Fast Fiber Assign: Emulating Fiber Assignment Effects for Realistic DESI Catalogs

by

Malik Muhammad Sikandar Hanif

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Physics)
in The University of Michigan
2025

Doctoral Committee:

Professor Dragan Huterer, Co-Chair
Professor Emeritus Gregory Tarlè, Co-Chair
Professor Eric F. Bell
Assistant Professor Ashley J. Ross, The Ohio State University
Professor Thomas A. Schwarz

Malik Muhammad Sikandar Hanif
sikandar@umich.edu
ORCID iD: 0009-0006-2583-5006

© Malik Muhammad Sikandar Hanif 2025

All Rights Reserved

ACKNOWLEDGMENTS

When I joined the Dark Energy Spectroscopic Instrument (DESI) group at the University of Michigan at the start of my program, I was told that the great thing about working in a collaboration is that you have so many advisors. The last five years have only confirmed that time and again, making me happily indebted to so many great colleagues and researchers. First, I would like to thank my advisors. Professors Dragan Huterer and Gregory Tarlé, who enthusiastically accepted me into their group when I was a first-year graduate student with very little research experience in cosmology and helped get me up to speed as we all worked from home in the midst of a pandemic. Their expertise was crucial to getting me started and they were very supportive as my work branched out. Greg is also the source of the comment at the start of this section.

One of those advisors from the wider collaboration is Assistant Professor Ashley Ross, towards whom my gratitude is large enough to take up more space than he has often asked me to clear up on Perlmutter's storage system. He has been an invaluable resource every step of the way as I've worked with the DESI Large Scale Structure (LSS) pipeline and he is the person I have worked most closely with on my projects. I am so thankful for his answering any and all questions I have sent his way.

I would also like to thank the rest of my dissertation committee — Professors Eric Bell and Tom Schwarz. Eric brought his expertise from Astronomy to the committee as a cognate member. Aside from being thankful for his role as a committee member, I am also grateful to Tom for being a wonderful supervisor during my time as a Graduate Student Instructor (GSI) for one of his courses. His passion for teaching and concern

for students is truly inspirational.

Within DESI, I would first like to thank Davide Bianchi for having me as part of his project to create the Fast Fiber Assign (FFA) emulator, which forms the crux of this dissertation. I thank Arnaud de Mattia for guiding me as I volunteered early on with the LSS pipeline, helping me familiarize myself with various codes that were used in this analysis. I also thank Cheng Zhao, Daniel Forero-Sanchez and Svyatoslav Trusov for all the discussions and ideas exchanged, and for their part in introducing me to various analysis methods used in this dissertation.

So many thanks to my colleagues and friends at the University of Michigan. In particular, to my office-mates, Otàvio, Jiaming and Tianke for always being a great time, no matter which permutation of us showed up on any given day. Also, a big thanks to Uendert for, aside from his crucial work in DESI, explaining what a template fit truly is.

I also acknowledge the US Department of Energy (DOE), Office of Science, Office of High-Energy Physics and the National Energy Research Scientific Computing Center (NERSC). Without the funding and computational resources to process such large amounts of data, this work would not have been possible. I was also funded via Graduate Student Instructor (GSI) roles in the Department of Physics at the University of Michigan. I am immensely grateful to both the department and the university as a whole for providing me this opportunity.

My journey to today would have been impossible without so many people I have known outside my life in physics. I cannot thank Ms. Maleeha from fifth-grade English enough for the one question she asked me that changed the course of my life. I also thank Mr. Mehmood for his infinite patience with me before mathematics finally clicked for me. I thank my friends, too numerous to name, who have provided me with support and friendship since middle school.

Finally, above all but in no particular order, I thank my family. My brother,

Shehryar, for being the best little brother one could ask for and for also geeking out over LaTeX. I thank both my parents for their hard work and emphasis on providing my brother and I the best education they could. In particular, no amount of words can encapsulate my appreciation for my mom for being a pillar of support and my biggest fan. And, of course, the greatest of gratitude to EC for being my best friend and confidant. Again, no words will suffice.

TABLE OF CONTENTS

ACKNOWLEDGMENTS ii

LIST OF FIGURES vii

LIST OF TABLES xiv

ABSTRACT xvi

CHAPTER

1. Introduction 1

1.1 The Universe — Homogeneous and Isotropic 1
1.1.1 Distances in Cosmology 4
1.1.2 The Friedmann Equations 6
1.1.3 Components of the Universe 9
1.2 The Universe — Not so Homogeneous 10
1.2.1 Cosmic Inflation 12
1.2.2 Large Scale Structure 12
1.3 The Dark Energy Spectroscopic Instrument (DESI) 25
1.4 Outline 31

2. Cosmology from Clustering 35

2.1 Extracting Information From Surveys 35
2.1.1 Classic BAO Approach 35
2.1.2 Compressed Full Shape Analysis 36
2.1.3 Full Modeling 38
2.1.4 ShapeFit 39
2.2 Covariance Estimation 41
2.2.1 Jackknife Covariances 43

3. Fiber Assignment and Fast Fiber Assignment 50

3.1 Inverse Probability Weights . . . 51
3.2 Simulating Fiber Assignment . . . 54
3.2.1 Alternative Merged Target Ledger (altMTL) . . . 54
3.2.2 Fast Fiber Assignment (FFA) . . . 55
3.3 Mock Catalogs . . . 57
3.4 Adapting FFA for Mock Catalogs . . . 60
3.5 Comparing altMTL and FFA Mock Catalogs . . . 61
3.6 Impact . . . 63

4. Fast Fiber Assign: Further Investigations . . . 67

4.1 Methods . . . 68
4.2 Results and Discussion . . . 70
4.2.1 Effect on Kernel . . . 70
4.2.2 Clustering Results . . . 74
4.2.3 Normed Trace Comparisons . . . 74
4.2.4 Comparing Fit Parameters . . . 75

5. DESI Y1 Cosmology . . . 103

5.1 DESI Data Release 1 . . . 103
5.1.1 LRG Sample . . . 104
5.1.2 ELG Sample . . . 104
5.1.3 QSO Sample . . . 104
5.1.4 BGS Sample . . . 104
5.2 BAO Analysis . . . 105
5.3 Cosmological Constraints From BAO . . . 107
5.3.1 ΛCDM . . . 109
5.3.2 Dark Energy Constraints . . . 110
5.4 Full Shape Analysis . . . 112
5.5 Dark Energy Constraints From Full Shape Analysis . . . 113
5.5.1 Flat ΛCDM . . . 114
5.5.2 Flat w_0w_aCDM . . . 115
5.6 Summary . . . 116

6. Concluding Remarks . . . 119

BIBLIOGRAPHY . . . 124

LIST OF FIGURES

Figure

1.1 Hubble's velocity-distance diagram for galaxies. Radial velocities, corrected for solar motion, are plotted against distances estimated from stars and mean luminosities of nebulae in a cluster. Figure from [50]. . 2

1.2 This figure represents the evolution of the densities of different components of a ΛCDM universe. The horizontal axis is the logarithm of the scale factor, here used as a measure of the passage of time. The vertical axis is the relative density parameter from Equation 1.15. 11

1.3 The distribution of galaxies in the Sloan Digital Sky Survey (SDSS) main galaxy sample, showing galaxies within 1.25 degrees of the Celestial Equator. Figure from [106]. 13

1.4 An artistic rendering of the baryon acoustic oscillations. Image by Zosia Rostomian and provided by Lawrence Berkeley National Laboratory is licensed under CC BY-NC-ND 4.0. ©2017 The Regents of the University of California, Lawrence Berkeley National Laboratory. 22

1.5 A graphic representing the quantities that are physically measured parallel and perpendicular to the line of sight of an observer, in this case, looking at a spherical feature. 22

1.6 A graphic representing the redshift-space distortions. On large scales, velocities are directed 'inward' towards large overdensities — denoted by the wide blue arrows. This manifests in redshift space as the Kaiser effect, where the object appears squashed. On smaller scales, galaxies have random motions (peculiar velocities) that cause an elongation along the line of sight in redshift space. This is often called the Fingers of God effect. 23

1.7 The large-scale redshift-space correlation function of the SDSS LRG sample. The colored lines are fits from different models. Note around comoving separation 100 h^{-1} Mpc the peak in the correlation function corresponding to the BAO feature. Figure reproduced from [38]. 24

1.8 An example of the linear matter power spectrum $P(k)$ given a cosmology. Note the power law dependence on either 'extreme' scale. On large scales (small k), the function scales as $k^{n_s} \approx k^1$ and on small scales (large k), it goes as $k^{n_s-4} \approx k^3$. 24

1.9 A model of the 4 m Mayall Telescope with the instrumentation built for DESI. Figure reproduced from [29]. 26
1.10 A view of a partially assembled focal plane petal with an array of robotic positioners that is each connected to a fiber-optic cable. Photo by Marilyn Chung and provided by Lawrence Berkeley National Laboratory is licensed under CC BY-NC-ND 4.0. ©2017 The Regents of the University of California, Lawrence Berkeley National Laboratory 27
1.11 A DESI fiber positioner. On the insert (top-right), the centers of the two axes are indicated. Figure reproduced from [90]. 28
1.12 A set of example LRG spectra from DESI in the left panels. They are ranked according to quality (increasing from 0 to 4). Qualities 4 and 3 have multiple spectral lines while qualities 2 and below do not have spectral features that are robustly identifiable. The right panels show the DESI Legacy Survey images of the LRG targets. The spectra in gray, black, and orange are the originally observed galaxy spectrum, the smoothed spectrum with a Gaussian filter, and the error spectrum, respectively. Caption adapted from and figure reproduced from [54]. . . 29
1.13 Flowchart of the spectroscopic pipeline for DESI, showing the many processing steps that the data has to undergo before it can be analyzed for cosmology. Figure reproduced from [47]. The section numbers in the figure refer to the original paper. 33
1.14 The left panel of this figure shows the eccentric axis $(\theta - \Phi)$ kinematics. Whenever R_2 is retracted within the dashed circle E, a positioner is guaranteed free rotation about θ without any obstruction from its neighbors. On the right, this figure shows how the patrol coverage extends beyond the pitch. This assures complete coverage but requires implementation of collision avoidance algorithms during the move. Caption adapted and figure reproduced from [90]. 34
2.1 The top panel of this figure shows the linear power spectrum template at $z = 0$ for various cosmologies. The middle panel isolates the BAO 'wiggle' part, $P_w(k)$, where we can see the effect of the different sound horizon scales for each cosmology. The bottom panel shows the alignment of the oscillatory pattern after correcting for the difference in the sound horizon scales. Figure from [77]. 37
2.2 A comparison of the diagonals of the estimated covariance matrix of the real-space correlation function. These were obtained from 1000 Poisson random cubic mocks. The quantity on the vertical axis is the ratio of the mean of the diagonals of the covariance obtained via jackknifing to the diagonals of a directly calculated mock covariance. The Mohammad-Percival correction takes care of the scale-dependent bias that is noticeable in the case of the 'no correction' approach. Note that these are Poisson random catalogs and the $\mathrm{cov}(CC, AA)$ term in Equation 2.10 is zero. Figure reproduced from [97]. 45
2.3 Schematic showing the fitting approach utilized by FitCov. Figure reproduced from [97]. 48

2.4 The average of the relative bias defined in Equation 2.12 plotted as a function of separation s for different number densities. Figure reproduced from [97]. 49
3.1 Schematic for visualizing the maximum anticorrelation algorithm for bitweights. Each of the columns under the 'step' heading corresponds to a galaxy. For each galaxy, the rows indicate whether it was assigned (orange circle) or not (blank box) on any given realization. Figure reproduced from [10]. 58
3.2 The angular assignment completeness of the dark-time galaxies observed by DESI in its first year. This is defined as the number of tracers assigned to a fiber divided by the total number of tracers in the corresponding full sample. Figure from my work in [10]. 62
3.3 The left column of this figure shows the angular assignment completeness of the three dark-time tracers, measured from altMTL mocks. The right column shows the relative behavior of FFA and altMTL. Figure from my work in [10]. 63
3.4 A comparison of the redshift distributions ($dz = 0.02$) from various versions of an Abacus mock, with and without fiber assignment run. The 'norm' versions are the simulated altMTL and FFA histograms rescaled by $N_{\mathrm{complete}}/N_{\mathrm{assigned}}$ to match the amplitude of the complete sample. Figure adapted from my work in [10]. 64
3.5 Spread of the fiducial weights from the altMTL and FFA mocks, using $dw = 0.33$. The 'discrete' weights are the fiducial FFA weights rounded to the nearest integer. Figure adapted from my work in [10]. 64
3.6 Normalized pair counts from altMTL mocks with different kinds of weights divided by normalized pair counts from the full sample of targets. Figure from [10]. 65
3.7 Normalized pair counts from FFA mocks with different kinds of weights divided by normalized pair counts from the complete sample of targets. Figure from [10]. 65
4.1 This figure contains three subplots that visualize what the kernel that the emulator (trained on real LRG data) learned looks like, with the color representing the probability of assignment. The first figure on the left shows the unmodified version of the kernel. The middle figure shows the same kernel with $\beta_1 = -0.1$ and the right figure shows the first kernel with β_0=1.2. 71
4.2 This figure shows the distribution of the number of assigned LRG targets for a single `Abacus` mock catalog when run through the emulation pipeline with different parameters. Note that some of the points are super-imposed with multiple data points, but the dependence on different parameters is clear. 72

