravetz, Kaike Pan, Will J. Percival, Francisco Prada, Ashley J. Ross, Nicholas P. Ross, David J. Schlegel, Don Schneider, Audrey E. Simmons, Ramin Skibba, Jeremy Tinker, Rita Tojeiro, Benjamin A. Weaver, Andrew Wetzel, Martin White, David H. Weinberg, Daniel Thomas, Idit Zehavi, and Zheng Zheng. The clustering of galaxies in the SDSS-III Baryon

Oscillation Spectroscopic Survey: the low-redshift sample. MNRAS, 429(1):98–112, February 2013. doi: 10.1093/mnras/sts314.

[126] Hong Guo, Zheng Zheng, Idit Zehavi, Haojie Xu, Daniel J. Eisenstein, David H. Weinberg, Neta A. Bahcall, Andreas A. Berlind, Johan Comparat, Cameron K. McBride, Ashley J. Ross, Donald P. Schneider, Ramin A. Skibba, Molly E. C. Swanson, Jeremy L. Tinker, Rita Tojeiro, and David A. Wake. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: modelling of the luminosity and colour dependence in the Data Release 10. MNRAS, 441(3):2398–2413, July 2014. doi: 10.1093/mnras/stu763.

[127] Sergio A. Rodríguez-Torres, Chia-Hsun Chuang, Francisco Prada, Hong Guo, Anatoly Klypin, Peter Behroozi, Chang Hoon Hahn, Johan Comparat, Gustavo Yepes, Antonio D. Montero-Dorta, Joel R. Brownstein, Claudia Maraston, Cameron K. McBride, Jeremy Tinker, Stefan Gottlöber, Ginevra Favole, Yiping Shu, Francisco-Shu Kitaura, Adam Bolton, Román Scoccimarro, Lado Samushia, David Schlegel, Donald P. Schneider, and Daniel Thomas. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: modelling the clustering and halo occupation distribution of BOSS CMASS galaxies in the Final Data Release. MNRAS, 460(2):1173–1187, August 2016. doi: 10.1093/mnras/stw1014.

[128] Johannes U. Lange, Xiaohu Yang, Hong Guo, Wentao Luo, and Frank C. van den Bosch. New perspectives on the BOSS small-scale lensing discrepancy for the Planck ΛCDM cosmology. MNRAS, 488(4):5771–5787, October 2019. doi: 10.1093/mnras/stz2124.

[129] Sihan Yuan, Boryana Hadzhiyska, Sownak Bose, Daniel J. Eisenstein, and Hong Guo. Evidence for galaxy assembly bias in BOSS CMASS redshift-space galaxy correlation function. MNRAS, 502(3):3582–3598, April 2021. doi: 10.1093/mnras/stab235.

[130] Kuan Wang, Yao-Yuan Mao, Andrew R. Zentner, Hong Guo, Johannes U. Lange, Frank C. van den Bosch, and Lorena Mezini. Evidence of galaxy assembly bias in SDSS

DR7 galaxy samples from count statistics. MNRAS, 516(3):4003–4024, November 2022. doi: 10.1093/mnras/stac2465.

[131] Laila Linke, Patrick Simon, Peter Schneider, Daniel J. Farrow, Jens Rödiger, and Angus H. Wright. KiDS+VIKING+GAMA: Halo occupation distributions and correlations of satellite numbers with a new halo model of the galaxy-matter bispectrum for galaxy-galaxy-galaxy lensing. A&A, 665:A38, September 2022. doi: 10.1051/0004-6361/202243711.

[132] Shadab Alam, Arnaud de Mattia, Amélie Tamone, S. Ávila, John A. Peacock, V. Gonzalez-Perez, Alex Smith, Anand Raichoor, Ashley J. Ross, Julian E. Bautista, Etienne Burtin, Johan Comparat, Kyle S. Dawson, Hélion du Mas des Bourboux, Stéphanie Escoffier, Héctor Gil-Marín, Salman Habib, Katrin Heitmann, Jiamin Hou, Faizan G. Mohammad, Eva-Maria Mueller, Richard Neveux, Romain Paviot, Will J. Percival, Graziano Rossi, Vanina Ruhlmann-Kleider, Rita Tojeiro, Mariana Vargas Magaña, Cheng Zhao, and Gong-Bo Zhao. The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: N-body mock challenge for the eBOSS emission line galaxy sample. MNRAS, 504(4):4667–4686, July 2021. doi: 10.1093/mnras/stab1150.

[133] Johannes U. Lange, Andrew P. Hearin, Alexie Leauthaud, Frank C. van den Bosch, Hong Guo, and Joseph DeRose. Five per cent measurements of the growth rate from simulation-based modelling of redshift-space clustering in BOSS LOWZ. MNRAS, 509 (2):1779–1804, January 2022. doi: 10.1093/mnras/stab3111.

[134] Yosuke Kobayashi, Takahiro Nishimichi, Masahiro Takada, and Hironao Miyatake. Full-shape cosmology analysis of the SDSS-III BOSS galaxy power spectrum using an emulator-based halo model: A 5% determination of σ_8. Phys. Rev. D, 105(8):083517, April 2022. doi: 10.1103/PhysRevD.105.083517.

[135] Michael J. Chapman, Faizan G. Mohammad, Zhongxu Zhai, Will J. Percival, Jeremy L. Tinker, Julian E. Bautista, Joel R. Brownstein, Etienne Burtin, Kyle S. Dawson, Héctor Gil-Marín, Axel de la Macorra, Ashley J. Ross, Graziano Rossi, Donald P.

Schneider, and Gong-Bo Zhao. The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the growth rate of structure from the small-scale clustering of the luminous red galaxy sample. MNRAS, 516(1):617–635, October 2022. doi: 10.1093/mnras/stac1923.

[136] Sihan Yuan, Lehman H. Garrison, Daniel J. Eisenstein, and Risa H. Wechsler. Stringent σ_8 constraints from small-scale galaxy clustering using a hybrid MCMC + emulator framework. MNRAS, 515(1):871–896, September 2022. doi: 10.1093/mnras/stac1830.

[137] Zhongxu Zhai, Jeremy L. Tinker, Arka Banerjee, Joseph DeRose, Hong Guo, Yao-Yuan Mao, Sean McLaughlin, Kate Storey-Fisher, and Risa H. Wechsler. The Aemulus Project. V. Cosmological Constraint from Small-scale Clustering of BOSS Galaxies. ApJ, 948(2):99, May 2023. doi: 10.3847/1538-4357/acc65b.

[138] Michael Levi, Chris Bebek, Timothy Beers, Robert Blum, Robert Cahn, Daniel Eisenstein, Brenna Flaugher, Klaus Honscheid, Richard Kron, Ofer Lahav, Patrick McDonald, Natalie Roe, David Schlegel, and representing the DESI collaboration. The DESI Experiment, a whitepaper for Snowmass 2013. *arXiv e-prints*, art. arXiv:1308.0847, August 2013. doi: 10.48550/arXiv.1308.0847.

[139] DESI Collaboration, A. G. Adame, J. Aguilar, S. Ahlen, S. Alam, G. Aldering, D. M. Alexander, R. Alfarsy, C. Allende Prieto, M. Alvarez, O. Alves, A. Anand, F. Andrade-Oliveira, E. Armengaud, J. Asorey, S. Avila, A. Aviles, S. Bailey, A. Balaguera-Antolínez, O. Ballester, C. Baltay, A. Bault, J. Bautista, J. Behera, S. F. Beltran, S. BenZvi, L. Beraldo e Silva, J. R. Bermejo-Climent, A. Berti, R. Besuner, F. Beutler, D. Bianchi, C. Blake, R. Blum, A. S. Bolton, S. Brieden, A. Brodzeller, D. Brooks, Z. Brown, E. Buckley-Geer, E. Burtin, L. Cabayol-Garcia, Z. Cai, R. Canning, L. Cardiel-Sas, A. Carnero Rosell, F. J. Castander, J. L. Cervantes-Cota, S. Chabanier, E. Chaussidon, J. Chaves-Montero, S. Chen, C. Chuang, T. Claybaugh, S. Cole, A. P. Cooper, A. Cuceu, T. M. Davis, K. Dawson, R. de Belsunce, R. de la Cruz, A. de la Macorra, A. de Mattia, R. Demina, U. Demirbozan, J. DeRose,

A. Dey, B. Dey, G. Dhungana, J. Ding, Z. Ding, P. Doel, R. Doshi, K. Douglass, A. Edge, S. Eftekharzadeh, D. J. Eisenstein, A. Elliott, S. Escoffier, P. Fagrelius, X. Fan, K. Fanning, V. A. Fawcett, S. Ferraro, J. Ereza, B. Flaugher, A. Font-Ribera, D. Forero-Sánchez, J. E. Forero-Romero, C. S. Frenk, B. T. Gänsicke, L. Á. García, J. García-Bellido, C. Garcia-Quintero, L. H. Garrison, H. Gil-Marín, J. Golden-Marx, S. Gontcho A Gontcho, A. X. Gonzalez-Morales, V. Gonzalez-Perez, C. Gordon, O. Graur, D. Green, D. Gruen, J. Guy, B. Hadzhiyska, C. Hahn, J. J. Han, M. M. S Hanif, H. K. Herrera-Alcantar, K. Honscheid, J. Hou, C. Howlett, D. Huterer, V. Iršič, M. Ishak, A. Jana, L. Jiang, J. Jimenez, Y. P. Jing, S. Joudaki, E. Jullo, S. Juneau, N. Kizhuprakkat, N. G. Karaçaylı, T. Karim, R. Kehoe, S. Kent, A. Khederlarian, S. Kim, D. Kirkby, T. Kisner, F. Kitaura, J. Kneib, S. E. Koposov, A. Kovács, A. Kremin, A. Krolewski, B. L'Huillier, A. Lambert, C. Lamman, T. W. Lan, M. Landriau, D. Lang, J. U. Lange, J. Lasker, L. Le Guillou, A. Leauthaud, M. E. Levi, T. S. Li, E. Linder, A. Lyons, C. Magneville, M. Manera, C. J. Manser, D. Margala, P. Martini, P. McDonald, G. E. Medina, L. Medina-Varela, A. Meisner, J. Mena-Fernández, J. Meneses-Rizo, M. Mezcua, R. Miquel, P. Montero-Camacho, J. Moon, S. Moore, J. Moustakas, E. Mueller, J. Mundet, A. Muñoz-Gutiérrez, A. D. Myers, S. Nadathur, L. Napolitano, R. Neveux, J. A. Newman, J. Nie, G. Niz, P. Norberg, H. E. Noriega, E. Paillas, N. Palanque-Delabrouille, A. Palmese, P. Zhiwei, D. Parkinson, S. Penmetsa, W. J. Percival, A. Pérez-Fernández, I. Pérez-Ràfols, M. Pieri, C. Poppett, A. Porredon, F. Prada, R. Pucha, A. Raichoor, C. Ramírez-Pérez, S. Ramirez-Solano, M. Rashkovetskyi, C. Ravoux, A. Rocher, C. Rockosi, A. J. Ross, G. Rossi, R. Ruggeri, V. Ruhlmann-Kleider, C. G. Sabiu, K. Said, A. Saintonge, L. Samushia, E. Sanchez, C. Saulder, E. Schaan, E. F. Schlafly, D. Schlegel, D. Scholte, M. Schubnell, H. Seo, A. Shafieloo, R. Sharples, W. Sheu, J. Silber, F. Sinigaglia, M. Siudek, Z. Slepian, A. Smith, D. Sprayberry, L. Stephey, J. Suárez-Pérez, Z. Sun, T. Tan, G. Tarlé, R. Tojeiro, L. A. Ureña-López, R. Vaisakh, D. Valcin, F. Valdes, M. Valluri, M. Vargas-Magaña, A. Variu, L. Verde, M. Walther, B. Wang, M. S. Wang, B. A. Weaver, N. Weaverdyck, R. H. Wechsler, M. White, Y. Xie, J. Yang, C. Yèche, J. Yu,

S. Yuan, H. Zhang, Z. Zhang, C. Zhao, Z. Zheng, R. Zhou, Z. Zhou, H. Zou, S. Zou, and Y. Zu. Validation of the Scientific Program for the Dark Energy Spectroscopic Instrument. *arXiv e-prints*, art. arXiv:2306.06307, June 2023.