4.3 This figure shows the distribution of the number of assigned QSO targets for a single `Abacus` mock catalog when run through the emulation pipeline with different parameters. Note that some of the points are super-imposed with multiple data points, but the dependence on different parameters is clear. 73
4.4 Figure showing the ratio of the number counts of assigned pairs to the number counts of pairs in the complete sample of LRG targets for `Abacus` mock 0 for different emulator runs. The groups in each subplot title refer to what the mean ratio of counts is inside the collision window for each graph. 83
4.5 Figure showing the ratio of the number counts of assigned pairs to the number counts of pairs in the complete sample of QSO targets for `Abacus` mock 0 for different emulator runs. The groups in each subplot title refer to what the mean ratio of counts is inside the collision window for each graph. 84
4.6 This figure shows the monopole ($l = 0$) and quadrupole ($l = 2$) of the configuration space two-point correlation function for LRGs in the $0.4 < z < 0.6$ redshift range for various emulated runs. The deviation of each clustering statistic relative to a 'fiducial' one (bottom-right panel with bold border) is also plotted. The blue and gray bands are the 1σ and 2σ intervals, respectively. 87
4.7 This figure shows the monopole ($l = 0$) and quadrupole ($l = 2$) of the configuration space two-point correlation function for LRGs in the $0.6 < z < 0.8$ redshift range for various emulated runs. The deviation of each clustering statistic relative to a 'fiducial' one (bottom-right panel with bold border) is also plotted. The blue and gray bands are the 1σ and 2σ intervals, respectively. 88
4.8 This figure shows the monopole ($l = 0$) and quadrupole ($l = 2$) of the configuration space two-point correlation function for LRGs in the $0.8 < z < 1.1$ redshift range for various emulated runs. The deviation of each clustering statistic relative to a 'fiducial' one (bottom-right panel with bold border) is also plotted. The blue and gray bands are the 1σ and 2σ intervals, respectively. 89
4.9 This figure shows the monopole ($l = 0$) and quadrupole ($l = 2$) of the configuration space two-point correlation function for QSOs in the $0.8 < z < 2.1$ redshift range for various emulated runs. The deviation of each clustering statistic relative to a 'fiducial' one (bottom-right panel with bold border) is also plotted. The blue and gray bands are the 1σ and 2σ intervals, respectively. 90
4.10 The solid histograms are the distribution of $\alpha_{\rm iso}$ from fitting 250 LRG `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix. 91

4.11 The solid histograms are the distribution of $\alpha_{\rm AP}$ from fitting 250 LRG `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix. 92
4.12 The solid histograms are the distribution of *df* from fitting 250 LRG `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix. 93
4.13 The solid histograms are the distribution of $\sigma_{\alpha_{\rm iso}}$ from fitting 250 LRG `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix. 94
4.14 The solid histograms are the distribution of $\sigma_{\alpha_{\rm AP}}$ from fitting 250 LRG `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix. 95
4.15 The solid histograms are the distribution of σ_{df} from fitting 250 LRG `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix. 96
4.16 The solid histograms are the distribution of $\alpha_{\rm iso}$ from fitting 250 QSO `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix. 97
4.17 The solid histograms are the distribution of $\alpha_{\rm AP}$ from fitting 250 QSO `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix. 98
4.18 The solid histograms are the distribution of *df* from fitting 250 QSO `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix. 99
4.19 The solid histograms are the distribution of $\sigma_{\alpha_{\rm iso}}$ from fitting 250 QSO `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix. 100
4.20 The solid histograms are the distribution of $\sigma_{\alpha_{\rm AP}}$ from fitting 250 QSO `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix. 101
4.21 The solid histograms are the distribution of σ_{df} from fitting 250 QSO `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix. 102

5.1 Measured and best-fit monopole and quadrupole from DESI DR1 measurements. Symbols display the measurements and solid lines the best-fitting power spectrum model to each of the individual redshift bins according to the DESI full modeling baseline model (see Table 10 in [2]). The errors are 1σ and include both the statistical and systematic error budget. The bottom sub-panels display the residuals between data and model in units of the diagonal error. Caption adapted and figure reproduced from [2]. 105
5.2 Top row: DESI measurements of the BAO distance scales at different redshifts, parametrized as (left) the ratio of the angle-averaged distance $D_V \equiv (zD_M^2 D_H)^{1/3}$ to the sound horizon at the baryon drag epoch, r_d, and (right) the ratio of transverse and line-of-sight comoving distances D_M/D_H, from all tracers and redshift bins as labeled. Caption adapted from and figure reproduced from [3]. 108
5.3 *Left panel:* 68% and 95% credible-interval contours for Ω_m and $r_d h$ for a flat ΛCDM cosmology from fits to BAO data from each DESI tracer individually. The results from all tracers are consistent with each other and the change in the degeneracy direction comes from the different effective redshifts of the samples. *Right panel:* The corresponding results for fits to data from the combined tracers from DESI, supplemented by information from external datasets. Caption adapted from and figure reproduced from [3]. 109
5.4 *Left panel:* Marginalized posterior constraints on Ω_M and H_0 from combining DESI BAO data with external data used to calibrate r_d in a flat ΛCDM model. *Right panel:* The marginalized 1D posteriors on Ω_m in a flat ΛCDM model from different probes, as labeled. Caption adapted from and figure reproduced from [3]. 110
5.5 *Left panel:* 68% and 95% marginalized posterior constraints in the $w_0 - w_a$ plane for the flat $w_0 w_a$CDM cosmology. *Right panel:* 68% and 95% marginalized posterior constraints from DESI BAO combined with CMB and different SN datasets. The significance of the tension with ΛCDM is estimated to be 2.5σ, 3.5σ and 3.9σ, respectively, with the addition of PantheonPlus, Union3 and DESY5 SN Ia datasets. Caption adapted from and figure reproduced from [3]. 112
5.6 68% and 95% credible intervals from the combined DESI full shape and BAO analysis, assuming ΛCDM. Caption adapted and figure reproduced from [4]. 115

5.7 Constraints on w_0 and w_a assuming a w_0w_aCDM model with a time-varying dark energy equation of state parametrization. The contours represent the 68% and 95% credible intervals. The solid blue, orange and green contours represent the combination of DESI (FS+BAO) and CMB with three respective SN Ia data sets: PantheonPlus, Union3 and DES-SN5YR. The dashed blue, orange and green contours show the same respective combinations, but with the DESI full shape clustering and BAO replaced by DESI (BAO). The figure shows how the addition of the full shape information to the BAO-only data improves the precision of the constraints. The measurements of these two parameters remain mutually consistent, and prefer $w_0 > -1$ and $w_a < 0$. Caption and figure reproduced from [4]. 117

LIST OF TABLES

Table

2.1 Summary of the three main approaches for extracting cosmological information from galaxy surveys. Table adapted from [16]. 41
3.1 The total number of galaxies measured from a sample Abacus mock. The 'complete' sample is before fiber assignment. The 'fraction assigned' refers to the ratio between the counts of galaxies assigned and the total counts of galaxies in the complete sample. Table from [10]. 64
4.1 The numbers of assigned counts for different tracers in the data catalogs that were used to train the FFA emulator. 71
4.2 The parameter combinations for LRG emulated runs used in our analysis. 73
4.3 The parameter combinations for QSO emulated runs used in our analysis. 73
4.4 Normed trace comparisons between different LRG covariance matrices. 'altMTL' and 'Data' refer to analytical covariance matrices for Abacus altMTL mocks and data, respectively. 'EZ' refers to the mock-based covariance from 1000 `EZmocks` and FFA is the FitCov covariance for each emulation run specified in the first column. 'FFA-fid' refers to the comparison between each FFA FitCov covariance matrix with the FitCov covariance matrix from the fiducial FFA choice. Note this is why the last row is composed of ones. The top row serves as a standard to compare the rest of the covariances to. 76
4.5 Normed trace comparisons between different QSO covariance matrices. 'altMTL' and 'Data' refer to analytical covariance matrices for Abacus altMTL mocks and data, respectively. 'EZ' refers to the mock-based covariance from 1000 `EZmocks` and FFA is the FitCov covariance for each emulation run specified in the first column. 'FFA-fid' refers to the comparison between each FFA FitCov covariance matrix with the FitCov covariance matrix from the fiducial FFA choice. Note this is why the last row is composed of one(s). The top row serves as a standard to compare the rest of the covariances to. 77
4.6 Mean deviation of LRG fit parameters from 250 `EZmocks`, using different FitCov covariance matrices. 78
4.7 Mean deviation of QSO fit parameters from 250 `EZmocks`, using different FitCov covariance matrices. 79

4.8 Mean deviation of LRG fit errors from 250 `EZmocks`, using different FitCov covariance matrices. 82
4.9 Mean deviation of LRG fit errors from 250 `EZmocks`, using different FitCov covariance matrices, relative to the FitCov covariance matrix obtained from the fiducial FFA choice. Note this is why all the entries for the last (fiducial) iteration are zero. 85
4.10 Mean deviation of QSO fit errors from 250 `EZmocks`, using different FitCov covariance matrices. 86
4.11 Mean deviation of LRG fits to α_{AP} and the corresponding errors on it from 250 `EZmocks`, using different FitCov covariance matrices. 86
5.1 Parameters and priors used in the DESI Y1 BAO analysis. The 'background-only' parametrization refers to using BAO and SN data only. 'CMB' refers to data from *Planck* and ACT being involved. Table reproduced from [3]. 107
5.2 Statistics for the DESI samples used for the DESI DR1 BAO measurements used for the BAO analysis. For each tracer and redshift range, N_{tracer} is the number of objects. The angular diameter distance D_{M} and the Hubble distance D_{H} are presented in units of the sound horizon at the drag epoch. When these two are measured, they are correlated with coefficient r. Otherwise, D_{V}/r_d is the measured quantity. Table adapted from [3]. 108
5.3 Parameters and priors used in DESI Y1 Full Shape analysis. $\mathcal{U}$ refers to a uniform prior within the given range, and $\mathcal{N}$ refers to a Gaussian normal prior with the given mean and standard deviation squared. Table adapted from [4]. 114

ABSTRACT

The data-driven science of modern cosmology relies on galaxy surveys to source the millions of data points that turn targets in the sky into numbers that hold valuable information. The Dark Energy Spectroscopic Instrument (DESI), with 5000 optical fibers, is one such survey with the overarching goal of a better understanding of dark energy, the mysterious component that drives the accelerated expansion of the universe. An understanding of the various systematics that form the DESI pipeline is crucial for a better understanding of the data itself and for the ability to generate mock data catalogs with survey realism. One such step is fiber-assignment, the process of selecting which targets on the sky are observed by the instrument, especially when fiber collisions are present. When not properly accounted for, this suppresses the observed clustering amplitude.

One way to simulate fiber-assignment, which is a complex and computationally expensive process, is to use an emulator, an algorithm which learns from reference datasets with the purpose of being able to emulate the effects of a process approximately but quickly. We specifically discuss the Fast Fiber Assign (FFA) algorithm which was employed by DESI to obtain clustering covariance matrices for the Data Release 1 (DR1) analysis. It was observed that the emulated covariances were not always a perfect match with analytical methods and that this could potentially be ascribed to the FFA process. Thus we especially focus on studying how tuning the emulator affects the outputs, particularly when looking at the effects on covariance matrices.

We find that small adjustments to the learned kernel can be beneficial. This kernel, which is a function of the counts of close neighbors a galaxy on the sky has and how

many times it could possibly be observed by the instrument, can be smoothed to reduce noise. Moreover, we find that the cosmological fitting constraints tend to be sensitive to the selection of galaxies on small scales within a 'collision window'. This corresponds to the minimum separation two DESI fibers can be placed at. Using these results, we recommend improved sets of emulator parameters for the next such DESI analysis.

CHAPTER 1

Introduction

1.1 The Universe — Homogeneous and Isotropic

Under the vast and silent theater of the night sky, humanity's gaze has forever been drawn upward, tracing patterns among the stars and pondering the infinite expanse beyond. The players in this theater have varied in scale and character: from minuscule (to the eye) twinkling stars to the Sun and Moon that dominate the sky at their respective times of day to vast constellations that different cultures have reflected their own stories in. From the earliest star charts to modern day astrophysics, astronomy has been a way for humanity to connect with the universe and seek out some reason, predictability and knowledge about questions pertaining to our place in everything, where we came from and where we are going.