[140] Rongpu Zhou, Jeffrey A. Newman, Kyle S. Dawson, Daniel J. Eisenstein, David D. Brooks, Arjun Dey, Biprateep Dey, Yutong Duan, Sarah Eftekharzadeh, Enrique Gaztañaga, Robert Kehoe, Martin Landriau, Michael E. Levi, Timothy C. Licquia, Aaron M. Meisner, John Moustakas, Adam D. Myers, Nathalie Palanque-Delabrouille, Claire Poppett, Francisco Prada, Anand Raichoor, David J. Schlegel, Michael Schubnell, Ryan Staten, Gregory Tarlé, and Christophe Yèche. Preliminary Target Selection for the DESI Luminous Red Galaxy (LRG) Sample. *Research Notes of the American Astronomical Society*, 4(10):181, October 2020. doi: 10.3847/2515-5172/abc0f4.

[141] Rongpu Zhou, Biprateep Dey, Jeffrey A. Newman, Daniel J. Eisenstein, K. Dawson, S. Bailey, A. Berti, J. Guy, Ting-Wen Lan, H. Zou, J. Aguilar, S. Ahlen, Shadab Alam, D. Brooks, A. de la Macorra, A. Dey, G. Dhungana, K. Fanning, A. Font-Ribera, S. Gontcho A. Gontcho, K. Honscheid, Mustapha Ishak, T. Kisner, A. Kovács, A. Kremin, M. Landriau, Michael E. Levi, C. Magneville, Marc Manera, P. Martini, Aaron M. Meisner, R. Miquel, J. Moustakas, Adam D. Myers, Jundan Nie, N. Palanque-Delabrouille, W. J. Percival, C. Poppett, F. Prada, A. Raichoor, A. J. Ross, E. Schlafly, D. Schlegel, M. Schubnell, Gregory Tarlé, B. A. Weaver, R. H. Wechsler, Christophe Yéche, and Zhimin Zhou. Target Selection and Validation of DESI Luminous Red Galaxies. AJ, 165(2):58, February 2023. doi: 10.3847/1538-3881/aca5fb.

[142] Anand Raichoor, Daniel J. Eisenstein, Tanveer Karim, Jeffrey A. Newman, John Moustakas, David D. Brooks, Kyle S. Dawson, Arjun Dey, Yutong Duan, Sarah Eftekharzadeh, Enrique Gaztañaga, Robert Kehoe, Martin Landriau, Dustin Lang, Jae H. Lee, Michael E. Levi, Aaron M. Meisner, Adam D. Myers, Nathalie Palanque-Delabrouille, Claire Poppett, Francisco Prada, Ashley J. Ross, David J. Schlegel, Michael Schubnell, Ryan Staten, Gregory Tarlé, Rita Tojeiro, Christophe Yèche, and

Rongpu Zhou. Preliminary Target Selection for the DESI Emission Line Galaxy (ELG) Sample. *Research Notes of the American Astronomical Society*, 4(10):180, October 2020. doi: 10.3847/2515-5172/abc078.

[143] A. Raichoor, J. Moustakas, Jeffrey A. Newman, T. Karim, S. Ahlen, Shadab Alam, S. Bailey, D. Brooks, K. Dawson, A. de la Macorra, A. de Mattia, A. Dey, Biprateep Dey, G. Dhungana, S. Eftekharzadeh, D. J. Eisenstein, K. Fanning, A. Font-Ribera, J. García-Bellido, E. Gaztañaga, S. Gontcho A Gontcho, J. Guy, K. Honscheid, M. Ishak, R. Kehoe, T. Kisner, Anthony Kremin, Ting-Wen Lan, M. Landriau, L. Le Guillou, Michael E. Levi, C. Magneville, M. Manera, P. Martini, Aaron M. Meisner, Adam D. Myers, Jundan Nie, N. Palanque-Delabrouille, W. J. Percival, C. Poppett, F. Prada, A. J. Ross, V. Ruhlmann-Kleider, C. G. Sabiu, E. F. Schlafly, D. Schlegel, Gregory Tarlé, B. A. Weaver, Christophe Yèche, Rongpu Zhou, Zhimin Zhou, and H. Zou. Target Selection and Validation of DESI Emission Line Galaxies. AJ, 165(3): 126, March 2023. doi: 10.3847/1538-3881/acb213.

[144] Christophe Yèche, Nathalie Palanque-Delabrouille, Charles-Antoine Claveau, David D. Brooks, Edmond Chaussidon, Tamara M. Davis, Kyle S. Dawson, Arjun Dey, Yutong Duan, Sarah Eftekharzadeh, Daniel J. Eisenstein, Enrique Gaztañaga, Robert Kehoe, Martin Landriau, Dustin Lang, Michael E. Levi, Aaron M. Meisner, Adam D. Myers, Jeffrey A. Newman, Claire Poppett, Francisco Prada, Anand Raichoor, David J. Schlegel, Michael Schubnell, Ryan Staten, Gregory Tarlé, and Rongpu Zhou. Preliminary Target Selection for the DESI Quasar (QSO) Sample. *Research Notes of the American Astronomical Society*, 4(10):179, October 2020. doi: 10.3847/2515-5172/abc01a.

[145] Edmond Chaussidon, Christophe Yèche, Nathalie Palanque-Delabrouille, David M. Alexander, Jinyi Yang, Steven Ahlen, Stephen Bailey, David Brooks, Zheng Cai, Solène Chabanier, Tamara M. Davis, Kyle Dawson, Axel de laMacorra, Arjun Dey, Biprateep Dey, Sarah Eftekharzadeh, Daniel J. Eisenstein, Kevin Fanning, Andreu Font-Ribera, Enrique Gaztañaga, Satya Gontcho A Gontcho, Alma X. Gonzalez-

Morales, Julien Guy, Hiram K. Herrera-Alcantar, Klaus Honscheid, Mustapha Ishak, Linhua Jiang, Stephanie Juneau, Robert Kehoe, Theodore Kisner, Andras Kovács, Anthony Kremin, Ting-Wen Lan, Martin Landriau, Laurent Le Guillou, Michael E. Levi, Christophe Magneville, Paul Martini, Aaron M. Meisner, John Moustakas, Andrea Muñoz-Gutiérrez, Adam D. Myers, Jeffrey A. Newman, Jundan Nie, Will J. Percival, Claire Poppett, Francisco Prada, Anand Raichoor, Corentin Ravoux, Ashley J. Ross, Edward Schlafly, David Schlegel, Ting Tan, Gregory Tarlé, Rongpu Zhou, Zhimin Zhou, and Hu Zou. Target Selection and Validation of DESI Quasars. ApJ, 944(1): 107, February 2023. doi: 10.3847/1538-4357/acb3c2.

[146] Omar Ruiz-Macias, Pauline Zarrouk, Shaun Cole, Peder Norberg, Carlton Baugh, David Brooks, Arjun Dey, Yutong Duan, Sarah Eftekharzadeh, Daniel J. Eisenstein, Jaime E. Forero-Romero, Enrique Gaztañaga, ChangHoon Hahn, Robert Kehoe, Martin Landriau, Dustin Lang, Michael E. Levi, John Lucey, Aaron M. Meisner, John Moustakas, Adam D. Myers, Nathalie Palanque-Delabrouille, Claire Poppett, Francisco Prada, Anand Raichoor, David J. Schlegel, Michael Schubnell, Gregory Tarlé, David H. Weinberg, M. J. Wilson, and Christophe Yèche. Preliminary Target Selection for the DESI Bright Galaxy Survey (BGS). *Research Notes of the American Astronomical Society*, 4(10):187, October 2020. doi: 10.3847/2515-5172/abc25a.

[147] ChangHoon Hahn, Michael J. Wilson, Omar Ruiz-Macias, Shaun Cole, David H. Weinberg, John Moustakas, Anthony Kremin, Jeremy L. Tinker, Alex Smith, Risa H. Wechsler, Steven Ahlen, Shadab Alam, Stephen Bailey, David Brooks, Andrew P. Cooper, Tamara M. Davis, Kyle Dawson, Arjun Dey, Biprateep Dey, Sarah Eftekharzadeh, Daniel J. Eisenstein, Kevin Fanning, Jaime E. Forero-Romero, Carlos S. Frenk, Enrique Gaztañaga, Satya Gontcho A Gontcho, Julien Guy, Klaus Honscheid, Mustapha Ishak, Stéphanie Juneau, Robert Kehoe, Theodore Kisner, Ting-Wen Lan, Martin Landriau, Laurent Le Guillou, Michael E. Levi, Christophe Magneville, Paul Martini, Aaron Meisner, Adam D. Myers, Jundan Nie, Peder Norberg, Nathalie Palanque-Delabrouille, Will J. Percival, Claire Poppett, Francisco Prada, Anand Raichoor, Ash-

ley J. Ross, Sasha Safonova, Christoph Saulder, Eddie Schlafly, David Schlegel, David Sierra-Porta, Gregory Tarle, Benjamin A. Weaver, Christophe Yèche, Pauline Zarrouk, Rongpu Zhou, Zhimin Zhou, and Hu Zou. DESI Bright Galaxy Survey: Final Target Selection, Design, and Validation. *arXiv e-prints*, art. arXiv:2208.08512, August 2022. doi: 10.48550/arXiv.2208.08512.