Within this theater, modern cosmology is the big picture — the study of the whole universe at its largest scales, its actors massive structures such as galaxies, clusters, superclusters, and the dark matter and dark energy that comprise it all. The start of cosmology as a modern science can arguably be traced back to Albert Einstein's publication of the theory of general relativity which prompted the development of the first modern models in cosmology [36]. Among these, Alexander Friedmann suggested the possibility of an expanding universe [42] with the subsequent introduction of the Big Bang model by Georges Lemaître [57]. Edwin Hubble's discovery, building on the work

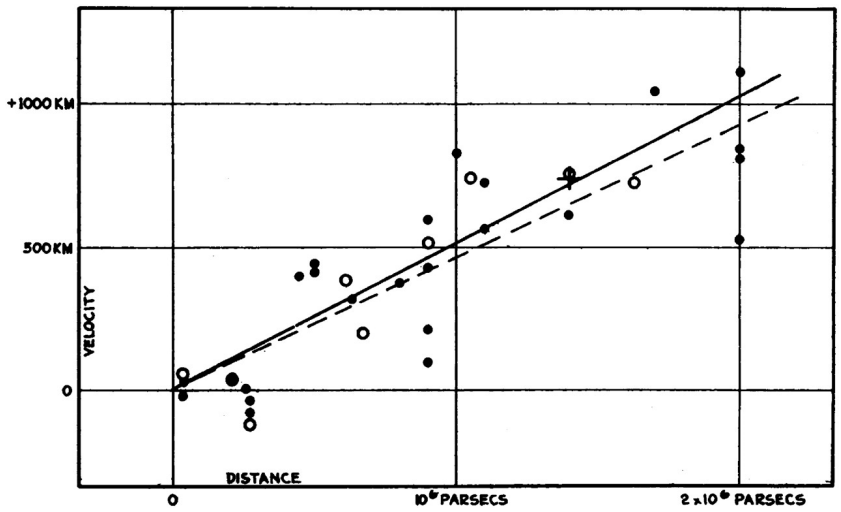


Figure 1.1: Hubble's velocity-distance diagram for galaxies. Radial velocities, corrected for solar motion, are plotted against distances estimated from stars and mean luminosities of nebulae in a cluster. Figure from [50].

of Vesto Slipher [92], of the expanding universe provided further evidence in favor of this paradigm [50]. Taking the field much farther than the 46 galaxies used by Hubble in his analysis, the COBE experiment's discovery of the Cosmic Microwave Background's (CMB) anisotropies [94] sparked the era of modern observational cosmology — a revolution for a field that was infamous for its massive error bars (for many decades after Hubble, the Hubble constant's value varied by a factor of two on either end of its range).

All the above cosmological observations are dependent on receiving light (across its spectrum) from sources far out in space. As spacetime itself is stretched by the universe's expansion, the wavelengths propagating through it are also extended, leading to a phenomenon called redshift, denoted by z:

$$z \equiv \frac{\lambda_0 - \lambda_e}{\lambda_e} \tag{1.1}$$

where λ_0 and λ_e are the received and emitted wavelengths, respectively. Note that we will use the term 'redshift of an object' to refer to the redshift of the light received from it. The greater the redshift of an object, the more space its light has traveled through,

meaning the object is farther away. Hence, the greater this redshift, the farther 'into the past' we are looking. Redshift, therefore, will often be used as a measure of time.

The mathematics of cosmology today is anchored in a few simple statements, primarily Einstein's equivalence principle, and the cosmological principle. The former leads to the development of a curved spacetime that appears locally flat. The latter is the assumption that the universe is both homogeneous and isotropic. Homogeneity means that the universe looks the same from every point in space and isotropy refers to the universe appearing the same in every direction from a given observation point. To the layperson, both these statements appear to be 'disproved' by looking up at the sky. And while it is true that the universe can possess lots of structure on small scales, appearing heterogeneous, on much larger scales, this is washed out and everything should follow the cosmological principle.

It was proven by Robertson [85] and Walker [99] that the following metric, used by Friedmann [42] and Lemaître [57], is the most general metric consistent with Einstein's general relativity and the cosmological principle for a spacetime with uniform curvature:

$$\begin{aligned} ds^2 &= -dt^2 + a^2(t)\Big[\frac{dr^2}{1-\kappa r^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2)\Big] \\ &= -dt^2 + a^2(t)\Big[\frac{dr^2}{1-\kappa r^2} + r^2 d\Omega^2\Big] \end{aligned} \tag{1.2}$$

where $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$. This is called the Friedmann-Lemaître-Robertson-Walker (FLRW) metric. Here, $a(t)$ is the scale factor which traces the expansion of space, and we are working with the convention that the speed of light $c = 1$. The scale factor comes from the fact that as the universe expands, the distances between individual points in space grow. Conventionally, a is taken to be zero at the Big Bang and $a(t_0) = 1$ where t_0 is the time today. κ is a dimensionless constant whose values

represent different kinds of spacetime curvature:

$$\kappa = \begin{cases} +1 & \text{positive curvature, closed (elliptic) universe} \\ 0 & \text{zero curvature, flat (Euclidean) universe} \\ -1 & \text{negative curvature, open (hyperbolic) universe} \end{cases}$$

Euclidean geometry is the 'familiar' one — the internal angles of a triangle add up to π, and lines parallel at a point always stay parallel. In an elliptic geometry, the internal angles of a triangle add up to be greater than π, and lines parallel at a point eventually intersect. In a hyperbolic geometry, the internal angles of a triangle add up to be less than π, and lines that are parallel at a point will diverge.

1.1.1 Distances in Cosmology

Now that we have a metric, it is natural to consider distances. The 'physical distance' between two objects or points is the 'real' distance between them. That is, as spacetime expands, this physical distance grows as the objects move apart. But if these two objects have no peculiar velocities (velocities because of reasons beyond the expansion, such as from gravitational attraction between structures), then we can imagine them to be at a constant distance from each other if we have a coordinate system that stretches along with the aforementioned expansion. This distance that 'ignores' the expansion is called the comoving distance d_c and can be related to the physical distance d_{ph} via

$$d_{ph} = a(t) d_c \tag{1.3}$$

Given that light propagates on null geodesics, we can integrate the line element of the FLRW metric (Equation 1.2) to get the proper distance d_p (which is the physical

distance at a given time):

$$d_p = a(t) \int_0^r \frac{dr'}{\sqrt{1 - \kappa r'^2}} \tag{1.4}$$

We define the distance that light covers as it travels from time t to the present time t_0 to be the coordinate distance χ:

$$\chi(t) = \int_t^{t_0} \frac{dt'}{a(t')}$$

With this, we can express the comoving distance as:

$$d_c = \begin{cases} R_0 \sin(\chi/R_0) & \text{(positive curvature)} \\ \chi & \text{(no curvature)} \\ R_0 \sinh(\chi/R_0) & \text{(negative curvature)} \end{cases} \tag{1.5}$$

Also, with the scale factor in hand, we have the following relationship, which helps connect distances and lengths to redshifts:

$$1 + z = \frac{1}{a}$$

It is vital to note that cosmologists do not actually observe or measure proper distances — this would, after all, be a technical challenge involving holding the universe constant for long enough to be able to make a measurement (the issue of getting an instrument out to measure this distance another difficulty). What one can measure are the luminosity and angular-diameter distances.

The luminosity distance d_L is given by

$$d_L(z) = (1 + z) d_c(z) \tag{1.6}$$

and is inferred from the observed flux f of an object with luminosity L, where

$$f = \frac{L}{4\pi d_L^2}$$

Luminosity is a measure of the radiated electromagnetic energy per unit time. The angular-diameter distance d_A is inferred from the observed angular extent $\Delta\theta$ of an object with proper size Δx. This can be related to the comoving distance as

$$d_A(z) = \frac{1}{1+z} d_c(z) \tag{1.7}$$

Finally, we introduce the particle horizon distance. This is the maximum possible distance from which light could reach an observer in the 'current' lifetime of the universe, starting from the Big Bang at time $t = 0$ to time t_0 today. This is given by

$$d_{\text{part hor}} = \int_0^{t_0} \frac{dt}{a(t)} \tag{1.8}$$

1.1.2 The Friedmann Equations

Before introducing the equations that govern the expansion of a homogeneous and isotropic universe, we pause to introduce the Hubble parameter. Hubble put forward that the recession velocities v of galaxies (the rate at which they move away from us) are proportional to their distances d from us [50]:

$$v = H_0 d \tag{1.9}$$

This statement is called Hubble's law and H_0 is the Hubble constant, the value of which is of extreme interest to cosmologists. If we extend this idea to any time in the

universe's history, we get:

$$\begin{aligned} v(t) &= \frac{d}{dt}d(t) \\ &= H(t)d(t) \end{aligned}$$

where $d(t)$ is the physical distance between the observer and galaxy (Equation 1.3), and we have defined

$$H(t) \equiv H \equiv \frac{\dot{a}(t)}{a(t)} \tag{1.10}$$

to be the Hubble parameter.

Solving the Einstein field equations with the FLRW metric and cosmological principle yields the first Friedmann equation:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} \tag{1.11}$$

where $k = \frac{\kappa}{R_0^2}$ is a dimensionless curvature parameter (and R_0 is the radius of curvature of space). Moreover, if we model the universe's contents as a perfect fluid defined by its isotropic pressure P and energy density ρ, we can use the conservation of energy-momentum in general relativity along with the cosmological principle to derive the continuity equation:

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + P) = 0 \tag{1.12}$$

This allows us to predict how the energy density of the components of the universe evolves. By combining Equations 1.11 and 1.12, we can get the second Friedmann

equation (often called the acceleration equation):

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) \tag{1.13}$$

From Equations 1.10 and 1.13, we deduce that there exists a critical density ρ_{crit} at which the curvature of the universe (parameterized by k) must be zero:

$$\rho_{\text{crit}} = \frac{3H^2}{8\pi G} \tag{1.14}$$

Using the subscript 0 to denote present-day quantities (a convention we will use throughout this dissertation), we can define $\rho_{\text{crit, 0}}$ to be the present-day critical density. We now introduce the (relative) critical density parameter $\Omega(t)$:

$$\Omega(t) = \frac{\rho(t)}{\rho_{\text{crit}}(t)} \tag{1.15}$$

We can use this to cast Equation 1.11 into the form:

$$1 - \Omega(t) = -\frac{\kappa}{R_0^2 a^2(t) H^2(t)} \tag{1.16}$$

Given that mainstream models do not predict a change in the sign of the curvature κ, the right hand side of Equation 1.16 cannot change sign. Therefore, we infer that if $\Omega(t) = 1$ at any point in time, it will always be so. And if it is either greater or smaller than 1 at any point, it will always stay so, respectively. The most up to date measurements of Ω_0 (also often just denoted Ω) point towards a flat universe.

1.1.3 Components of the Universe

Before we can solve the Friedmann and continuity equations, we need to be able to relate pressure and energy density. For a perfect fluid i, we have the equation of state

$$P_i = w_i \rho_i$$

where w_i is the equation of state parameter. Solving Equation 1.12 for a constant w gives us

$$\rho(a) = \rho_0 a^{-3(1+w)} \tag{1.17}$$

For non-relativistic matter, $w_m = 0$ and we get the scaling relation $\rho_m \propto a^{-3}$, representing the dilution of matter density with the volume of an expanding universe. For radiation (relativistic particles such as photons), $w_r = 1/3$, giving us $\rho_r \propto a^{-4}$. The additional factor of a^{-1} on top of the volume dilution from expansion comes from the stretching of the radiation's wavelength. Finally, for the cosmological constant Λ, $w_\Lambda = -1$ and ρ_Λ is constant. Λ represents the energy of the vacuum itself and this constant energy density naturally follows because as the universe expands and more vacuum is 'created', there is a proportional increase in vacuum energy, leading to a constant energy density.

Today, the concordance model in cosmology is ΛCDM (CDM referring to Cold Dark Matter). Its major components are:

- **Matter:** This is 'ordinary matter', such as protons and neutrons, and it constitutes roughly 5% of the total composition of the universe.
- **Radiation:** Radiation consists of relativistic particles such as photons and, in the early universe, relativistic neutrinos. Radiation is a tiny fraction of all the components of the universe today.

- **Dark Matter:** Dark matter is matter that does not interact electromagnetically and thus cannot be seen directly. It either interacts with other objects only through gravitation or else it interacts so weakly that the interactions have not been observed. Evidence for dark matter comes from a variety of sources, including galaxy rotation curves, gravitational lensing and its influence on large scale structure [111]. 'Cold' dark matter (CDM) refers to its non-relativistic nature, which allowed CDM to clump together even in the early times of the universe when baryonic matter was still interacting with radiation. Dark matter comprises around 25% of the total content and around 80% of the matter content in the universe.

- **Dark Energy:** Dark energy is the unidentified component of the universe that is believed to drive the expansion of the universe [84, 78]. It is denoted by Λ, the symbol for Einstein's cosmological constant. Considering it the intrinsic energy of the vacuum is just one of the hypotheses for its nature. In fact, looking for deviations from $w = -1$, the equation of state parameter for the cosmological constant is one of the biggest quests in the search for the nature of dark energy. Dark energy comprises approximately 70% of the total content in the universe.