[148] Antoine Rocher, Vanina Ruhlmann-Kleider, Etienne Burtin, Sihan Yuan, Arnaud de Mattia, Ashley J. Ross, Jessica Aguilar, Steven Ahlen, Shadab Alam, Davide Bianchi, David Brooks, Shaun Cole, Kyle Dawson, Axel de la Macorra, Peter Doel, Daniel J. Eisenstein, Kevin Fanning, Jaime E. Forero-Romero, Lehman H. Garrison, Satya Gontcho A Gontcho, Violeta Gonzalez-Perez, Julien Guy, Boryana Hadzhiyska, ChangHoon Hahn, Klaus Honscheid, Theodore Kisner, Martin Landriau, James Lasker, Michael E. Levi, Marc Manera, Aaron Meisner, Ramon Miquel, John Moustakas, Eva-Maria Mueller, Jeffrey A. Newman, Jundan Nie, Will J. Percival, Claire Poppett, Fei Qin, Graziano Rossi, Lado Samushia, Eusebio Sanchez, David Schlegel, Michael Schubnell, Hee-Jong Seo, Gregory Tarlé, Mariana Vargas-Magaña, Benjamin A. Weaver, Jiaxi Yu, Hanyu Zhang, Zheng Zheng, Zhimin Zhou, and Hu Zou. The DESI One-Percent survey: exploring the Halo Occupation Distribution of Emission Line Galaxies with AbacusSummit simulations. *arXiv e-prints*, art. arXiv:2306.06319, June 2023.

[149] F. Prada, J. Ereza, A. Smith, J. Lasker, R. Vaisakh, R. Kehoe, C. A. Dong-Páez, M. Siudek, M. S. Wang, S. Alam, F. Beutler, D. Bianchi, S. Cole, B. Dey, D. Kirkby, P. Norberg, J. Aguilar, S. Ahlen, D. Brooks, T. Claybaugh, K. Dawson, A. de la Macorra, K. Fanning, J. E. Forero-Romero, S. Gontcho A Gontcho, C. Hahn, K. Honscheid, M. Ishak, T. Kisner, M. Landriau, M. Manera, A. Meisner, R. Miquel, J. Moustakas, E. Mueller, J. Nie, W. J. Percival, C. Poppett, M. Rezaie, G. Rossi, E. Sanchez, M. Schubnell, G. Tarlé, M. Vargas-Magaña, B. A. Weaver, S. Yuan, and Z. Zhou. The DESI One-Percent Survey: Modelling the clustering and halo occupation of all four DESI tracers with Uchuu. *arXiv e-prints*, art. arXiv:2306.06315, June 2023.

[150] Tomoaki Ishiyama, Francisco Prada, Anatoly A. Klypin, Manodeep Sinha, R. Benton Metcalf, Eric Jullo, Bruno Altieri, Sofía A. Cora, Darren Croton, Sylvain de la Torre, David E. Millán-Calero, Taira Oogi, José Ruedas, and Cristian A. Vega-Martínez. The Uchuu simulations: Data Release 1 and dark matter halo concentrations. MNRAS, 506(3):4210–4231, September 2021. doi: 10.1093/mnras/stab1755.

[151] Jiaxi Yu, Cheng Zhao, Violeta Gonzalez-Perez, Chia-Hsun Chuang, Allyson Brodzeller, Arnaud de Mattia, Jean-Paul Kneib, Alex Krolewski, Antoine Rocher, Ashley Ross, Yunchong Wang, Sihan Yuan, Hanyu Zhang, Rongpu Zhou, Jessica Nicole Aguilar, Steven Ahlen, David Brooks, Kyle Dawson, Alex de la Macorra, Peter Doel, Kevin Fanning, Andreu Font-Ribera, Jaime Forero-Romero, Satya Gontcho A Gontcho, Klaus Honscheid, Robert Kehoe, Theodore Kisner, Anthony Kremin, Martin Landriau, Marc Manera, Paul Martini, Aaron Meisner, Ramon Miquel, John Moustakas, Jundan Nie, Will Percival, Claire Poppett, Anand Raichoor, Graziano Rossi, Hee-Jong Seo, Gregory Tarlé, Zhimin Zhou, and Hu Zou. The DESI One-Percent Survey: Exploring A Generalized SHAM for Multiple Tracers with the UNIT Simulation. *arXiv e-prints*, art. arXiv:2306.06313, June 2023.

[152] Chia-Hsun Chuang, Gustavo Yepes, Francisco-Shu Kitaura, Marcos Pellejero-Ibanez, Sergio Rodríguez-Torres, Yu Feng, Robert Benton Metcalf, Risa H. Wechsler, Cheng Zhao, Chun-Hao To, Shadab Alam, Arka Banerjee, Joseph DeRose, Carlo Giocoli, Alexander Knebe, and Guillermo Reyes. UNIT project: Universe N-body simulations for the Investigation of Theoretical models from galaxy surveys. MNRAS, 487(1): 48–59, July 2019. doi: 10.1093/mnras/stz1233.

[153] Hongyu Gao, Y. P. Jing, Shanquan Gui, Kun Xu, Yun Zheng, Donghai Zhao, Jessica Nicole Aguilar, Steven Ahlen, David Brooks, Todd Claybaugh, Kyle Dawson, Axel de la Macorra, Peter Doel, Kevin Fanning, Jaime E. Forero-Romero, Satya Gontcho A Gontcho, Julien Guy, Klaus Honscheid, Robert Kehoe, Martin Landriau, Marc Manera, Aaron Meisner, Ramon Miquel, John Moustakas, Jeffrey A. Newman, Jundan

Nie, Will Percival, Graziano Rossi, Michael Schubnell, Hee-Jong Seo, Gregory Tarlé, Benjamin Alan Weaver, Jiaxi Yu, and Zhimin Zhou. The DESI One-Percent survey: constructing galaxy-halo connections for ELGs and LRGs using auto and cross correlations. *arXiv e-prints*, art. arXiv:2306.06317, June 2023.

[154] DESI Collaboration, A. G. Adame, J. Aguilar, S. Ahlen, S. Alam, G. Aldering, D. M. Alexander, R. Alfarsy, C. Allende Prieto, M. Alvarez, O. Alves, A. Anand, F. Andrade-Oliveira, E. Armengaud, J. Asorey, S. Avila, A. Aviles, S. Bailey, A. Balaguera-Antolínez, O. Ballester, C. Baltay, A. Bault, J. Bautista, J. Behera, S. F. Beltran, S. BenZvi, L. Beraldo e Silva, J. R. Bermejo-Climent, A. Berti, R. Besuner, F. Beutler, D. Bianchi, C. Blake, R. Blum, A. S. Bolton, S. Brieden, A. Brodzeller, D. Brooks, Z. Brown, E. Buckley-Geer, E. Burtin, L. Cabayol-Garcia, Z. Cai, R. Canning, L. Cardiel-Sas, A. Carnero Rosell, F. J. Castander, J. L. Cervantes-Cota, S. Chabanier, E. Chaussidon, J. Chaves-Montero, S. Chen, C. Chuang, T. Claybaugh, S. Cole, A. P. Cooper, A. Cuceu, T. M. Davis, K. Dawson, R. de Belsunce, R. de la Cruz, A. de la Macorra, A. de Mattia, R. Demina, U. Demirbozan, J. DeRose, A. Dey, B. Dey, G. Dhungana, J. Ding, Z. Ding, P. Doel, R. Doshi, K. Douglass, A. Edge, S. Eftekharzadeh, D. J. Eisenstein, A. Elliott, S. Escoffier, P. Fagrelius, X. Fan, K. Fanning, V. A. Fawcett, S. Ferraro, J. Ereza, B. Flaugher, A. Font-Ribera, D. Forero-Sánchez, J. E. Forero-Romero, C. S. Frenk, B. T. Gänsicke, L. Á. García, J. García-Bellido, C. Garcia-Quintero, L. H. Garrison, H. Gil-Marín, J. Golden-Marx, S. Gontcho A Gontcho, A. X. Gonzalez-Morales, V. Gonzalez-Perez, C. Gordon, O. Graur, D. Green, D. Gruen, J. Guy, B. Hadzhiyska, C. Hahn, J. J. Han, M. M. S Hanif, H. K. Herrera-Alcantar, K. Honscheid, J. Hou, C. Howlett, D. Huterer, V. Iršič, M. Ishak, A. Jacques, A. Jana, L. Jiang, J. Jimenez, Y. P. Jing, S. Joudaki, E. Jullo, S. Juneau, N. Kizhuprakkat, N. G. Karaçaylı, T. Karim, R. Kehoe, S. Kent, A. Khederlarian, S. Kim, D. Kirkby, T. Kisner, F. Kitaura, J. Kneib, S. E. Koposov, A. Kovács, A. Kremin, A. Krolewski, B. L'Huillier, A. Lambert, C. Lamman, T. W. Lan, M. Landriau, D. Lang, J. U. Lange, J. Lasker, L. Le Guillou, A. Leauthaud, M. E. Levi, T. S. Li, E. Linder,

A. Lyons, C. Magneville, M. Manera, C. J. Manser, D. Margala, P. Martini, P. McDonald, G. E. Medina, L. Medina-Varela, A. Meisner, J. Mena-Fernández, J. Meneses-Rizo, M. Mezcua, R. Miquel, P. Montero-Camacho, J. Moon, S. Moore, J. Moustakas, E. Mueller, J. Mundet, A. Muñoz-Gutiérrez, A. D. Myers, S. Nadathur, L. Napolitano, R. Neveux, J. A. Newman, J. Nie, R. Nikutta, G. Niz, P. Norberg, H. E. Noriega, E. Paillas, N. Palanque-Delabrouille, A. Palmese, P. Zhiwei, D. Parkinson, S. Penmetsa, W. J. Percival, A. Pérez-Fernández, I. Pérez-Ràfols, M. Pieri, C. Poppett, A. Porredon, S. Pothier, F. Prada, R. Pucha, A. Raichoor, C. Ramírez-Pérez, S. Ramirez-Solano, M. Rashkovetskyi, C. Ravoux, A. Rocher, C. Rockosi, A. J. Ross, G. Rossi, R. Ruggeri, V. Ruhlmann-Kleider, C. G. Sabiu, K. Said, A. Saintonge, L. Samushia, E. Sanchez, C. Saulder, E. Schaan, E. F. Schlafly, D. Schlegel, D. Scholte, M. Schubnell, H. Seo, A. Shafieloo, R. Sharples, W. Sheu, J. Silber, F. Sinigaglia, M. Siudek, Z. Slepian, A. Smith, D. Sprayberry, L. Stephey, J. Suárez-Pérez, Z. Sun, T. Tan, G. Tarlé, R. Tojeiro, L. A. Ureña-López, R. Vaisakh, D. Valcin, F. Valdes, M. Valluri, M. Vargas-Magaña, A. Variu, L. Verde, M. Walther, B. Wang, M. S. Wang, B. A. Weaver, N. Weaverdyck, R. H. Wechsler, M. White, Y. Xie, J. Yang, C. Yèche, J. Yu, S. Yuan, H. Zhang, Z. Zhang, C. Zhao, Z. Zheng, R. Zhou, Z. Zhou, H. Zou, S. Zou, and Y. Zu. The Early Data Release of the Dark Energy Spectroscopic Instrument. *arXiv e-prints*, art. arXiv:2306.06308, June 2023.