1.2 The Universe — Not so Homogeneous

The standard Big Bang model that we have discussed so far is, at best, incomplete. There are three major problems that we list below. All of them represent measurements that appear all too coincidental to have occurred without a mechanism behind them:

- **The Flatness Problem:** As mentioned before, the total energy density of the universe at the present day has been measured to be very close to the critical density [5]: $\Omega_0 = 0.999 \pm 0.002$. Moreover, any deviation from $\Omega_0 = 1$ scales as a power law in time. This means that any deviation from the critical density today

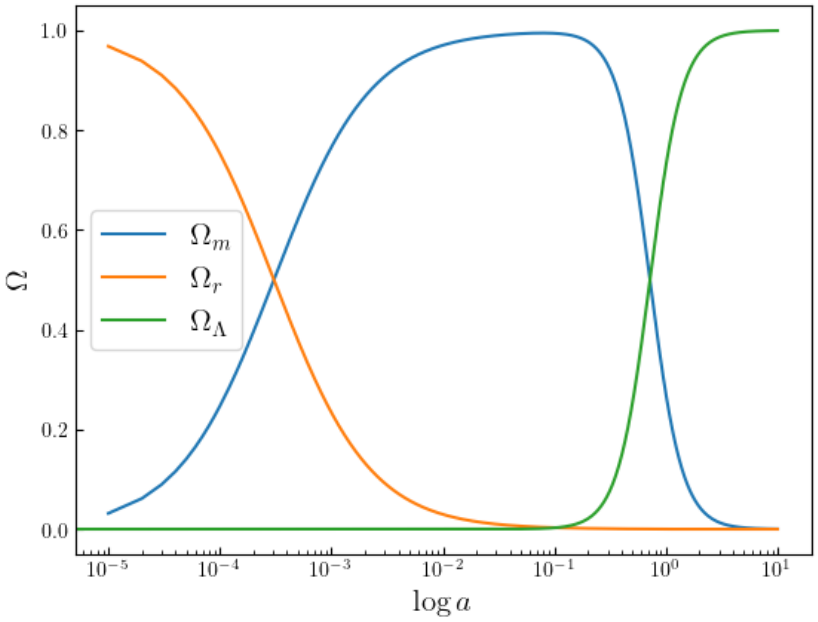


Figure 1.2: This figure represents the evolution of the densities of different components of a ΛCDM universe. The horizontal axis is the logarithm of the scale factor, here used as a measure of the passage of time. The vertical axis is the relative density parameter from Equation 1.15.

would have been smaller in the past. So for the universe to be so flat today, it would have to be even more so in the past. The flatness problem is thus the quest for the reason behind what appears to be a very finely tuned number.

- **The Horizon Problem:** On the largest scales, the universe is very homogeneous. A powerful example of this is the CMB which is uniform to one part in 100,000 across the sky. Such homogeneity would imply that different parts of the sky were once in causal contact with each other as opposed to this being an unlikely coincidence. But as the horizon grows, we find that regions that are not 'supposed' to have ever been in causal contact in the past still show remarkable uniformity in temperature. Perhaps there was a mechanism that allowed for causal contact in the far past between seemingly disconnected regions of the universe.

- **The Magnetic Monopole Problem:** Magnetic monopoles are theoretical constructs that act as a single magnetic charge (as opposed to a magnetic dipole). A

magnetic monopole has never been observed but they are predicted by a number of Grand Unified Theories. If they exist, we require a paradigm that explains why they have never been observed.

1.2.1 Cosmic Inflation

A solution proposed to these problems is cosmic inflation [46]. Inflation describes a rapid exponential expansion of the universe during its early times, around 10^{-36} to 10^{-32} seconds after the Big Bang. In a universe influenced mainly by a fluid with equation of state parameter $w < -1/3$, we get accelerated cosmic expansion.

Inflation solves the flatness problem by providing a mechanism where a very rapid expansion of space (of at least a certain duration) makes the universe appear locally flat even if it starts out with some curvature. This is akin to taking a ball and expanding its size to, say, planetary scales where any patch appears flat to a local observer. For the horizon problem, inflation could be the reason why parts of the universe that did indeed start out in causal contact were 'inflated away' from each other much faster than they would have in a non-inflation Big Bang cosmology — thus appearing as if they were never in causal contact before as they re-enter the horizon at late times. Finally, for the magnetic monopole problem, inflation dilutes the density of monopoles, making them so difficult to find.

1.2.2 Large Scale Structure

Looking at a map of galaxies (such as in Figure 1.3) makes it clear that galaxies are not located at random positions in the sky but instead form filamentary structures. To be able to explain the origin of and describe the distribution of this structure, we lay out some of the required formalism. At the heart of this is the concept of an overdensity — an excess mass density compared to the mean density. At a given location $\vec{x}$, taking $\bar{\rho}(t)$ to be the mean density over all space, this real space (also referred to as configuration

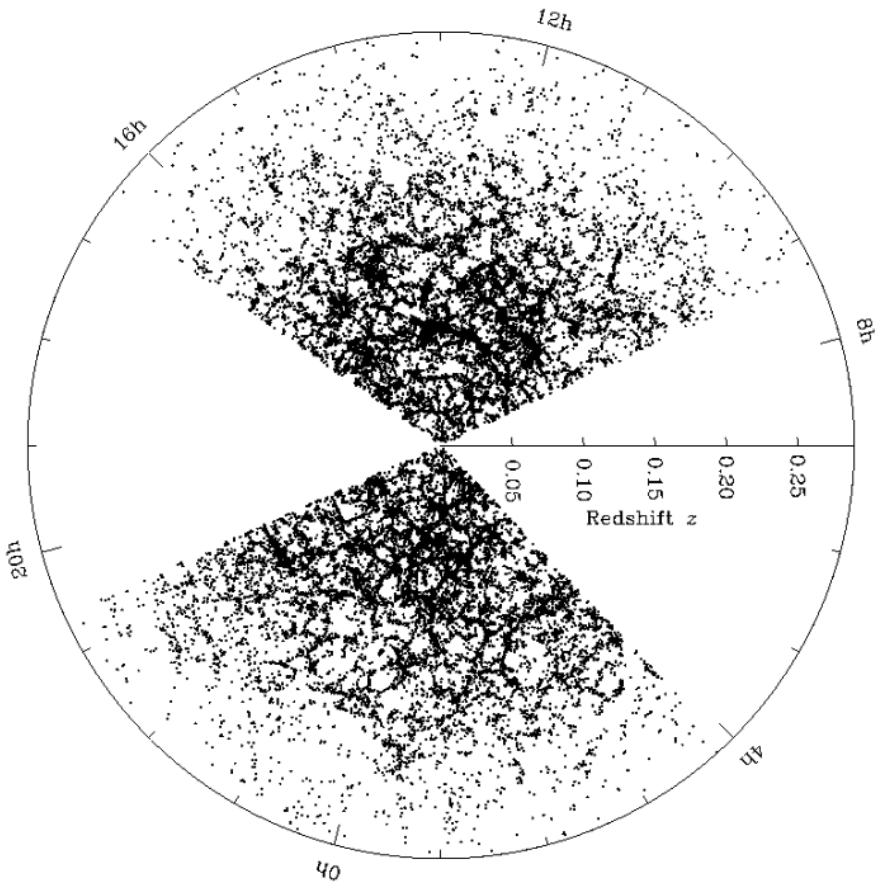


Figure 1.3: The distribution of galaxies in the Sloan Digital Sky Survey (SDSS) main galaxy sample, showing galaxies within 1.25 degrees of the Celestial Equator. Figure from [106].

or position space) overdensity $\delta(\vec{x}, t)$ is defined to be

$$\delta(\vec{x}, t) \equiv \frac{\rho(\vec{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)}$$

These overdensities are seeded by the inflation discussed earlier and are predicted to be approximately Gaussian. Because of the definition of the overdensity, as fluctuations grow with time, they eventually enter a non-Gaussian regime when $\delta \gtrsim 1$. We confine ourselves to the Gaussian regime for mathematical tractability. Also, we will find it helpful to view the overdensity in Fourier space because each Fourier mode $\delta_{\mathbf{k}}$ evolves

independently:

$$\delta_{\mathbf{k}}(t) = \frac{1}{\sqrt{V}} \int \delta(\mathbf{r}, t) \exp\left(-i\mathbf{k} \cdot \mathbf{r}\right) d^3\mathbf{r}$$

The evolution of fluctuations can be shown to follow:

$$\frac{\partial^2 \delta}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial \delta}{\partial t} = \left(4\pi G\bar{\rho} + \frac{c_s^2}{a^2}\nabla^2\right)\delta + \frac{2T}{3a^2}\nabla^2 S \tag{1.18}$$

in real space, where T is the temperature, S is the entropy of the system and c_s is the speed of sound ($c_s = \sqrt{\frac{d\bar{P}}{d\bar{\rho}}}$). In Fourier space, this becomes:

$$\frac{\partial^2 \delta_{\mathbf{k}}}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial \delta_{\mathbf{k}}}{\partial t} = \left(4\pi G\bar{\rho} + \frac{k^2 c_s^2}{a^2}\right)\delta_{\mathbf{k}} - \frac{2Tk^2}{3a^2} S_{\mathbf{k}}$$

Under adiabatic conditions, the S terms vanish. Note that at a particular scale $k = \sqrt{4\pi G\bar{\rho}}(a/c_s) \equiv k_J$ (called the Jeans scale), we have a boundary for the behavior of perturbations. On small scales (where $k > k_J$), we have oscillating solutions and so perturbations do not grow. On scales larger than the Jeans wavenumber, we have growth of perturbations.

In general, solving the differential equations for the growth of perturbations requires numerical techniques rather than deriving analytic solutions. One parameter that is used to reduce the above equations to dimensionless form for numerical solving is the linear growth function D, which we introduce for the sake of its connection to other parameters that we can measure:

$$D(a) = \frac{\delta(a)}{\delta(1)} \tag{1.19}$$

From here, we can define the growth rate f of structure as

$$f = \frac{d \ln D}{d \ln a} \tag{1.20}$$

We will return to this quantity later in our analysis.

1.2.2.1 Quantifying Structure

We now lay out the primary method of quantifying structure through the measurement of galaxy clustering. Let n be the mean density of galaxies in space. The probability of finding a galaxy in volume dV is given by

$$dP = n\, dV$$

Then the probability of finding a galaxy in volume dV_1 within a distance r_{12} of another galaxy in volume dV_2 is

$$dP = n^2(1 + \xi(r_{12}))dV_1 dV_2 \tag{1.21}$$

where ξ represents the extra probability of finding another galaxy within a given distance of a galaxy. This quantifies the concept of correlations between galaxies — if we find one galaxy, what are the increased odds that we find another one within some distance? That is, how much do galaxies cluster? In a continuous density field $\rho(\mathbf{x})$, the two-point correlation function is defined as

$$\begin{aligned} \xi(r) &\equiv \langle \delta(\mathbf{x} + \mathbf{r}) \delta(\mathbf{x}) \rangle_{\mathbf{x}} \\ &= \frac{\langle [\rho(\mathbf{x} + \mathbf{r}) - \langle \rho \rangle][\rho(\mathbf{x}) - \langle \rho \rangle] \rangle_{\mathbf{x}}}{\langle \rho \rangle^2} \end{aligned}$$

An example of the configuration space two-point correlation function is given in

Figure 1.7. In Fourier space, with comoving volume V, the components of the overdensity are

$$\delta_{\mathbf{k}} = \frac{1}{\sqrt{V}} \int \delta(\mathbf{r}) \exp\left(i\mathbf{k} \cdot \mathbf{r}\right) d^3r$$

and the Fourier transform of the two-point correlation function is called the power spectrum $P(k)$:

$$\langle \delta_{\mathbf{k}} \delta^*_{\mathbf{k}'} \rangle = P(k)(2\pi)^3 \delta^{(3)}(\mathbf{k} - \mathbf{k}')$$

where $\delta^{(3)}$ is a Dirac delta function. These two clustering statistics are essential components for measuring and fitting models of galaxy clustering in major surveys [6, 7].

To be able to empirically measure this clustering, in particular the two-point correlation function, we need an unbiased (the estimator's expected value should be equal to the parameter's true value) and minimum variance estimator. For practical purposes in this project, we use the popular Landy-Szalay estimator [55]:

$$\hat{\xi}(r) = \left(\frac{N_r}{N_d}\right)^2 \frac{DD(r)}{RR(r)} - 2\frac{N_r}{N_d}\frac{DR(r)}{RR(r)} + 1 \tag{1.22}$$

where $DD(r)$ is the number count of pairs in the data catalog within the interval $r \pm \Delta r$. $RR(r)$ is the corresponding number count of pairs in a randomly distributed catalog, and $DR(r)$ represents cross-pair (between data and randoms) counts. N_r and N_d are the numbers of objects in the random and data catalogs, respectively. For completeness, we note that we can rewrite Equation 1.22 using the corresponding normalized pair counts for each term to get

$$\hat{\xi}(r) = \frac{DD(r)}{RR(r)} - 2\frac{DR(r)}{RR(r)} + 1 \tag{1.23}$$

where we have absorbed the normalization terms into the DD, RR and DR counts. This is purely to match convention as this is the form of the estimator that is usually quoted in the literature.

Observe also that an unsmoothed density field has infinite variance — roughly saying that there is lots of structure no matter how far one 'zooms in'. Practically, it is necessary to smooth (or filter or course grain) the theoretical density field before it can be connected to observations. For this, we use window functions W to create weighted averages over regions of space. This corresponds to convolutions in configuration space (which translate to multiplication in Fourier space), giving us the density field at a point $\mathbf{r}$ smoothed over a region $\mathbf{r}'$ around it (scale R):

$$\delta(\mathbf{r}, R) = \int W(|\mathbf{r} - \mathbf{r}'|)\delta(\mathbf{r}')\, d^3\mathbf{r}'$$
$$\delta_{\mathbf{k}}(R) = W(k, R)\delta_{\mathbf{k}}$$

The power spectrum then conveniently transforms as

$$P(k, R) = |W(k, R)|^2 P(k)$$

In cosmology, conventionally, a top-hat window function is used:

$$W_{\mathrm{TH}}(k, R) = \frac{3j_1(kR)}{kR}$$

where $j_1(x)$ is the order one Bessel function.