[155] Joseph Harry Silber, Parker Fagrelius, Kevin Fanning, Michael Schubnell, Jessica Nicole Aguilar, Steven Ahlen, Jon Ameel, Otger Ballester, Charles Baltay, Chris Bebek, Dominic Benton Beard, Robert Besuner, Laia Cardiel-Sas, Ricard Casas, Francisco Javier Castander, Todd Claybaugh, Carl Dobson, Yutong Duan, Patrick Dunlop, Jerry Edelstein, William T. Emmet, Ann Elliott, Matthew Evatt, Irena Gershkovich, Julien Guy, Stu Harris, Henry Heetderks, Ian Heetderks, Klaus Honscheid, Jose Maria Illa, Patrick Jelinsky, Sharon R. Jelinsky, Jorge Jimenez, Armin Karcher, Stephen Kent, David Kirkby, Jean-Paul Kneib, Andrew Lambert, Mike Lampton, Daniela Leitner, Michael Levi, Jeremy McCauley, Aaron Meisner, Timothy N. Miller, Ra-

mon Miquel, Juliá Mundet, Claire Poppett, David Rabinowitz, Kevin Reil, David Roman, David Schlegel, Santiago Serrano, William Van Shourt, David Sprayberry, Gregory Tarlé, Suk Sien Tie, Curtis Weaverdyck, Kai Zhang, Marco Azzaro, Stephen Bailey, Santiago Becerril, Tami Blackwell, Mohamed Bouri, David Brooks, Elizabeth Buckley-Geer, Jose Peñate Castro, Mark Derwent, Arjun Dey, Govinda Dhungana, Peter Doel, Daniel J. Eisenstein, Nasib Fahim, Juan Garcia-Bellido, Enrique Gaztañaga, Satya Gontcho A Gontcho, Gaston Gutierrez, Philipp Hörler, Robert Kehoe, Theodore Kisner, Anthony Kremin, Luzius Kronig, Martin Landriau, Laurent Le Guillou, Paul Martini, John Moustakas, Nathalie Palanque-Delabrouille, Xiyan Peng, Will Percival, Francisco Prada, Carlos Allende Prieto, Guillermo Gonzalez de Rivera, Eusebio Sanchez, Justo Sanchez, Ray Sharples, Marcelle Soares-Santos, Edward Schlafly, Benjamin Alan Weaver, Zhimin Zhou, Yaling Zhu, Hu Zou, and DESI Collaboration. The Robotic Multiobject Focal Plane System of the Dark Energy Spectroscopic Instrument (DESI). AJ, 165(1):9, January 2023. doi: 10.3847/1538-3881/ac9ab1.

[156] E. F. Schlafly, D. Kirkby, D. J. Schlegel, A. D. Myers, A. Raichoor, K. Dawson, J. Aguilar, C. Allende Prieto, S. Bailey, S. BenZvi, J. Bermejo-Climent, D. Brooks, A. de la Macorra, Arjun Dey, P. Doel, K. Fanning, A. Font-Ribera, J. E. Forero-Romero, J. García-Bellido, S. Gontcho A Gontcho, J. Guy, C. Hahn, K. Honscheid, M. Ishak, S. Juneau, R. Kehoe, T. Kisner, A. Kremin, M. Landriau, D. A. Lang, J. Lasker, M. E. Levi, C. Magneville, C. J. Manser, P. Martini, A. M. Meisner, R. Miquel, J. Moustakas, J. A. Newman, Jundan Nie, N. Palanque-Delabrouille, W. J. Percival, C. Poppett, C. Rockosi, A. J. Ross, G. Rossi, G. Tarlé, B. A. Weaver, C. Yèche, and R. Zhou. Survey Operations for the Dark Energy Spectroscopic Instrument. *arXiv e-prints*, art. arXiv:2306.06309, June 2023.

[157] J. Guy, S. Bailey, A. Kremin, Shadab Alam, D. M. Alexander, C. Allende Prieto, S. BenZvi, A. S. Bolton, D. Brooks, E. Chaussidon, A. P. Cooper, K. Dawson, A. de la Macorra, A. Dey, Biprateep Dey, G. Dhungana, D. J. Eisenstein, A. Font-Ribera, J. E. Forero-Romero, E. Gaztañaga, S. Gontcho A Gontcho, D. Green, K. Honscheid,

M. Ishak, R. Kehoe, D. Kirkby, T. Kisner, Sergey E. Koposov, Ting-Wen Lan, M. Landriau, L. Le Guillou, Michael E. Levi, C. Magneville, Christopher J. Manser, P. Martini, Aaron M. Meisner, R. Miquel, J. Moustakas, Adam D. Myers, Jeffrey A. Newman, Jundan Nie, N. Palanque-Delabrouille, W. J. Percival, C. Poppett, F. Prada, A. Raichoor, C. Ravoux, A. J. Ross, E. F. Schlafly, D. Schlegel, M. Schubnell, Ray M. Sharples, Gregory Tarlé, B. A. Weaver, Christophe Yéche, Rongpu Zhou, Zhimin Zhou, and H. Zou. The Spectroscopic Data Processing Pipeline for the Dark Energy Spectroscopic Instrument. AJ, 165(4):144, April 2023. doi: 10.3847/1538-3881/acb212.

[158] Adam D. Myers, John Moustakas, Stephen Bailey, Benjamin A. Weaver, Andrew P. Cooper, Jaime E. Forero-Romero, Bela Abolfathi, David M. Alexander, David Brooks, Edmond Chaussidon, Chia-Hsun Chuang, Kyle Dawson, Arjun Dey, Biprateep Dey, Govinda Dhungana, Peter Doel, Kevin Fanning, Enrique Gaztañaga, Satya Gontcho A Gontcho, Alma X. Gonzalez-Morales, ChangHoon Hahn, Hiram K. Herrera-Alcantar, Klaus Honscheid, Mustapha Ishak, Tanveer Karim, David Kirkby, Theodore Kisner, Sergey E. Koposov, Anthony Kremin, Ting-Wen Lan, Martin Landriau, Dustin Lang, Michael E. Levi, Christophe Magneville, Lucas Napolitano, Paul Martini, Aaron Meisner, Jeffrey A. Newman, Nathalie Palanque-Delabrouille, Will Percival, Claire Poppett, Francisco Prada, Anand Raichoor, Ashley J. Ross, Edward F. Schlafly, David Schlegel, Michael Schubnell, Ting Tan, Gregory Tarle, Michael J. Wilson, Christophe Yèche, Rongpu Zhou, Zhimin Zhou, and Hu Zou. The Target-selection Pipeline for the Dark Energy Spectroscopic Instrument. AJ, 165(2):50, February 2023. doi: 10.3847/1538-3881/aca5f9.

[159] J. Lasker. Dark Energy Spectroscopic Instrument Survey Alternative Targeting Assignment Histories. *XXX*, x in prep.

[160] Davide Bianchi and Will J. Percival. Unbiased clustering estimation in the presence of missing observations. MNRAS, 472(1):1106–1118, November 2017. doi: 10.1093/mnras/stx2053.

[161] Will J. Percival and Davide Bianchi. Using angular pair upweighting to improve 3D clustering measurements. MNRAS, 472(1):L40–L44, November 2017. doi: 10.1093/mnrasl/slx135.

[162] Faizan G. Mohammad, Will J. Percival, Hee-Jong Seo, Michael J. Chapman, D. Bianchi, Ashley J. Ross, Cheng Zhao, Dustin Lang, Julian Bautista, Jonathan Brinkmann, Joel R. Brownstein, Etienne Burtin, Chia-Hsun Chuang, Kyle S. Dawson, Sylvain de la Torre, Arnaud de Mattia, Sarah Eftekharzadeh, Sebastien Fromenteau, Héctor Gil-Marín, Jiamin Hou, Eva-Maria Mueller, Richard Neveux, Romain Paviot, Anand Raichoor, Graziano Rossi, Donald P. Schneider, Amélie Tamone, Jeremy L. Tinker, Rita Tojeiro, Mariana Vargas Magaña, and Gong-Bo Zhao. The completed SDSS-IV extended baryon oscillation spectroscopic survey: pairwise-inverse probability and angular correction for fibre collisions in clustering measurements. MNRAS, 498(1):128–143, October 2020. doi: 10.1093/mnras/staa2344.

[163] Hume A. Feldman, Nick Kaiser, and John A. Peacock. Power-Spectrum Analysis of Three-dimensional Redshift Surveys. ApJ, 426:23, May 1994. doi: 10.1086/174036.

[164] Kyle S. Dawson, David J. Schlegel, Christopher P. Ahn, Scott F. Anderson, Éric Aubourg, Stephen Bailey, Robert H. Barkhouser, Julian E. Bautista, Alessandra Beifiori, Andreas A. Berlind, Vaishali Bhardwaj, Dmitry Bizyaev, Cullen H. Blake, Michael R. Blanton, Michael Blomqvist, Adam S. Bolton, Arnaud Borde, Jo Bovy, W. N. Brandt, Howard Brewington, Jon Brinkmann, Peter J. Brown, Joel R. Brownstein, Kevin Bundy, N. G. Busca, William Carithers, Aurelio R. Carnero, Michael A. Carr, Yanmei Chen, Johan Comparat, Natalia Connolly, Frances Cope, Rupert A. C. Croft, Antonio J. Cuesta, Luiz N. da Costa, James R. A. Davenport, Timothée Delubac, Roland de Putter, Saurav Dhital, Anne Ealet, Garrett L. Ebelke, Daniel J. Eisenstein, S. Escoffier, Xiaohui Fan, N. Filiz Ak, Hayley Finley, Andreu Font-Ribera, R. Génova-Santos, James E. Gunn, Hong Guo, Daryl Haggard, Patrick B. Hall, Jean-Christophe Hamilton, Ben Harris, David W. Harris, Shirley Ho, David W. Hogg, Diana

Holder, Klaus Honscheid, Joe Huehnerhoff, Beatrice Jordan, Wendell P. Jordan, Guinevere Kauffmann, Eyal A. Kazin, David Kirkby, Mark A. Klaene, Jean-Paul Kneib, Jean-Marc Le Goff, Khee-Gan Lee, Daniel C. Long, Craig P. Loomis, Britt Lundgren, Robert H. Lupton, Marcio A. G. Maia, Martin Makler, Elena Malanushenko, Viktor Malanushenko, Rachel Mandelbaum, Marc Manera, Claudia Maraston, Daniel Margala, Karen L. Masters, Cameron K. McBride, Patrick McDonald, Ian D. McGreer, Richard G. McMahon, Olga Mena, Jordi Miralda-Escudé, Antonio D. Montero-Dorta, Francesco Montesano, Demitri Muna, Adam D. Myers, Tracy Naugle, Robert C. Nichol, Pasquier Noterdaeme, Sebastián E. Nuza, Matthew D. Olmstead, Audrey Oravetz, Daniel J. Oravetz, Russell Owen, Nikhil Padmanabhan, Nathalie Palanque-Delabrouille, Kaike Pan, John K. Parejko, Isabelle Pâris, Will J. Percival, Ismael Pérez-Fournon, Ignasi Pérez-Ràfols, Patrick Petitjean, Robert Pfaffenberger, Janine Pforr, Matthew M. Pieri, Francisco Prada, Adrian M. Price-Whelan, M. Jordan Raddick, Rafael Rebolo, James Rich, Gordon T. Richards, Constance M. Rockosi, Natalie A. Roe, Ashley J. Ross, Nicholas P. Ross, Graziano Rossi, J. A. Rubiño-Martin, Lado Samushia, Ariel G. Sánchez, Conor Sayres, Sarah J. Schmidt, Donald P. Schneider, C. G. Scóccola, Hee-Jong Seo, Alaina Shelden, Erin Sheldon, Yue Shen, Yiping Shu, Anže Slosar, Stephen A. Smee, Stephanie A. Snedden, Fritz Stauffer, Oliver Steele, Michael A. Strauss, Alina Streblyanska, Nao Suzuki, Molly E. C. Swanson, Tomer Tal, Masayuki Tanaka, Daniel Thomas, Jeremy L. Tinker, Rita Tojeiro, Christy A. Tremonti, M. Vargas Magaña, Licia Verde, Matteo Viel, David A. Wake, Mike Watson, Benjamin A. Weaver, David H. Weinberg, Benjamin J. Weiner, Andrew A. West, Martin White, W. M. Wood-Vasey, Christophe Yeche, Idit Zehavi, Gong-Bo Zhao, and Zheng Zheng. The Baryon Oscillation Spectroscopic Survey of SDSS-III. AJ, 145(1): 10, January 2013. doi: 10.1088/0004-6256/145/1/10.