Using the fact that the two-point correlation function and power spectrum are Fourier transform pairs, we write the zero-lag correlation function (that is, evaluated at zero

separation):

$$\xi(0) = \frac{1}{2\pi^2}\int_0^\infty P(k)\lim_{r\to 0}\frac{\sin(kr)}{kr}k^2\,dk$$
$$= \int_0^\infty \Delta^2(k)\,d\ln k \tag{1.24}$$

where we use the following definition of the dimensionless power spectrum:

$$\Delta^2(k) \equiv \frac{k^3 P(k)}{2\pi^2}$$

With the top-hat window function and renaming $\xi(0)$ smoothed over scales R as $\sigma^2(R)$ to agree with convention, Equation 1.24 reads

$$\sigma^2(R) = \int_0^\infty \Delta^2(k)\Big(\frac{3j_1(kR)}{kR}\Big)^2 d\ln k \tag{1.25}$$

This is the root mean square (rms) amplitude squared of linear mass fluctuations smoothed over scale R. Traditionally, a smoothing scale of $R = 8h^{-1}$ Mpc is used — this is approximately the size of a galaxy cluster:

$$\sigma_8 \equiv \sigma(R = 8h^{-1}) \tag{1.26}$$

The importance of this quantity lies in that it is a way for us to normalize the power spectrum by observing the distribution of galaxies.

Calculating the power spectrum computationally is a vital part of analyzing large scale structure. In standard FLRW cosmology, the dimensionless linear power spectrum of matter density perturbations is given by

$$\Delta^2(k,a) = A_s\frac{4}{25\Omega_m^2}\Big(\frac{k}{k_{\rm piv}}\Big)^{n_s-1}\Big(\frac{k}{H_0}\Big)^4[ag(a)]^2T^2(k) \tag{1.27}$$

where A_s is the amplitude of scalar fluctuations, n_s is the scalar spectral index, and $k_{\rm piv}$ is the pivot wavenumber around which the power spectrum is a power law. $ag(a)$ describes the linear growth of perturbations where $D(a) \equiv ag(a)/g(1)$. $D(a)$ was introduced in Equation 1.19. Finally, $T(k)$ is the linear transfer function.

The transfer function describes the effect of the transition from a radiation-dominated universe to one that was matter-dominated (around $z \approx 3500$). Cosmic inflation predicts that fluctuations enter the horizon with amplitudes independent of their wavelengths. For sub-horizon scales and for matter at late times, in linear theory, Equation 1.18 can be recast as

$$\ddot{\delta} + 2H\dot{\delta} - 4\pi G\rho_m(t)\delta = 0 \tag{1.28}$$

Equation 1.28 can be used to show that fluctuations grow only logarithmically during the radiation-dominated era. So fluctuations that enter the horizon during this epoch have their growth suppressed. These fluctuations were not free to grow quicker until the onset of matter domination. Comparatively, modes entering the horizon during the matter-dominated era did not experience this growth suppression. It is these effects that transfer functions encode. These functions can be deduced from codes that solve the Einstein-Boltzmann equations for density perturbation evolution.

We close this subsection by noting that it is common to decompose the two-point correlation function and the power spectrum into multipoles to make clearer the effects of isotropic and anisotropic clustering. In the case of the two-point correlation function, the multipoles $\xi_l(s)$ are given by

$$\xi_l(s) = \frac{2l+1}{2} \int_{-1}^{1} \xi(s,\mu) L_l(\mu)\, d\mu \tag{1.29}$$

where μ is the cosine of the angle between the line of sight and the separation vector

s and L_l are the Legendre polynomials of order l.

1.2.2.2 Baryon Acoustic Oscillations

We move to study a very crucial feature in the clustering of galaxies. Let us consider an overdensity in the early universe when conditions are hot enough for baryons and radiation (photons) to be coupled together. As matter is drawn into this overdensity, there is increased pressure from the baryon-photon interactions. This contest between gravity and pressure gives rise to sound waves or oscillations that travel spherically outwards at the speed of sound c_s. Once the universe cools and gets to recombination, the mean free path of photons becomes much longer and they can free-stream away, while the baryons are trapped in the potential wells of dark matter. Note that dark matter, being pressureless, did not travel outwards along with the baryons. This phenomenon takes place at the end of the drag epoch ($z_d \approx 1060$). This drag epoch is the time of baryon decoupling and it is slightly later than the time of photon decoupling (around a redshift of 1090) because of the relative number of baryons and photons. As there are many more photons than baryons, the photons stop 'perceiving' the baryons earlier than the opposite and the baryons are still dragged along by the pressure.

An imprint of this pressure wave is left behind (see Figure 1.4), frozen in at the time when baryons and photons decoupled when the spherical pressure wave had radius r_d (called the sound horizon at the drag epoch):

$$r_d = \int_{z_d}^{\infty} \frac{c_s(z)}{H(z)}\, dz$$

This manifests as an increased probability of finding pairs of galaxies at distances corresponding to r_d. This feature is what is called the Baryon Acoustic Oscillations (BAO). It manifests as a peak in the real space two-point correlation function and as oscillations in the Fourier space power spectrum. An example of this is shown in Figure

1.7. The well-understood physics of BAO provides a 'standard ruler' for measuring distances in the universe because cosmologists can predict the BAO scale at the drag epoch.

When measuring the galaxy correlation function, we are essentially observing the angle at which the BAO feature appears from our location. We can also measure clustering in the radial direction when we have accurate spectroscopic redshifts. These can be connected to the transverse (perpendicular relative to the line of sight) separation $s_\perp$ by

$$s_\perp(z) = d_A(z)\theta$$

and to the radial (or parallel) separation via

$$s_\parallel = \frac{\Delta z}{H(z)}$$

For an illustration of this, see Figure 1.5. Note that for the BAO analysis, $s_\perp(z) = s_\parallel(z) = r_d$. Therefore, measuring the BAO gives us insight into the angular-diameter distance and the Hubble parameter, scaled by the sound horizon.

1.2.2.3 Redshift-Space Distortions

When the positions of galaxies are observed in redshift space (that is, say, plotted as a function of their redshifts), their distribution appears distorted. This is generally referred to as redshift-space distortions (RSD) and manifests in two forms, as visualized in Figure 1.6. First, galaxies are affected by the gravitational pull of structures around them. Objects physically close to each other will be pulled by the same large-scale structures, and are therefore more likely to have similar velocities. These are called peculiar velocities and this corresponds to the elongation of structures along the line of sight on small scales. This is often called the Fingers of God effect.

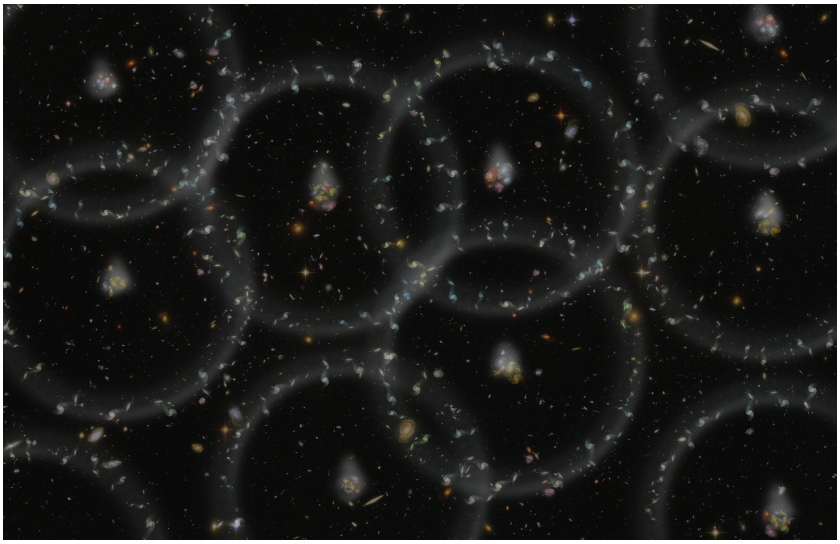


Figure 1.4: An artistic rendering of the baryon acoustic oscillations. Image by Zosia Rostomian and provided by Lawrence Berkeley National Laboratory is licensed under CC BY-NC-ND 4.0. ©2017 The Regents of the University of California, Lawrence Berkeley National Laboratory.

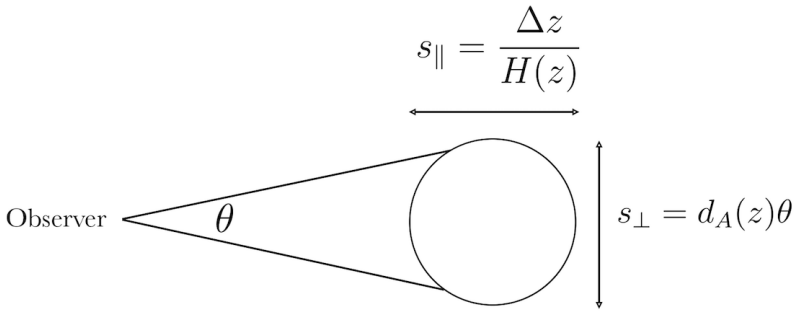


Figure 1.5: A graphic representing the quantities that are physically measured parallel and perpendicular to the line of sight of an observer, in this case, looking at a spherical feature.

The dominant effect is the apparent flattening or squashing of structures along the line of sight. This is called the Kaiser effect [52]. Briefly, if the line of sight velocity of a galaxy in natural units is $v_\parallel$, the observed redshift z_0 is related to the true redshift z

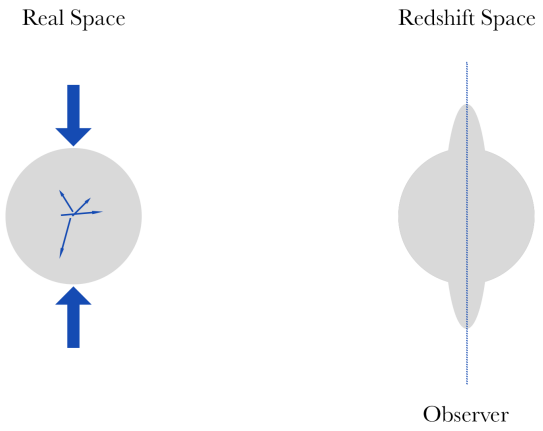


Figure 1.6: A graphic representing the redshift-space distortions. On large scales, velocities are directed 'inward' towards large overdensities — denoted by the wide blue arrows. This manifests in redshift space as the Kaiser effect, where the object appears squashed. On smaller scales, galaxies have random motions (peculiar velocities) that cause an elongation along the line of sight in redshift space. This is often called the Fingers of God effect.

by the Doppler effect formula:

$$1 + z_0 = (1 + z)\sqrt{\frac{1 + v_\parallel}{1 - v_\parallel}}$$
$$\approx (1 + z)(1 + v_\parallel)$$

where the approximation is made in the non-relativistic limit. It can be shown that the overdensities in real space δ and redshift space $\delta^{(s)}$ are related by $\delta^{(s)} = (1 + f\mu^2)\delta$ where μ is the cosine of the angle between the line of sight direction and the wavenumber, and f is the linear growth rate. Thus, RSD effects leave their imprint on the correlation function and power spectrum of the matter distribution.

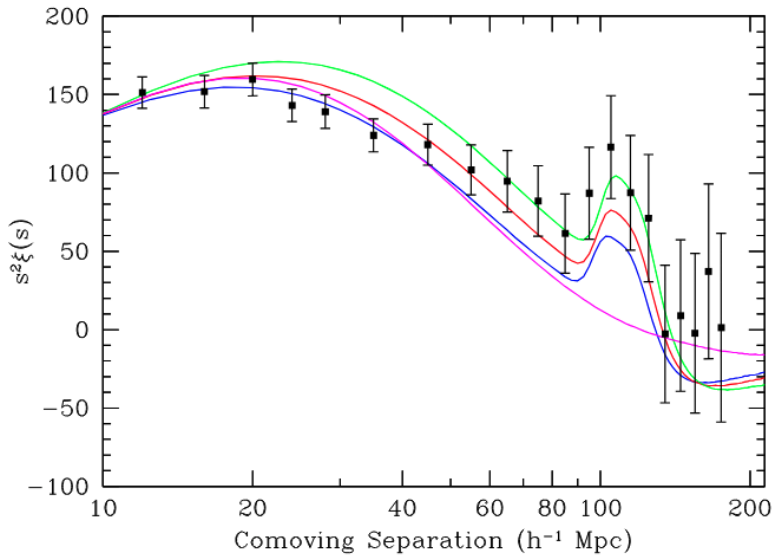


Figure 1.7: The large-scale redshift-space correlation function of the SDSS LRG sample. The colored lines are fits from different models. Note around comoving separation 100 h^{-1} Mpc the peak in the correlation function corresponding to the BAO feature. Figure reproduced from [38].