[165] Kyle S. Dawson, Jean-Paul Kneib, Will J. Percival, Shadab Alam, Franco D. Albareti, Scott F. Anderson, Eric Armengaud, Éric Aubourg, Stephen Bailey, Julian E. Bautista, Andreas A. Berlind, Matthew A. Bershady, Florian Beutler, Dmitry Bizyaev,

Michael R. Blanton, Michael Blomqvist, Adam S. Bolton, Jo Bovy, W. N. Brandt, Jon Brinkmann, Joel R. Brownstein, Etienne Burtin, N. G. Busca, Zheng Cai, Chia-Hsun Chuang, Nicolas Clerc, Johan Comparat, Frances Cope, Rupert A. C. Croft, Irene Cruz-Gonzalez, Luiz N. da Costa, Marie-Claude Cousinou, Jeremy Darling, Axel de la Macorra, Sylvain de la Torre, Timothée Delubac, Hélion du Mas des Bourboux, Tom Dwelly, Anne Ealet, Daniel J. Eisenstein, Michael Eracleous, S. Escoffier, Xiaohui Fan, Alexis Finoguenov, Andreu Font-Ribera, Peter Frinchaboy, Patrick Gaulme, Antonis Georgakakis, Paul Green, Hong Guo, Julien Guy, Shirley Ho, Diana Holder, Joe Huehnerhoff, Timothy Hutchinson, Yipeng Jing, Eric Jullo, Vikrant Kamble, Karen Kinemuchi, David Kirkby, Francisco-Shu Kitaura, Mark A. Klaene, Russ R. Laher, Dustin Lang, Pierre Laurent, Jean-Marc Le Goff, Cheng Li, Yu Liang, Marcos Lima, Qiufan Lin, Weipeng Lin, Yen-Ting Lin, Daniel C. Long, Britt Lundgren, Nicholas MacDonald, Marcio Antonio Geimba Maia, Elena Malanushenko, Viktor Malanushenko, Vivek Mariappan, Cameron K. McBride, Ian D. McGreer, Brice Ménard, Andrea Merloni, Andres Meza, Antonio D. Montero-Dorta, Demitri Muna, Adam D. Myers, Kirpal Nandra, Tracy Naugle, Jeffrey A. Newman, Pasquier Noterdaeme, Peter Nugent, Ricardo Ogando, Matthew D. Olmstead, Audrey Oravetz, Daniel J. Oravetz, Nikhil Padmanabhan, Nathalie Palanque-Delabrouille, Kaike Pan, John K. Parejko, Isabelle Pâris, John A. Peacock, Patrick Petitjean, Matthew M. Pieri, Alice Pisani, Francisco Prada, Abhishek Prakash, Anand Raichoor, Beth Reid, James Rich, Jethro Ridl, Sergio Rodriguez-Torres, Aurelio Carnero Rosell, Ashley J. Ross, Graziano Rossi, John Ruan, Mara Salvato, Conor Sayres, Donald P. Schneider, David J. Schlegel, Uros Seljak, Hee-Jong Seo, Branimir Sesar, Sarah Shandera, Yiping Shu, Anže Slosar, Flavia Sobreira, Alina Streblyanska, Nao Suzuki, Donna Taylor, Charling Tao, Jeremy L. Tinker, Rita Tojeiro, Mariana Vargas-Magaña, Yuting Wang, Benjamin A. Weaver, David H. Weinberg, Martin White, W. M. Wood-Vasey, Christophe Yeche, Zhongxu Zhai, Cheng Zhao, Gong-bo Zhao, Zheng Zheng, Guangtun Ben Zhu, and Hu Zou. The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data. AJ, 151(2):44, February 2016. doi:

10.3847/0004-6256/151/2/44.

[166] Faizan G. Mohammad and Will J. Percival. Creating jackknife and bootstrap estimates of the covariance matrix for the two-point correlation function. MNRAS, 514(1):1289–1301, July 2022. doi: 10.1093/mnras/stac1458.

[167] Nina A. Maksimova, Lehman H. Garrison, Daniel J. Eisenstein, Boryana Hadzhiyska, Sownak Bose, and Thomas P. Satterthwaite. ABACUSSUMMIT: a massive set of high-accuracy, high-resolution N-body simulations. MNRAS, 508(3):4017–4037, December 2021. doi: 10.1093/mnras/stab2484.

[168] Boryana Hadzhiyska, Lehman H. Garrison, Daniel Eisenstein, and Sownak Bose. The halo light-cone catalogues of ABACUSSUMMIT. MNRAS, 509(2):2194–2208, January 2022. doi: 10.1093/mnras/stab3066.

[169] Risa H. Wechsler, James S. Bullock, Joel R. Primack, Andrey V. Kravtsov, and Avishai Dekel. Concentrations of Dark Halos from Their Assembly Histories. ApJ, 568(1):52–70, March 2002. doi: 10.1086/338765.

[170] Liang Gao and Simon D. M. White. Assembly bias in the clustering of dark matter haloes. MNRAS, 377(1):L5–L9, April 2007. doi: 10.1111/j.1745-3933.2007.00292.x.

[171] Yen-Ting Lin, Rachel Mandelbaum, Yun-Hsin Huang, Hung-Jin Huang, Neal Dalal, Benedikt Diemer, Hung-Yu Jian, and Andrey Kravtsov. On Detecting Halo Assembly Bias with Galaxy Populations. ApJ, 819(2):119, March 2016. doi: 10.3847/0004-637X/819/2/119.

[172] Xiaoju Xu, Idit Zehavi, and Sergio Contreras. Dissecting and modelling galaxy assembly bias. MNRAS, 502(3):3242–3263, April 2021. doi: 10.1093/mnras/stab100.

[173] Xiaoju Xu, Saurabh Kumar, Idit Zehavi, and Sergio Contreras. Predicting halo occupation and galaxy assembly bias with machine learning. MNRAS, 507(4):4879–4899, November 2021. doi: 10.1093/mnras/stab2464.

[174] Ana Maria Delgado, Digvijay Wadekar, Boryana Hadzhiyska, Sownak Bose, Lars Hernquist, and Shirley Ho. Modelling the galaxy-halo connection with machine learning. MNRAS, 515(2):2733–2746, September 2022. doi: 10.1093/mnras/stac1951.

[175] Andrés N. Salcedo, Ying Zu, Youcai Zhang, Huiyuan Wang, Xiaohu Yang, Yiheng Wu, Yipeng Jing, Houjun Mo, and David H. Weinberg. Elucidating galaxy assembly bias in SDSS. *Science China Physics, Mechanics, and Astronomy*, 65(10):109811, October 2022. doi: 10.1007/s11433-022-1955-7.

[176] Sihan Yuan, Boryana Hadzhiyska, Sownak Bose, and Daniel J. Eisenstein. Illustrating galaxy-halo connection in the DESI era with ILLUSTRISTNG. MNRAS, 512(4):5793–5811, June 2022. doi: 10.1093/mnras/stac830.

[177] Esteban Jiménez, Sergio Contreras, Nelson Padilla, Idit Zehavi, Carlton M. Baugh, and Violeta Gonzalez-Perez. Extensions to the halo occupation distribution model for more accurate clustering predictions. MNRAS, 490(3):3532–3544, December 2019. doi: 10.1093/mnras/stz2790.

[178] J. Hartlap, P. Simon, and P. Schneider. Why your model parameter confidences might be too optimistic. Unbiased estimation of the inverse covariance matrix. A&A, 464(1): 399–404, March 2007. doi: 10.1051/0004-6361:20066170.

[179] J. Speagle and K. Barbary. dynesty: Dynamic Nested Sampling package. Astrophysics Source Code Library, September 2018.

[180] Joshua S. Speagle. DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. MNRAS, 493(3):3132–3158, April 2020. doi: 10.1093/mnras/staa278.

[181] Julien Lesgourgues. The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview. *arXiv e-prints*, art. arXiv:1104.2932, April 2011. doi: 10.48550/arXiv.1104.2932.

[182] Annalisa Pillepich, Dylan Nelson, Lars Hernquist, Volker Springel, Rüdiger Pakmor, Paul Torrey, Rainer Weinberger, Shy Genel, Jill P. Naiman, Federico Marinacci, and Mark Vogelsberger. First results from the IllustrisTNG simulations: the stellar mass content of groups and clusters of galaxies. MNRAS, 475(1):648–675, March 2018. doi: 10.1093/mnras/stx3112.

[183] Volker Springel, Rüdiger Pakmor, Annalisa Pillepich, Rainer Weinberger, Dylan Nelson, Lars Hernquist, Mark Vogelsberger, Shy Genel, Paul Torrey, Federico Marinacci, and Jill Naiman. First results from the IllustrisTNG simulations: matter and galaxy clustering. MNRAS, 475(1):676–698, March 2018. doi: 10.1093/mnras/stx3304.

[184] Dylan Nelson, Annalisa Pillepich, Volker Springel, Rainer Weinberger, Lars Hernquist, Rüdiger Pakmor, Shy Genel, Paul Torrey, Mark Vogelsberger, Guinevere Kauffmann, Federico Marinacci, and Jill Naiman. First results from the IllustrisTNG simulations: the galaxy colour bimodality. MNRAS, 475(1):624–647, March 2018. doi: 10.1093/mnras/stx3040.

[185] DESI collaboration et al. *XXX*, in prep.

[186] David J. Setton, Biprateep Dey, Gourav Khullar, Rachel Bezanson, Jeffrey A. Newman, Jessica N. Aguilar, Steven Ahlen, Brett H. Andrews, David Brooks, Axel de la Macorra, Arjun Dey, Sarah Eftekharzadeh, Andreu Font-Ribera, Satya Gontcho A Gontcho, Anthony Kremin, Stephanie Juneau, Martin Landriau, Aaron Meisner, Ramon Miquel, John Moustakas, Alan Pearl, Francisco Prada, Gregory Tarlé, Małgorzata Siudek, Benjamin Alan Weaver, Zhimin Zhou, and Hu Zou. DESI Survey Validation Spectra Reveal an Increasing Fraction of Recently Quenched Galaxies at z 1. ApJ, 947 (2):L31, April 2023. doi: 10.3847/2041-8213/acc9b5.

[187] DESI collaboration et al. *XXX*, in prep.

[188] Pauline Zarrouk, Etienne Burtin, Héctor Gil-Marín, Ashley J. Ross, Rita Tojeiro, Isabelle Pâris, Kyle S. Dawson, Adam D. Myers, Will J. Percival, Chia-Hsun Chuang,

Gong-Bo Zhao, Julian Bautista, Johan Comparat, Violeta González-Pérez, Salman Habib, Katrin Heitmann, Jiamin Hou, Pierre Laurent, Jean-Marc Le Goff, Francisco Prada, Sergio A. Rodríguez-Torres, Graziano Rossi, Rossana Ruggeri, Ariel G. Sánchez, Donald P. Schneider, Jeremy L. Tinker, Yuting Wang, Christophe Yèche, Falk Baumgarten, Joel R. Brownstein, Sylvain de la Torre, Hélion du Mas des Bourboux, Jean-Paul Kneib, Vivek Mariappan, Nathalie Palanque-Delabrouille, John Peacock, Patrick Petitjean, Hee-Jong Seo, and Cheng Zhao. The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: measurement of the growth rate of structure from the anisotropic correlation function between redshift 0.8 and 2.2. MNRAS, 477(2):1639–1663, June 2018. doi: 10.1093/mnras/sty506.

[189] Gordon T. Richards, Nicholas E. Kruczek, S. C. Gallagher, Patrick B. Hall, Paul C. Hewett, Karen M. Leighly, Rajesh P. Deo, Rachael M. Kratzer, and Yue Shen. Unification of Luminous Type 1 Quasars through C IV Emission. AJ, 141(5):167, May 2011. doi: 10.1088/0004-6256/141/5/167.

[190] Gordon T. Richards, Daniel E. Vanden Berk, Timothy A. Reichard, Patrick B. Hall, Donald P. Schneider, Mark SubbaRao, Anirudda R. Thakar, and Donald G. York. Broad Emission-Line Shifts in Quasars: An Orientation Measure for Radio-Quiet Quasars? AJ, 124(1):1–17, July 2002. doi: 10.1086/341167.

[191] Sergio A. Rodríguez-Torres, Johan Comparat, Francisco Prada, Gustavo Yepes, Etienne Burtin, Pauline Zarrouk, Pierre Laurent, ChangHoon Hahn, Peter Behroozi, Anatoly Klypin, Ashley Ross, Rita Tojeiro, and Gong-Bo Zhao. Clustering of quasars in the first year of the SDSS-IV eBOSS survey: interpretation and halo occupation distribution. MNRAS, 468(1):728–740, June 2017. doi: 10.1093/mnras/stx454.

[192] Pierre Laurent, Sarah Eftekharzadeh, Jean-Marc Le Goff, Adam Myers, Etienne Burtin, Martin White, Ashley J. Ross, Jeremy Tinker, Rita Tojeiro, Julian Bautista, Jonathan Brinkmann, Johan Comparat, Kyle Dawson, Hélion du Mas des Bourboux, Jean-Paul Kneib, Ian D. McGreer, Nathalie Palanque-Delabrouille, Will J. Percival,

Francisco Prada, Graziano Rossi, Donald P. Schneider, David Weinberg, Christophe Yèche, Pauline Zarrouk, and Gong-Bo Zhao. Clustering of quasars in SDSS-IV eBOSS: study of potential systematics and bias determination. J. Cosmology Astropart. Phys., 2017(7):017, July 2017. doi: 10.1088/1475-7516/2017/07/017.

[193] Yue Shen, Cameron K. McBride, Martin White, Zheng Zheng, Adam D. Myers, Hong Guo, Jessica A. Kirkpatrick, Nikhil Padmanabhan, John K. Parejko, Nicholas P. Ross, David J. Schlegel, Donald P. Schneider, Alina Streblyanska, Molly E. C. Swanson, Idit Zehavi, Kaike Pan, Dmitry Bizyaev, Howard Brewington, Garrett Ebelke, Viktor Malanushenko, Elena Malanushenko, Daniel Oravetz, Audrey Simmons, and Stephanie Snedden. Cross-correlation of SDSS DR7 Quasars and DR10 BOSS Galaxies: The Weak Luminosity Dependence of Quasar Clustering at z ~0.5. ApJ, 778(2):98, December 2013. doi: 10.1088/0004-637X/778/2/98.

[194] Chuang et al. The DESI Mock Challenge: Covariance Matrix Comparison Project. *XXX*, in prep.

[195] Sihan Yuan, Boryana Hadzhiyska, and Tom Abel. Full forward model of galaxy clustering statistics with ABACUSSUMMIT light cones. MNRAS, 520(4):6283–6298, April 2023. doi: 10.1093/mnras/stad550.

[196] W. L. W. Sargent and E. L. Turner. A statistical method for determining the cosmological density parameter from the redshifts of a complete sample of galaxies. ApJ, 212:L3–L7, February 1977. doi: 10.1086/182362.

[197] Nick Kaiser. Clustering in real space and in redshift space. MNRAS, 227:1–21, July 1987. doi: 10.1093/mnras/227.1.1.

[198] Atsushi Taruya, Takahiro Nishimichi, and Shun Saito. Baryon acoustic oscillations in 2D: Modeling redshift-space power spectrum from perturbation theory. Phys. Rev. D, 82(6):063522, September 2010. doi: 10.1103/PhysRevD.82.063522.

[199] Uroš Seljak and Patrick McDonald. Distribution function approach to redshift space distortions. J. Cosmology Astropart. Phys., 2011(11):039, November 2011. doi: 10.1088/1475-7516/2011/11/039.

[200] Beth A. Reid and Martin White. Towards an accurate model of the redshift-space clustering of haloes in the quasi-linear regime. MNRAS, 417(3):1913–1927, November 2011. doi: 10.1111/j.1365-2966.2011.19379.x.

[201] Teppei Okumura, Uroš Seljak, and Vincent Desjacques. Distribution function approach to redshift space distortions. Part III: halos and galaxies. J. Cosmology Astropart. Phys., 2012(11):014, November 2012. doi: 10.1088/1475-7516/2012/11/014.

[202] Jordan Carlson, Beth Reid, and Martin White. Convolution Lagrangian perturbation theory for biased tracers. MNRAS, 429(2):1674–1685, February 2013. doi: 10.1093/mnras/sts457.

[203] Takahiko Matsubara. Integrated perturbation theory and one-loop power spectra of biased tracers. Phys. Rev. D, 90(4):043537, August 2014. doi: 10.1103/PhysRevD.90.043537.

[204] Lile Wang, Beth Reid, and Martin White. An analytic model for redshift-space distortions. MNRAS, 437(1):588–599, January 2014. doi: 10.1093/mnras/stt1916.

[205] Martin White. The Zel'dovich approximation. MNRAS, 439(4):3630–3640, April 2014. doi: 10.1093/mnras/stu209.

[206] Matthew Lewandowski, Ashley Perko, and Leonardo Senatore. Analytic prediction of baryonic effects from the EFT of large scale structures. J. Cosmology Astropart. Phys., 2015(5):019–019, May 2015. doi: 10.1088/1475-7516/2015/05/019.

[207] J. A. Peacock and S. J. Dodds. Reconstructing the Linear Power Spectrum of Cosmological Mass Fluctuations. MNRAS, 267:1020, April 1994. doi: 10.1093/mnras/267.4.1020.

[208] Román Scoccimarro. Redshift-space distortions, pairwise velocities, and nonlinearities. Phys. Rev. D, 70(8):083007, October 2004. doi: 10.1103/PhysRevD.70.083007.

[209] Beth A. Reid, Hee-Jong Seo, Alexie Leauthaud, Jeremy L. Tinker, and Martin White. A 2.5 per cent measurement of the growth rate from small-scale redshift space clustering of SDSS-III CMASS galaxies. MNRAS, 444(1):476–502, October 2014. doi: 10.1093/mnras/stu1391.

[210] Zhongxu Zhai, Jeremy L. Tinker, Matthew R. Becker, Joseph DeRose, Yao-Yuan Mao, Thomas McClintock, Sean McLaughlin, Eduardo Rozo, and Risa H. Wechsler. The Aemulus Project. III. Emulation of the Galaxy Correlation Function. ApJ, 874(1):95, March 2019. doi: 10.3847/1538-4357/ab0d7b.

[211] Johannes U. Lange, Frank C. van den Bosch, Andrew R. Zentner, Kuan Wang, Andrew P. Hearin, and Hong Guo. Cosmological Evidence Modelling: a new simulation-based approach to constrain cosmology on non-linear scales. MNRAS, 490(2):1870–1878, December 2019. doi: 10.1093/mnras/stz2664.

[212] Joseph DeRose, Risa H. Wechsler, Jeremy L. Tinker, Matthew R. Becker, Yao-Yuan Mao, Thomas McClintock, Sean McLaughlin, Eduardo Rozo, and Zhongxu Zhai. The AEMULUS Project. I. Numerical Simulations for Precision Cosmology. ApJ, 875(1):69, April 2019. doi: 10.3847/1538-4357/ab1085.

[213] Naoyuki Tamura, Naruhisa Takato, Atsushi Shimono, Yuki Moritani, Kiyoto Yabe, Yuki Ishizuka, Akitoshi Ueda, Yukiko Kamata, Hrand Aghazarian, Stéphane Arnouts, Gabriel Barban, Robert H. Barkhouser, Renato C. Borges, David F. Braun, Michael A. Carr, Pierre-Yves Chabaud, Yin-Chang Chang, Hsin-Yo Chen, Masashi Chiba, Richard C. Y. Chou, You-Hua Chu, Judith Cohen, Rodrigo P. de Almeida, Antonio C. de Oliveira, Ligia S. de Oliveira, Richard G. Dekany, Kjetil Dohlen, Jesulino B. dos Santos, Leandro H. dos Santos, Richard Ellis, Maximilian Fabricius, Didier Ferrand, Décio Ferreira, Mirek Golebiowski, Jenny E. Greene, Johannes Gross, James E. Gunn,

Randolph Hammond, Albert Harding, Murdock Hart, Timothy M. Heckman, Christopher M. Hirata, Paul Ho, Stephen C. Hope, Larry Hovland, Shu-Fu Hsu, Yen-Shan Hu, Ping-Jie Huang, Marc Jaquet, Yipeng Jing, Jennifer Karr, Masahiko Kimura, Matthew E. King, Eiichiro Komatsu, Vincent Le Brun, Olivier Le Fèvre, Arnaud Le Fur, David Le Mignant, Hung-Hsu Ling, Craig P. Loomis, Robert H. Lupton, Fabrice Madec, Peter Mao, Lucas S. Marrara, Claudia Mendes de Oliveira, Yosuke Minowa, Chaz Morantz, Hitoshi Murayama, Graham J. Murray, Youichi Ohyama, Joseph Orndorff, Sandrine Pascal, Jefferson M. Pereira, Daniel Reiley, Martin Reinecke, Andreas Ritter, Mitsuko Roberts, Mark A. Schwochert, Michael D. Seiffert, Stephen A. Smee, Laerte Sodre, David N. Spergel, Aaron J. Steinkraus, Michael A. Strauss, Christian Surace, Yasushi Suto, Nao Suzuki, John Swinbank, Philip J. Tait, Masahiro Takada, Tomonori Tamura, Yoko Tanaka, Laurence Tresse, Orlando Verducci, Didier Vibert, Clement Vidal, Shiang-Yu Wang, Chih-Yi Wen, Chi-Hung Yan, and Naoki Yasuda. Prime Focus Spectrograph (PFS) for the Subaru telescope: overview, recent progress, and future perspectives. In Christopher J. Evans, Luc Simard, and Hideki Takami, editors, *Ground-based and Airborne Instrumentation for Astronomy VI*, volume 9908 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 99081M, August 2016. doi: 10.1117/12.2232103.