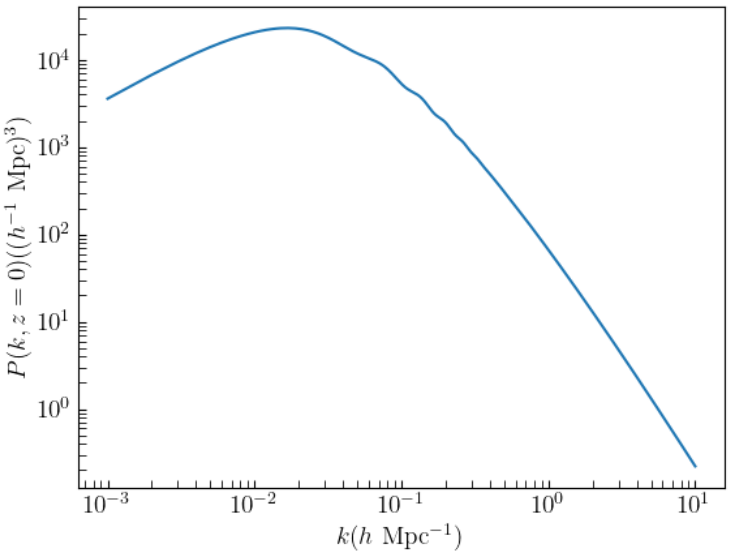


Figure 1.8: An example of the linear matter power spectrum $P(k)$ given a cosmology. Note the power law dependence on either 'extreme' scale. On large scales (small k), the function scales as $k^{n_s} \approx k^1$ and on small scales (large k), it goes as $k^{n_s-4} \approx k^3$.

1.3 The Dark Energy Spectroscopic Instrument (DESI)

As a data-driven and statistics-heavy science, modern cosmology requires huge numbers of measurements. Today, galaxy surveys have become instrumental in our pursuit of understanding the universe. These surveys map the distribution of galaxies across vast stretches of space, providing insights into the underlying cosmological models and the evolution of cosmic structure. With measurements from such surveys, cosmologists can construct three-dimensional maps of the universe and extract the statistics of various processes, such as the previously mentioned BAO feature in the clustering of galaxies.

The Dark Energy Spectroscopic Instrument (DESI) is one such galaxy survey (and instrument) whose overarching goal is to further our understanding of dark energy by greatly improving upon the precision and scope of cosmic mapping. Deployed at the Mayall 4-meter Telescope at Kitt Peak National Observatory in Arizona, DESI aims to provide critical insights into the nature of this component that dominates the energy budget of the universe. Figure 1.9 shows a model of the instrument.

DESI employs 5000 optical fibers [91] that are maneuvered by robotically-controlled fiber positioners that allow for rapid and accurate targeting of objects in the sky [70]. The instrument will capture light from 40 million galaxies and quasars, divided into four primary target classes: Bright Galaxy Survey (BGS) targets, Luminous Red Galaxies (LRGs), Emission Line Galaxies (ELGs), and Quasars (QSOs). LRGs have old stellar populations with low rates of new star formation. ELGs have prominent emission lines in their spectra, indicating the presence of hot, ionized gas and higher rates of star formation. QSOs are extremely luminous active galactic nuclei. Their very high brightness means that they can be used to probe very distant redshifts. Along with these 'primary' targets, there are also secondary and tertiary targets to facilitate science programs whose goals extend beyond the primary DESI ones [32, 34, 88, 103, 41, 95, 33, 98].

This categorization enables a comprehensive examination of the universe across an extensive redshift range. The BGS provides a magnitude-limited sample of about 14

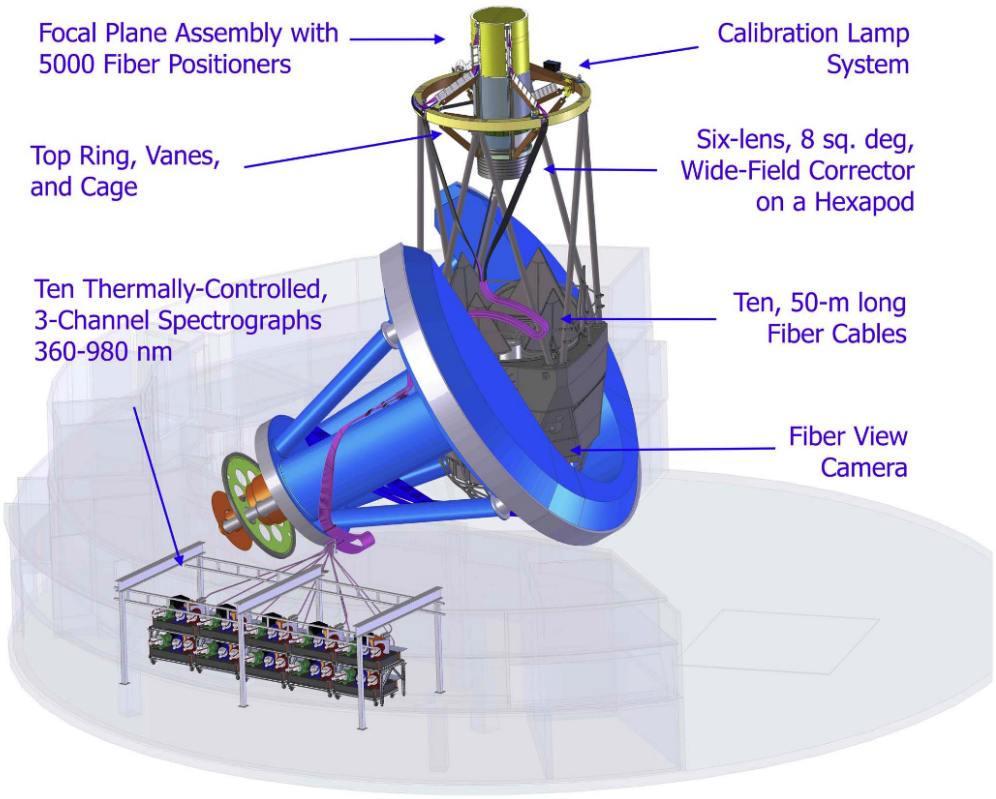


Figure 1.9: A model of the 4 m Mayall Telescope with the instrumentation built for DESI. Figure reproduced from [29].

million galaxies at a median redshift of $z \sim 0.2$, while the LRGs extend to $z = 1$, employing absorption lines for their identification. The ELG sample extends to $z = 1.6$, relying on the detection of the OII emission-line doublet, and the QSO selection reaches even higher redshifts, probing the universe beyond $z = 2.1$ through the Lyman-alpha forest. The Lyman-alpha forest refers to a series of absorption lines that appear in the spectra of light originating from very distant sources [67, 93]. This light passes through various clouds of gas at different redshifts on its way to us and this creates absorption lines at different wavelengths. For cosmological inference, DESI divides the various targets into the following bins:

- BGS: $z \in (0.1, 0.4)$
- ELG: $z \in (0.8, 1.1)$, $z \in (1.1, 1.6)$

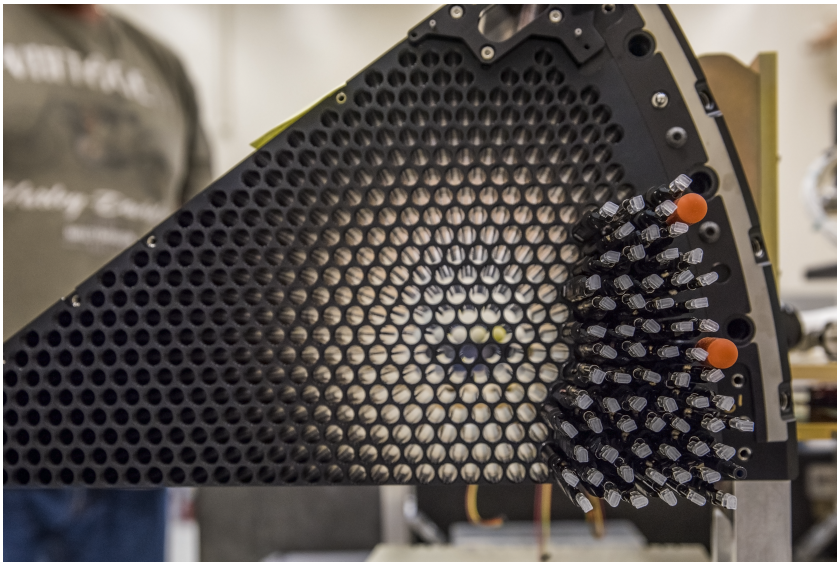


Figure 1.10: A view of a partially assembled focal plane petal with an array of robotic positioners that is each connected to a fiber-optic cable. Photo by Marilyn Chung and provided by Lawrence Berkeley National Laboratory is licensed under CC BY-NC-ND 4.0. ©2017 The Regents of the University of California, Lawrence Berkeley National Laboratory

- LRG: $z \in (0.4, 0.6)$, $z \in (0.6, 0.8)$, $z \in (0.8, 1.1)$
- QSO: $z \in (0.8, 2.1)$, $z \in (2.1, 3.5)$

DESI's focal plane is designed to optimize the collection and analysis of spectra and is structured to subtend a radius of 1.6 degrees on the sky. It hosts 5020 robotically controlled fiber positioners, each capable of precise movements to align an optical fiber with a specific target. These fibers channel the light collected from the sky into spectrographs for analysis. The focal plane is divided into 10 wedge-shaped segments known as 'petals.' Each petal contains 500 fiber positioners and directs the collected light to its own dedicated spectrograph. This segmentation helps manage the vast amount of data being collected and ensures that each spectrograph operates within its designed capacity. On the edge of each petal, there is a rectangular area reserved for a 'guide and focus array' camera — these cameras track stars to maintain the telescope's focus and correct aim during observations. This ensures that the telescope remains precisely

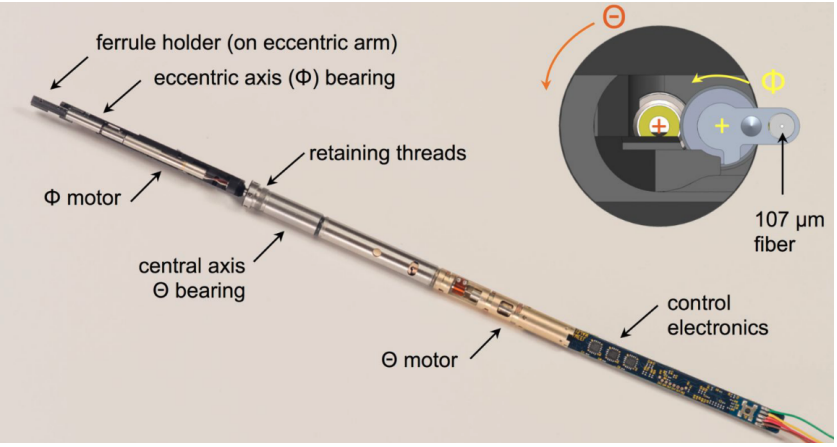


Figure 1.11: A DESI fiber positioner. On the insert (top-right), the centers of the two axes are indicated. Figure reproduced from [90].

targeted on the desired sky region even as external conditions change. See Figure 1.10 for a visual.

When an observation has been made (see Figure 1.12 for an example of spectra), it is run through the spectroscopic pipeline described in [47]. This pipeline measures the redshifts, uncertainties and spectral classes of targets. The schematic in Figure 1.13 summarizes this process.

For a detailed overview of the DESI survey strategy, we refer the reader to [89]. We also introduce the term 'tile', where each tile denotes a particular pointing in celestial coordinates, with an associated set of targets assigned to fiber positioners. The survey strategy prioritizes which tiles should be observed, provided information such as current observing conditions. DESI observing time is divided into 'dark' and 'bright' time programs. The former encompasses LRGs [110], ELGs [81] and QSOs [20] while BGS targets are observed in bright time when conditions are not optimal for dark viewing [48].

A given area on the sky is covered by multiple tiles to ensure observation of the different target classes to a high degree of completeness. DESI prioritizes the completion of observations at a given region of the sky over covering the full footprint. DESI targets are selected based on photometry from the DESI Legacy Imaging Survey [35].

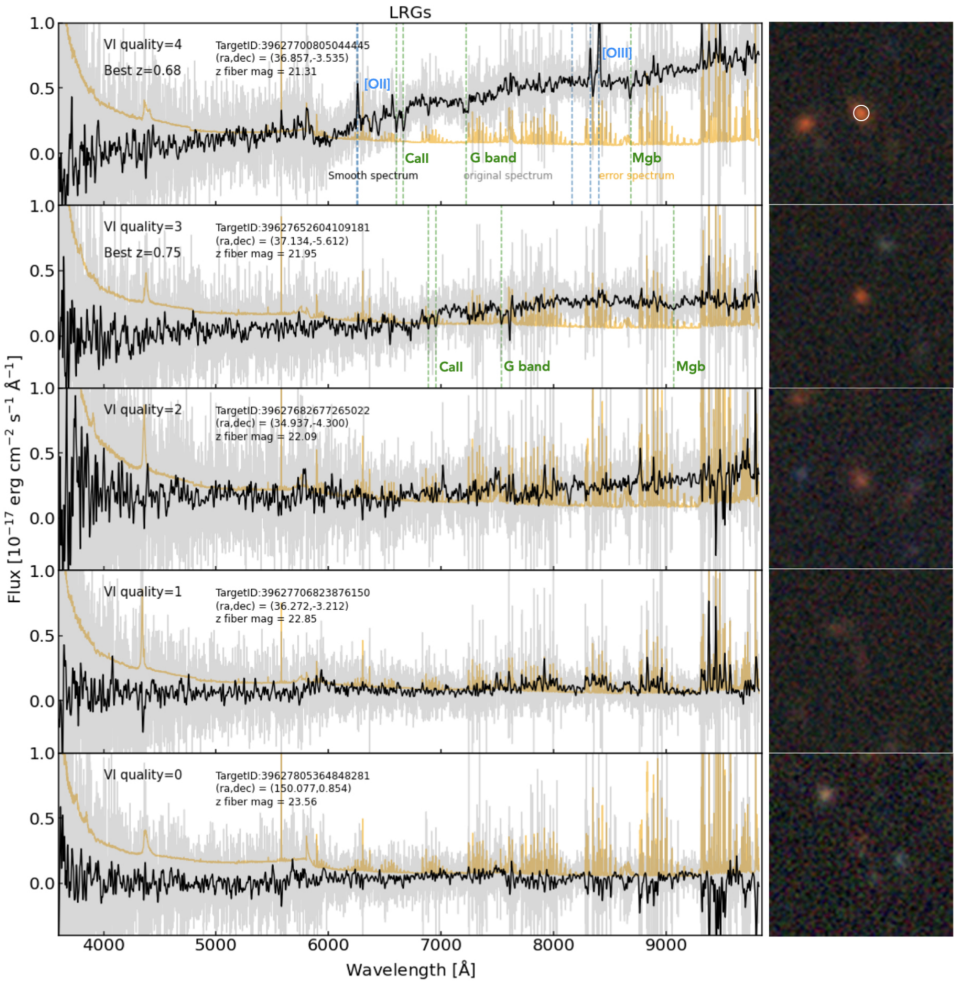


Figure 1.12: A set of example LRG spectra from DESI in the left panels. They are ranked according to quality (increasing from 0 to 4). Qualities 4 and 3 have multiple spectral lines while qualities 2 and below do not have spectral features that are robustly identifiable. The right panels show the DESI Legacy Survey images of the LRG targets. The spectra in gray, black, and orange are the originally observed galaxy spectrum, the smoothed spectrum with a Gaussian filter, and the error spectrum, respectively. Caption adapted from and figure reproduced from [54].

We introduce the term 'fiber assignment'. This is the process of selecting which targets on the sky are actually observed by the instrument. The DESI FiberAssign software is used to define the DESI survey footprint and to gather data on all targets within that area. For each tile, the properties of each scenario where a DESI target is accessible to a specific fiber positioner are determined. These are identified as 'potential assignments.' This information is used for calculating completeness statistics and

properly adjusting for their variations. While all these scenarios are considered potential assignments, only one target per positioner receives the actual assignment. The assignment is given to the target with the highest `PRIORITY`, and in cases where multiple targets share the same `PRIORITY`, the decision is made based on `SUBPRIORITY`. All this information is contained in files called 'merged target ledgers' (MTLs). These are lists of potential DESI targets (each having a unique identifier or `TARGETID`), coordinates, and priorities. In particular, in decreasing order of priority, we have QSOs, LRGs and ELGs. A 10% subsample of ELGs is boosted to match the priority of LRGs to help extract the LRG-ELG cross-correlation signals.

For fiber assignment, it is crucial to be aware of some real world constraints that necessitate the procedure to begin with. While DESI has 5000 fiber positioners, each individual fiber has a finite area it can cover. Therefore, in areas of the sky where there is a high concentration of targets, it may not be possible to observe every target even with multiple passes of the region on multiple tiles. Figure 1.14 shows the mechanics of a fiber positioner. Note the θ and Φ axes that describe the motion of a positioner. Motion along the former direction can be visualized as a rotation of the positioner, while the latter direction corresponds to an 'arm' that can swing out to provide additional coverage.

Also, there is a minimum separation that two fibers can be placed at physically, which can prevent observations of target pairs with separation less than this distance in one pass. These effects are called fiber collisions. Note that this does not have to refer to real, physical collisions, especially in the case of the first constraint described here. It is these collisions that have to be taken into account by the full fiber assignment algorithm and that contribute to making it a computationally expensive process, which drives the need for developing alternate methods of simulating fiber assignment. This is especially important for the development of multiple large mock catalogs that the collaboration uses for a variety of reasons, including testing code and inferring covariance matrices. We will return to this in future chapters and solving this problem is at the core of this

dissertation.

1.4 Outline

This goal of this dissertation is to discuss the fiber assignment problem, particularly in the context of being able to replicate its effects in an efficient manner. This is important for two main reasons: It helps to understand the effects of the fiber assignment process itself and also provides us with methods to simulate the process in ways that are computationally economical and expeditious. Chiefly, I will discuss ways to emulate fiber assignment via an algorithm we call 'fast fiber assignment' (FFA). The package that implements this is called Fast Fiber Assign and we will use the two terms interchangeably. This method has been employed for generating mock catalogs for DESI with very good success but with certain challenges such as covariance matrices that don't exactly match the real data.

Until now, a detailed analysis of the way the emulator behaves depending on how the parameters that control its behavior are adjusted was a topic that remained to be studied. The primary goal of this dissertation is to present the important work I did in emulating fiber assignment for the DESI Y1 analysis, ultimately contributing to cosmological constraints from baryon acoustic oscillations and full shape analysis. As an additional investigation, I will also present results from my study of how the FFA emulator can be tuned for better performance. To lead up to this, we structure this dissertation as follows.

In Chapter 2, I will discuss how the clustering statistics introduced in Chapter 1 are connected to cosmological analyses. I will focus on techniques that involve analyzing the BAO peak and also the broadband shape of the power spectrum or two-point correlation function. Along with this, there will be a discussion of covariance matrices and different ways to calculate them. This will set us up with the analysis techniques I will use later on.

In Chapter 3, I consider the fiber assignment problem in detail, focusing on ways to mitigate it and how it can be simulated. The content of this chapter is heavily based on content from [10], a paper on which I am the second author and which first formally introduced the Fast Fiber Assign algorithm. I will describe the algorithm and show results that reflect its performance, along with an explanation of the work I did in contributing to DESI Y1 analysis.

In Chapter 4, I will present all the tests and discussion for the examination I conducted into the behavior of the emulator under parameter adjustment, along with the effects of this on estimated covariance matrices. This analysis in this context is original. It is based on combining various other techniques, such as those described in Chapter 2, which are appropriately cited. Credit is also due to other members of the DESI collaboration for providing ideas on how to proceed with this, as mentioned in the Acknowledgments section.

Chapter 5 will be a short coda on DESI Y1 cosmology results from baryon acoustic oscillations and full shape analysis. I will present dark energy constraints from DESI Y1 papers to show the final results that were made possible in part by efforts towards developing fast fiber assignment. Finally, I will summarize the dissertation and put forward my closing remarks in Chapter 6.

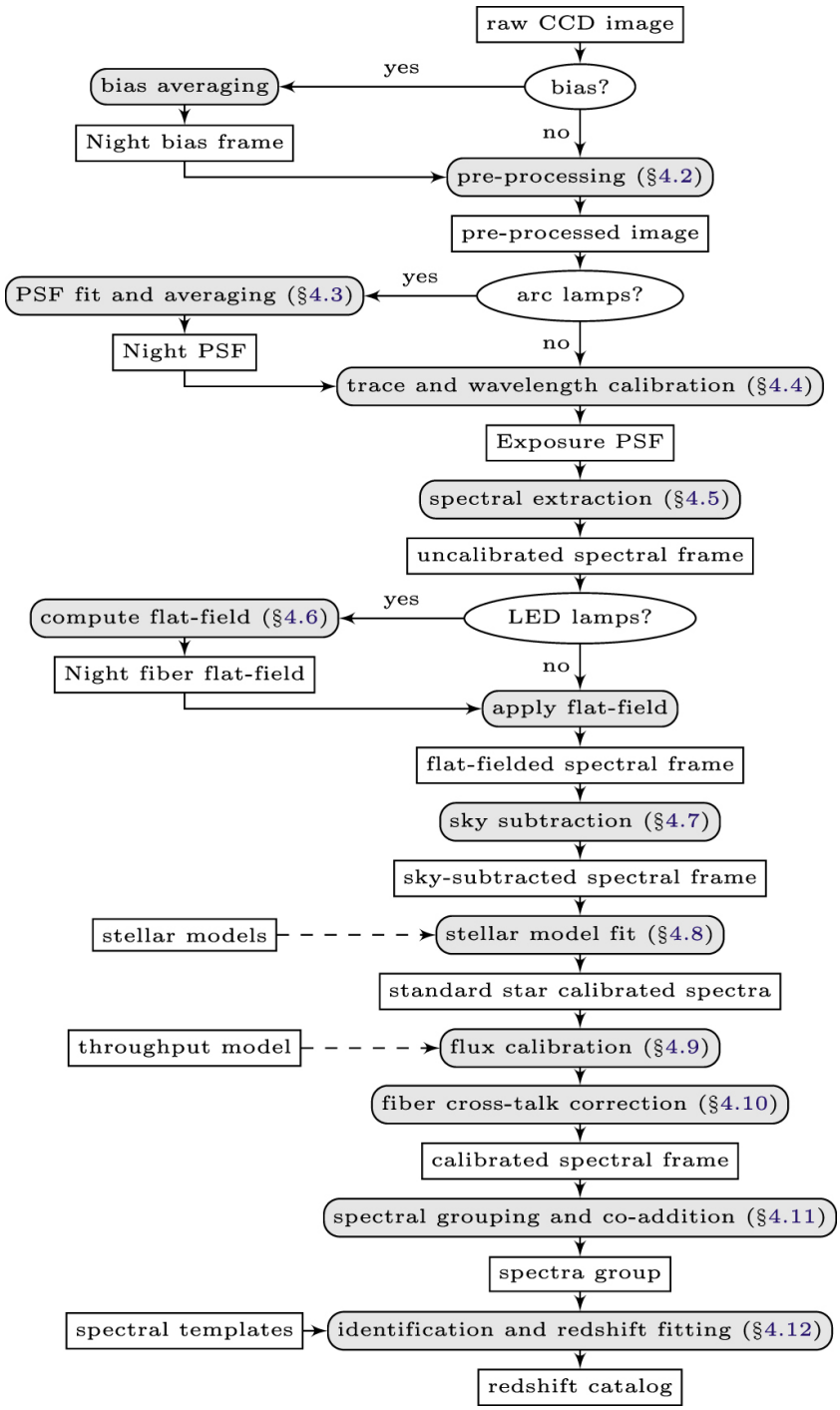


Figure 1.13: Flowchart of the spectroscopic pipeline for DESI, showing the many processing steps that the data has to undergo before it can be analyzed for cosmology. Figure reproduced from [47]. The section numbers in the figure refer to the original paper.

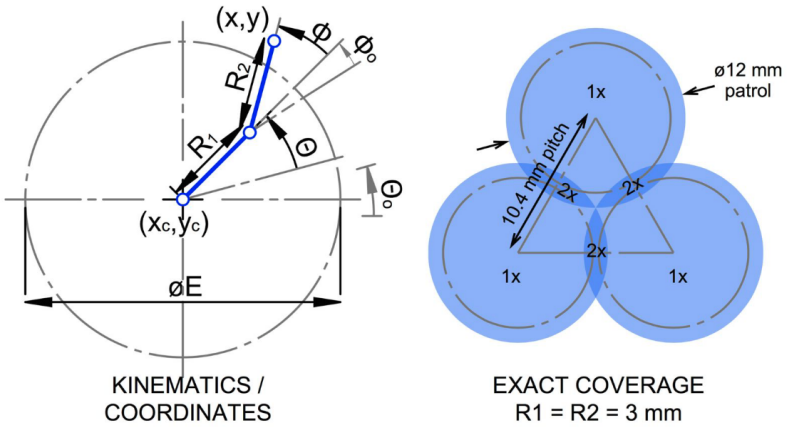


Figure 1.14: The left panel of this figure shows the eccentric axis $(\theta - \Phi)$ kinematics. Whenever R_2 is retracted within the dashed circle E, a positioner is guaranteed free rotation about θ without any obstruction from its neighbors. On the right, this figure shows how the patrol coverage extends beyond the pitch. This assures complete coverage but requires implementation of collision avoidance algorithms during the move. Caption adapted and figure reproduced from [90].

CHAPTER 2

Cosmology from Clustering

2.1 Extracting Information From Surveys

In a spectroscopic survey such as DESI, data consists of measured redshifts from galaxies in the sky. For analysis, galaxies are distributed into bins of different effective redshifts. Within these bins, the clustering statistics such as ξ and $P(k)$ are measured. Note here we will use the term 'clustering statistic' when referring to either of these functions generally. To be able to calculate these, redshifts need to be converted into distances. For this, we need to make a choice of cosmology to use and this is typically called the 'grid cosmology'. Once the two-point correlation function or the power spectrum has been computed, there are multiple methods of extracting cosmological information from these statistics. Broadly, they can be split into analyzing the BAO peak specifically (henceforth called the 'classic' BAO approach) and analyzing the broader shape of the clustering function.