[214] R. Laureijs, J. Amiaux, S. Arduini, J. L. Augueères, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet, B. Garilli, P. Gondoin, L. Guzzo, J. Hoar, H. Hoekstra, R. Holmes, T. Kitching, T. Maciaszek, Y. Mellier, F. Pasian, W. Percival, J. Rhodes, G. Saavedra Criado, M. Sauvage, R. Scaramella, L. Valenziano, S. Warren, R. Bender, F. Castander, A. Cimatti, O. Le Fèvre, H. Kurki-Suonio, M. Levi, P. Lilje, G. Meylan, R. Nichol, K. Pedersen, V. Popa, R. Rebolo Lopez, H. W. Rix, H. Rottgering, W. Zeilinger, F. Grupp, P. Hudelot, R. Massey, M. Meneghetti, L. Miller, S. Paltani, S. Paulin-Henriksson, S. Pires, C. Saxton, T. Schrabback, G. Seidel, J. Walsh, N. Aghanim, L. Amendola, J. Bartlett, C. Baccigalupi, J. P. Beaulieu, K. Benabed, J. G. Cuby, D. Elbaz, P. Fosalba, G. Gavazzi, A. Helmi, I. Hook, M. Irwin, J. P. Kneib,

M. Kunz, F. Mannucci, L. Moscardini, C. Tao, R. Teyssier, J. Weller, G. Zamorani, M. R. Zapatero Osorio, O. Boulade, J. J. Foumond, A. Di Giorgio, P. Guttridge, A. James, M. Kemp, J. Martignac, A. Spencer, D. Walton, T. Blümchen, C. Bonoli, F. Bortoletto, C. Cerna, L. Corcione, C. Fabron, K. Jahnke, S. Ligori, F. Madrid, L. Martin, G. Morgante, T. Pamplona, E. Prieto, M. Riva, R. Toledo, M. Trifoglio, F. Zerbi, F. Abdalla, M. Douspis, C. Grenet, S. Borgani, R. Bouwens, F. Courbin, J. M. Delouis, P. Dubath, A. Fontana, M. Frailis, A. Grazian, J. Koppenhöfer, O. Mansutti, M. Melchior, M. Mignoli, J. Mohr, C. Neissner, K. Noddle, M. Poncet, M. Scodeggio, S. Serrano, N. Shane, J. L. Starck, C. Surace, A. Taylor, G. Verdoes-Kleijn, C. Vuerli, O. R. Williams, A. Zacchei, B. Altieri, I. Escudero Sanz, R. Kohley, T. Oosterbroek, P. Astier, D. Bacon, S. Bardelli, C. Baugh, F. Bellagamba, C. Benoist, D. Bianchi, A. Biviano, E. Branchini, C. Carbone, V. Cardone, D. Clements, S. Colombi, C. Conselice, G. Cresci, N. Deacon, J. Dunlop, C. Fedeli, F. Fontanot, P. Franzetti, C. Giocoli, J. Garcia-Bellido, J. Gow, A. Heavens, P. Hewett, C. Heymans, A. Holland, Z. Huang, O. Ilbert, B. Joachimi, E. Jennins, E. Kerins, A. Kiessling, D. Kirk, R. Kotak, O. Krause, O. Lahav, F. van Leeuwen, J. Lesgourgues, M. Lombardi, M. Magliocchetti, K. Maguire, E. Majerotto, R. Maoli, F. Marulli, S. Maurogordato, H. McCracken, R. McLure, A. Melchiorri, A. Merson, M. Moresco, M. Nonino, P. Norberg, J. Peacock, R. Pello, M. Penny, V. Pettorino, C. Di Porto, L. Pozzetti, C. Quercellini, M. Radovich, A. Rassat, N. Roche, S. Ronayette, E. Rossetti, B. Sartoris, P. Schneider, E. Semboloni, S. Serjeant, F. Simpson, C. Skordis, G. Smadja, S. Smartt, P. Spano, S. Spiro, M. Sullivan, A. Tilquin, R. Trotta, L. Verde, Y. Wang, G. Williger, G. Zhao, J. Zoubian, and E. Zucca. Euclid Definition Study Report. *arXiv e-prints*, art. arXiv:1110.3193, October 2011. doi: 10.48550/arXiv.1110.3193.

[215] Uroš Seljak. Extracting Primordial Non-Gaussianity without Cosmic Variance. Phys. Rev. Lett., 102(2):021302, January 2009. doi: 10.1103/PhysRevLett.102.021302.

[216] Patrick McDonald and Uroš Seljak. How to evade the sample variance limit on mea-

surements of redshift-space distortions. J. Cosmology Astropart. Phys., 2009(10):007, October 2009. doi: 10.1088/1475-7516/2009/10/007.

[217] DESI Collaboration, Amir Aghamousa, Jessica Aguilar, Steve Ahlen, Shadab Alam, Lori E. Allen, Carlos Allende Prieto, James Annis, Stephen Bailey, Christophe Balland, Otger Ballester, Charles Baltay, Lucas Beaufore, Chris Bebek, Timothy C. Beers, Eric F. Bell, José Luis Bernal, Robert Besuner, Florian Beutler, Chris Blake, Hannes Bleuler, Michael Blomqvist, Robert Blum, Adam S. Bolton, Cesar Briceno, David Brooks, Joel R. Brownstein, Elizabeth Buckley-Geer, Angela Burden, Etienne Burtin, Nicolas G. Busca, Robert N. Cahn, Yan-Chuan Cai, Laia Cardiel-Sas, Raymond G. Carlberg, Pierre-Henri Carton, Ricard Casas, Francisco J. Castander, Jorge L. Cervantes-Cota, Todd M. Claybaugh, Madeline Close, Carl T. Coker, Shaun Cole, Johan Comparat, Andrew P. Cooper, M. C. Cousinou, Martin Crocce, Jean-Gabriel Cuby, Daniel P. Cunningham, Tamara M. Davis, Kyle S. Dawson, Axel de la Macorra, Juan De Vicente, Timothée Delubac, Mark Derwent, Arjun Dey, Govinda Dhungana, Zhejie Ding, Peter Doel, Yutong T. Duan, Anne Ealet, Jerry Edelstein, Sarah Eftekharzadeh, Daniel J. Eisenstein, Ann Elliott, Stéphanie Escoffier, Matthew Evatt, Parker Fagrelius, Xiaohui Fan, Kevin Fanning, Arya Farahi, Jay Farihi, Ginevra Favole, Yu Feng, Enrique Fernandez, Joseph R. Findlay, Douglas P. Finkbeiner, Michael J. Fitzpatrick, Brenna Flaugher, Samuel Flender, Andreu Font-Ribera, Jaime E. Forero-Romero, Pablo Fosalba, Carlos S. Frenk, Michele Fumagalli, Boris T. Gaensicke, Giuseppe Gallo, Juan Garcia-Bellido, Enrique Gaztanaga, Nicola Pietro Gentile Fusillo, Terry Gerard, Irena Gershkovich, Tommaso Giannantonio, Denis Gillet, Guillermo Gonzalez-de-Rivera, Violeta Gonzalez-Perez, Shelby Gott, Or Graur, Gaston Gutierrez, Julien Guy, Salman Habib, Henry Heetderks, Ian Heetderks, Katrin Heitmann, Wojciech A. Hellwing, David A. Herrera, Shirley Ho, Stephen Holland, Klaus Honscheid, Eric Huff, Timothy A. Hutchinson, Dragan Huterer, Ho Seong Hwang, Joseph Maria Illa Laguna, Yuzo Ishikawa, Dianna Jacobs, Niall Jeffrey, Patrick Jelinsky, Elise Jennings, Linhua Jiang, Jorge Jimenez, Jennifer

Johnson, Richard Joyce, Eric Jullo, Stéphanie Juneau, Sami Kama, Armin Karcher, Sonia Karkar, Robert Kehoe, Noble Kennamer, Stephen Kent, Martin Kilbinger, Alex G. Kim, David Kirkby, Theodore Kisner, Ellie Kitanidis, Jean-Paul Kneib, Sergey Koposov, Eve Kovacs, Kazuya Koyama, Anthony Kremin, Richard Kron, Luzius Kronig, Andrea Kueter-Young, Cedric G. Lacey, Robin Lafever, Ofer Lahav, Andrew Lambert, Michael Lampton, Martin Landriau, Dustin Lang, Tod R. Lauer, Jean-Marc Le Goff, Laurent Le Guillou, Auguste Le Van Suu, Jae Hyeon Lee, Su-Jeong Lee, Daniela Leitner, Michael Lesser, Michael E. Levi, Benjamin L'Huillier, Baojiu Li, Ming Liang, Huan Lin, Eric Linder, Sarah R. Loebman, Zarija Lukić, Jun Ma, Niall MacCrann, Christophe Magneville, Laleh Makarem, Marc Manera, Christopher J. Manser, Robert Marshall, Paul Martini, Richard Massey, Thomas Matheson, Jeremy McCauley, Patrick McDonald, Ian D. McGreer, Aaron Meisner, Nigel Metcalfe, Timothy N. Miller, Ramon Miquel, John Moustakas, Adam Myers, Milind Naik, Jeffrey A. Newman, Robert C. Nichol, Andrina Nicola, Luiz Nicolati da Costa, Jundan Nie, Gustavo Niz, Peder Norberg, Brian Nord, Dara Norman, Peter Nugent, Thomas O'Brien, Minji Oh, Knut A. G. Olsen, Cristobal Padilla, Hamsa Padmanabhan, Nikhil Padmanabhan, Nathalie Palanque-Delabrouille, Antonella Palmese, Daniel Pappalardo, Isabelle Pâris, Changbom Park, Anna Patej, John A. Peacock, Hiranya V. Peiris, Xiyan Peng, Will J. Percival, Sandrine Perruchot, Matthew M. Pieri, Richard Pogge, Jennifer E. Pollack, Claire Poppett, Francisco Prada, Abhishek Prakash, Ronald G. Probst, David Rabinowitz, Anand Raichoor, Chang Hee Ree, Alexandre Refregier, Xavier Regal, Beth Reid, Kevin Reil, Mehdi Rezaie, Constance M. Rockosi, Natalie Roe, Samuel Ronayette, Aaron Roodman, Ashley J. Ross, Nicholas P. Ross, Graziano Rossi, Eduardo Rozo, Vanina Ruhlmann-Kleider, Eli S. Rykoff, Cristiano Sabiu, Lado Samushia, Eusebio Sanchez, Javier Sanchez, David J. Schlegel, Michael Schneider, Michael Schubnell, Aurélia Secroun, Uros Seljak, Hee-Jong Seo, Santiago Serrano, Arman Shafieloo, Huanyuan Shan, Ray Sharples, Michael J. Sholl, William V. Shourt, Joseph H. Silber, David R. Silva, Martin M. Sirk, Anze Slosar, Alex Smith, George F. Smoot, Debopam Som, Yong-Seon Song, David Sprayberry, Ryan Staten, Andy Ste-

fanik, Gregory Tarle, Suk Sien Tie, Jeremy L. Tinker, Rita Tojeiro, Francisco Valdes, Octavio Valenzuela, Monica Valluri, Mariana Vargas-Magana, Licia Verde, Alistair R. Walker, Jiali Wang, Yuting Wang, Benjamin A. Weaver, Curtis Weaverdyck, Risa H. Wechsler, David H. Weinberg, Martin White, Qian Yang, Christophe Yeche, Tianmeng Zhang, Gong-Bo Zhao, Yi Zheng, Xu Zhou, Zhimin Zhou, Yaling Zhu, Hu Zou, and Ying Zu. The DESI Experiment Part II: Instrument Design. *arXiv e-prints*, art. arXiv:1611.00037, October 2016. doi: 10.48550/arXiv.1611.00037.

[218] J. Carron. On the assumption of Gaussianity for cosmological two-point statistics and parameter dependent covariance matrices. A&A, 551:A88, March 2013. doi: 10.1051/0004-6361/201220538.

[219] Darsh Kodwani, David Alonso, and Pedro Ferreira. The effect on cosmological parameter estimation of a parameter dependent covariance matrix. *The Open Journal of Astrophysics*, 2(1):3, March 2019. doi: 10.21105/astro.1811.11584.

[220] Peter Coles and Bernard Jones. A lognormal model for the cosmological mass distribution. MNRAS, 248:1–13, January 1991. doi: 10.1093/mnras/248.1.1.

[221] Pierluigi Monaco, Tom Theuns, and Giuliano Taffoni. The pinocchio algorithm: pinpointing orbit-crossing collapsed hierarchical objects in a linear density field. MNRAS, 331(3):587–608, April 2002. doi: 10.1046/j.1365-8711.2002.05162.x.

[222] P. Monaco, E. Sefusatti, S. Borgani, M. Crocce, P. Fosalba, R. K. Sheth, and T. Theuns. An accurate tool for the fast generation of dark matter halo catalogues. MNRAS, 433(3):2389–2402, August 2013. doi: 10.1093/mnras/stt907.

[223] F. S. Kitaura, G. Yepes, and F. Prada. Modelling baryon acoustic oscillations with perturbation theory and stochastic halo biasing. MNRAS, 439:L21–L25, March 2014. doi: 10.1093/mnrasl/slt172.

[224] Francisco-Shu Kitaura, Héctor Gil-Marín, Claudia G. Scóccola, Chia-Hsun Chuang, Volker Müller, Gustavo Yepes, and Francisco Prada. Constraining the halo bispectrum

in real and redshift space from perturbation theory and non-linear stochastic bias. MNRAS, 450(2):1836–1845, June 2015. doi: 10.1093/mnras/stv645.

[225] Román Scoccimarro and Ravi K. Sheth. PTHALOS: a fast method for generating mock galaxy distributions. MNRAS, 329(3):629–640, January 2002. doi: 10.1046/j.1365-8711.2002.04999.x.

[226] Marc Manera, Roman Scoccimarro, Will J. Percival, Lado Samushia, Cameron K. McBride, Ashley J. Ross, Ravi K. Sheth, Martin White, Beth A. Reid, Ariel G. Sánchez, Roland de Putter, Xiaoying Xu, Andreas A. Berlind, Jonathan Brinkmann, Claudia Maraston, Bob Nichol, Francesco Montesano, Nikhil Padmanabhan, Ramin A. Skibba, Rita Tojeiro, and Benjamin A. Weaver. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: a large sample of mock galaxy catalogues. MNRAS, 428(2):1036–1054, January 2013. doi: 10.1093/mnras/sts084.

[227] Marc Manera, Lado Samushia, Rita Tojeiro, Cullan Howlett, Ashley J. Ross, Will J. Percival, Hector Gil-Marín, Joel R. Brownstein, Angela Burden, and Francesco Montesano. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: mock galaxy catalogues for the low-redshift sample. MNRAS, 447(1):437–445, February 2015. doi: 10.1093/mnras/stu2465.

[228] Svetlin Tassev, Matias Zaldarriaga, and Daniel J. Eisenstein. Solving large scale structure in ten easy steps with COLA. J. Cosmology Astropart. Phys., 2013(6):036, June 2013. doi: 10.1088/1475-7516/2013/06/036.

[229] Martin White, Jeremy L. Tinker, and Cameron K. McBride. Mock galaxy catalogues using the quick particle mesh method. MNRAS, 437(3):2594–2606, January 2014. doi: 10.1093/mnras/stt2071.

[230] Chia-Hsun Chuang, Francisco-Shu Kitaura, Francisco Prada, Cheng Zhao, and Gustavo Yepes. EZmocks: extending the Zel'dovich approximation to generate mock

galaxy catalogues with accurate clustering statistics. MNRAS, 446(3):2621–2628, January 2015. doi: 10.1093/mnras/stu2301.

[231] Yu Feng, Man-Yat Chu, Uroš Seljak, and Patrick McDonald. FASTPM: a new scheme for fast simulations of dark matter and haloes. MNRAS, 463(3):2273–2286, December 2016. doi: 10.1093/mnras/stw2123.

[232] Roland de Putter, Christian Wagner, Olga Mena, Licia Verde, and Will J. Percival. Thinking outside the box: effects of modes larger than the survey on matter power spectrum covariance. J. Cosmology Astropart. Phys., 2012(4):019, April 2012. doi: 10.1088/1475-7516/2012/04/019.

[233] Xiaoying Xu, Nikhil Padmanabhan, Daniel J. Eisenstein, Kushal T. Mehta, and Antonio J. Cuesta. A 2 per cent distance to z = 0.35 by reconstructing baryon acoustic oscillations - II. Fitting techniques. MNRAS, 427(3):2146–2167, December 2012. doi: 10.1111/j.1365-2966.2012.21573.x.

[234] Adrian C. Pope and István Szapudi. Shrinkage estimation of the power spectrum covariance matrix. MNRAS, 389(2):766–774, September 2008. doi: 10.1111/j.1365-2966.2008.13561.x.

[235] Dante J. Paz and Ariel G. Sánchez. Improving the precision matrix for precision cosmology. MNRAS, 454(4):4326–4334, December 2015. doi: 10.1093/mnras/stv2259.

[236] David W. Pearson and Lado Samushia. Estimating the power spectrum covariance matrix with fewer mock samples. MNRAS, 457(1):993–999, March 2016. doi: 10.1093/mnras/stw062.

[237] Ji Meng Loh. A Valid and Fast Spatial Bootstrap for Correlation Functions. ApJ, 681(1):726–734, July 2008. doi: 10.1086/588631.

[238] O. Friedrich, S. Seitz, T. F. Eifler, and D. Gruen. Performance of internal covariance estimators for cosmic shear correlation functions. MNRAS, 456(3):2662–2680, March 2016. doi: 10.1093/mnras/stv2833.

[239] Ginevra Favole, Benjamin R. Granett, Javier Silva Lafaurie, and Domenico Sapone. Does jackknife scale really matter for accurate large-scale structure covariances? MNRAS, 505(4):5833–5845, August 2021. doi: 10.1093/mnras/stab1720.

[240] Cameron Grove, Chia-Hsun Chuang, Ningombam Chandrachani Devi, Lehman Garrison, Benjamin L'Huillier, Yu Feng, John Helly, César Hernández-Aguayo, Shadab Alam, Hanyu Zhang, Yu Yu, Shaun Cole, Daniel Eisenstein, Peder Norberg, Risa Wechsler, David Brooks, Kyle Dawson, Martin Landriau, Aaron Meisner, Claire Poppett, Gregory Tarlé, and Octavio Valenzuela. The DESI N-body simulation project - I. Testing the robustness of simulations for the DESI dark time survey. MNRAS, 515 (2):1854–1870, September 2022. doi: 10.1093/mnras/stac1947.

[241] John Wishart. THE GENERALISED PRODUCT MOMENT DISTRIBUTION IN SAMPLES FROM A NORMAL MULTIVARIATE POPULATION. *Biometrika*, 20A (1-2):32–52, 12 1928. ISSN 0006-3444. doi: 10.1093/biomet/20A.1-2.32. URL https://doi.org/10.1093/biomet/20A.1-2.32.

[242] S. Escoffier, M. C. Cousinou, A. Tilquin, A. Pisani, A. Aguichine, S. de la Torre, A. Ealet, W. Gillard, and E. Jullo. Jackknife resampling technique on mocks: an alternative method for covariance matrix estimation. *arXiv e-prints*, art. arXiv:1606.00233, June 2016. doi: 10.48550/arXiv.1606.00233.

[243] Shadab Alam, John A. Peacock, Daniel J. Farrow, J. Loveday, and A. M. Hopkins. Using GAMA to probe the impact of small-scale galaxy physics on nonlinear redshift-space distortions. MNRAS, 503(1):59–76, May 2021. doi: 10.1093/mnras/stab409.

[244] Davide Bianchi, Héctor Gil-Marín, Rossana Ruggeri, and Will J. Percival. Measuring line-of-sight-dependent Fourier-space clustering using FFTs. MNRAS, 453(1):L11–L15, October 2015. doi: 10.1093/mnrasl/slv090.

[245] Nick Hand, Yu Feng, Florian Beutler, Yin Li, Chirag Modi, Uroš Seljak, and Zachary

Slepian. nbodykit: An Open-source, Massively Parallel Toolkit for Large-scale Structure. AJ, 156(4):160, October 2018. doi: 10.3847/1538-3881/aadae0.

[246] Variu et al. DESI Mock Challenge: Constructing DESI galaxy catalogues based on FastPM simulations. *XXX*, in prep.

[247] Balaguera-Antolínez et al. DESI Mock Challenge: BAM. *XXX*, in prep.

[248] EZ et al. DESI Mock Challenge: EZ. *XXX*, in prep.

[249] Chen et al. DESI Mock Challenge: analytical. *XXX*, in prep.