2.1.1 Classic BAO Approach

In the classic BAO approach, a clustering statistic 'template' is evaluated using a reference cosmology. This reference cosmology does not, in principle, have to be the same as the earlier grid cosmology, but they are often chosen to be the same. After separating this template into a 'wiggle' (containing BAO information) and 'non-wiggle'

(containing the broadband statistic) component, the non-wiggle part is marginalized over and the wiggle part is used to fit the data. See Figure 2.1 for examples of various templates. In particular, the position of the BAO peak is measured via rescaling using the following free 'dilation' parameters [72]:

$$\begin{aligned} \alpha_\perp &= \frac{D_M(z) r_d^{\rm ref}}{D_M^{\rm ref}(z) r_d} \\ \alpha_\parallel &= \frac{H^{\rm ref}(z) r_d}{H(z) r_d^{\rm ref}} \end{aligned} \tag{2.1}$$

These are the ratios between the 'true' and reference values of the comoving angular diameter distance D_M and the Hubble distance D_H, respectively, relative to the sound horizon r_d.

$$\begin{aligned} D_M(z) &= \int_0^z \frac{dz'}{H(z')} \\ D_H(z) &= \frac{1}{H(z)} \end{aligned}$$

Within ΛCDM , these can be interpreted in terms of the matter density Ω_m and $H_0 r_d$. However, a BAO-only approach like this does not capture the information contained in the matter transfer function and extracting that requires examining the broadband clustering statistic.

2.1.2 Compressed Full Shape Analysis

Naturally then, an alternative approach is to use the full statistic template to perform the fitting, without any wiggle and non-wiggle component separation. This first iteration of analyzing the full shape of the power spectrum is what we will call the classic or compressed Full Shape (FS) analysis. Note that in the current literature, the various ways of examining the full shape of the power spectrum and correlation function go by different, often inconsistent names. FS analyses rely on perturbation theory codes that

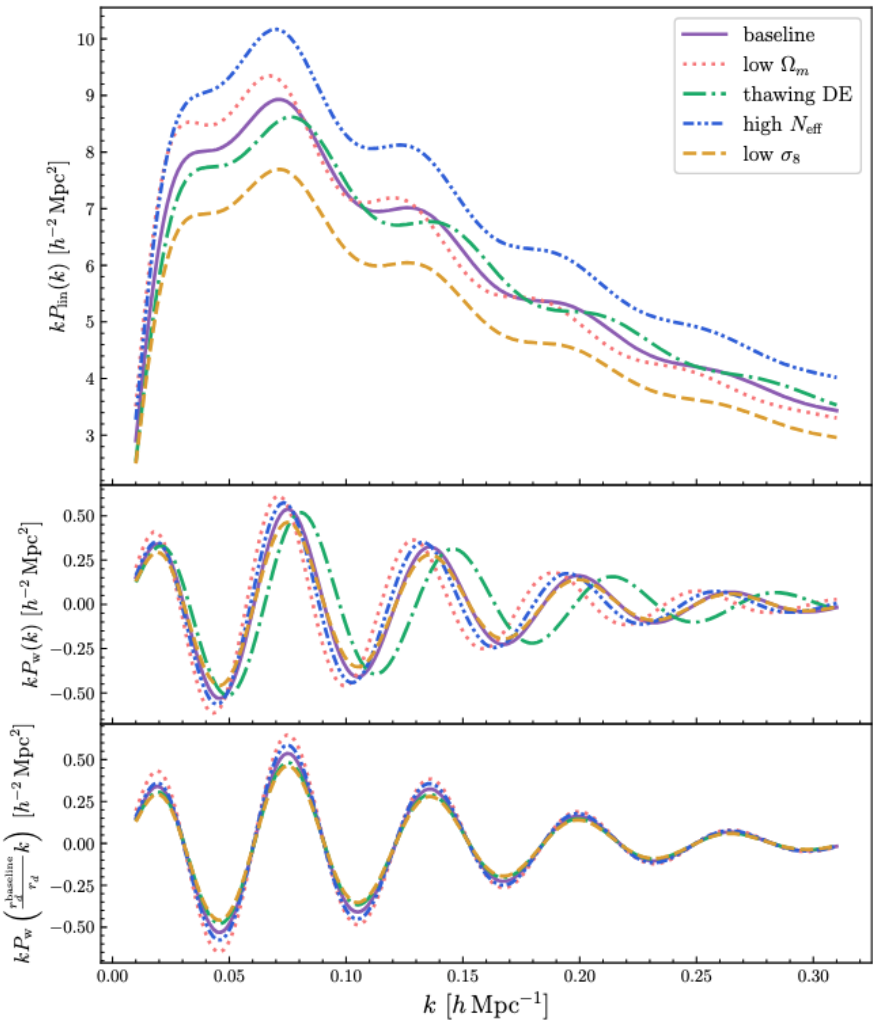


Figure 2.1: The top panel of this figure shows the linear power spectrum template at $z = 0$ for various cosmologies. The middle panel isolates the BAO 'wiggle' part, $P_w(k)$, where we can see the effect of the different sound horizon scales for each cosmology. The bottom panel shows the alignment of the oscillatory pattern after correcting for the difference in the sound horizon scales. Figure from [77].

predict the redshift-space galaxy power spectrum [23, 24, 66]. In this classic FS analysis, we gain sensitivity to the the $f\sigma_8$ parameter. Note that f and σ_8 were introduced via Equations 1.20 and 1.26, respectively. f, the growth rate of structure, is responsible for the large-scale bulk velocity component along the line-of-sight, causing an increased clustering signal in this direction. σ_8 gives the absolute amplitude of the power spectrum

and is set by the template. Both parameters are highly degenerate but their product is template-independent. This increased clustering along the line-of-sight does not change the BAO peak position and we can thus compress the information from the power spectrum multipoles into the set of parameters $\{\alpha_\perp, \alpha_\parallel, f\sigma_8\}$. The BAO dilation parameters from Equation 2.1 are often re-parametrized into a purely isotropic term $\alpha_{\rm iso}$ and an anisotropic term $\alpha_{\rm AP}$:

$$\begin{aligned} \alpha_{\rm iso}(z) &\equiv [\alpha_\parallel(z)\alpha_\perp^2(z)]^{1/3} = \frac{D_V(z) r^{\rm fid}(z)}{D_V^{\rm fid}(z) r_d} \\ \alpha_{\rm AP}(z) &\equiv \alpha_\parallel(z)/\alpha_\perp(z) = \frac{D_H(z) D_M^{\rm fid}(z)}{D_H^{\rm fid}(z) D_M(z)} \end{aligned} \tag{2.2}$$

where

$$D_V(z) \equiv [z D_H(z) D_M^2(z)]^{1/3} \tag{2.3}$$

is often called the isotropic BAO distance. This is not a fundamental quantity but has units of distance.

2.1.3 Full Modeling

Given a power spectrum $P(k)$, it is also possible to bypass the compression and directly fit cosmological models to the multipoles of the power spectrum. This comprises information from the BAO and RSD signals along with other features (for example, the slope of the power spectrum). This approach is often called Full Modeling (FM). At each analysis step, a linear power spectrum is generated via an Einstein-Boltzmann code (a code to compute perturbations in the universe, such as CAMB [60, 49] or CLASS [14]) and a non-linear theory code evaluates order-one loop corrections in perturbation theory to create the non-linear galaxy power spectrum. The reference cosmology is chosen at the start and kept fixed. The FM approach has to account for the distortions caused if this cosmology does not match the one being tested. To do so, the modeled power

spectrum is transformed via the following rescaling parameters [31]:

$$q_\perp(z) = \frac{D_M(z)}{D_M^{\rm ref}(z)}$$
$$q_\parallel(z) = \frac{D_H(z)}{D_H^{\rm ref}(z)} \tag{2.4}$$

In the FM analysis, the scaling parameters are internally derived parameters used to correct for a fixed reference cosmology. This is different from the scaling parameters of the classic BAO and Full Shape approaches, which are parameters of compression. The differences in these approaches are summarized in Table 2.1.

2.1.4 ShapeFit

There have been efforts to connect the classic and FM approaches. These have been motivated by the goals of achieving a model-independent analysis similar to the classic methods but with the tighter constraints of the latter without the large computational cost. The scheme we discuss that accomplishes this is ShapeFit [16]. The two major components of this bridge are: having the correct definition and application of the scaling parameters, and the capacity to model early-time physics on large scales of the full shape of the matter power spectrum.

In addition to the scaling parameters from Equations 2.1 or 2.2, which transform the modeled power spectra in the horizontal direction, ShapeFit considers how to capture 'vertical' information in more detail. This is contained in σ_8 (Equation 1.26). ShapeFit also introduces two shape parameters, m and n, whose meaning and effect we discuss below.

As talked about before, the linear power spectrum's broadband shape is dependent on the transfer function, which contains information from after the end of inflation and encodes the effect that the matter-radiation equality has on the growth of density fluctuations. ShapeFit proposes compressing the extra signal from this into a few effective

parameters such that the early and late time information is decoupled but can be easily combined when interpreting results in a cosmological light. The logarithmic slope of the linear power spectrum is assumed to consist of two components: the overall scale-independent slope n and the scale-dependent slope m. n captures the information on the primordial scaling of power and its deviation from scale invariance. This is analogous to the spectral index n_s. m is the maximum slope at a chosen pivot scale k_p, which is selected to be around the matter-radiation equality scale, letting m describe the physics of that era. Using these, the reference power spectrum template $P_{\rm ref}(k)$ is transformed into a new template $P'_{\rm ref}(k)$:

$$P'_{\rm ref}(k) = P_{\rm ref}(k) \exp\left\{\frac{m}{a}\tanh\left[a\ln\left(\frac{k}{k_p}\right)\right] + n\ln\left(\frac{k}{k_p}\right)\right\} \tag{2.5}$$

where k_p and a are calibrated using the formalism in [37]:

$$a \approx 0.6$$
$$k_p \approx 0.03h\text{Mpc}^{-1}$$

Now σ_8 not only becomes dependent on m, but there is also uncertainty in defining a scale of 8 h^{-1}Mpc when the the cosmology defining h is unknown. ShapeFit accounts for this by rescaling this in units of the sound horizon:

$$\sigma^2_{s8} = \int_0^\infty (ks)^3 P(ks, z) W^2_{\rm TH}(ks\ 8h^{-1}\text{Mpc}) \tag{2.6}$$

where $s = r_d/r_d^{\rm ref}$. In practice, ShapeFit constrains $fA_{sp}^{1/2}$ where A_{sp} is the power spectrum amplitude calculated at k_p. With the amplitude of the template fixed, we only fit for f. We can apply the following transformation to get an approximation for $f\sigma_{s8}$

[15]:

$$f\sigma_8 \approx f\sigma_{sd}^{\rm ref} \frac{fA_{sp}^{1/2}}{(fA_{sp}^{1/2})^{\rm fid}} \exp\left[\frac{m}{2a_m}\tanh\left(a_m \ln\left(\frac{r_d^{\rm fid}[h^{-1}{\rm Mpc}]}{8h^{-1}{\rm Mpc}}\right)\right)\right] \tag{2.7}$$

This method is exact in first-order perturbation theory and a good approximation at higher orders [2]. Usually, $df = f/f^{\rm ref}$ is quoted when reporting results on mocks and this dissertation will follow the same trend when ShapeFit is used for our analysis.

Fit Approach	**BAO-only**	**Full Shape**	
	Classic	**Compressed**	**Full Modeling**
Information Source	BAO wiggles	BAO + P(k) amplitude	Full P(k)
$P(k)$ **Template**	fixed	fixed	varies
Scaling Parameters	free $\alpha_\perp, \alpha_\parallel$	free $\alpha_\perp, \alpha_\parallel$	derived $q_\perp, q_\parallel$
Linear RSD	marginalized over	free f	derived f
Global Amplitude	marginalized over	fixed or free σ_8	free σ_8 or A_s

Table 2.1: Summary of the three main approaches for extracting cosmological information from galaxy surveys. Table adapted from [16].

2.2 Covariance Estimation

Any data only has meaning when we also know how spread out or uncertain it is. That is, we require an accurate and precise estimate of the data covariance matrix to be able to judge how well our models fit the data. This is a challenge in cosmology as we live only in one universe and have access to a single realization of data. We can calculate covariance matrices for any set of parameters or measured quantities that we are dealing with. In this dissertation, we will primarily be concerned with the covariance matrix of the two-point correlation function (or the power spectrum). When we measure the two-point correlation function ξ as a function of separation s between galaxies, we divide our observations into separation s bins. The diagonals of the covariance matrix

of xi represent the variance of the function in a given bin. This is the variance that we expect in that measurement due, say, to random effects like shot noise or cosmic variance. The off-diagonal entries represent the covariances between the measurements in different bins. This can be attributed to the fact that the same large-scale structures influence the galaxy distribution over a range of scales or to the way the survey geometry and selection effects affect measurement. There are three broad categories into which current methods of estimating covariance matrices in cosmology can be split.

First, there are analytic methods. These use theoretical models of clustering to derive the underlying statistics, for example in [13]. An obvious advantage of these is that they can be computationally inexpensive. On the other hand, calculating analytical covariances is not an easy task. To compute the covariance of two-point statistics (ξ), we need to be able to describe the non-Gaussianity in the four-point correlation function. Because it depends on the knowledge and accurate modeling of non-linear effects using perturbation theory, this is a complicated endeavor.

Second, there are mock-based covariances. Mocks are simulated datasets with underlying assumptions that often make them simpler (though less accurate) to compute compared to full N-body simulations [27, 28, 62]. With a large number (to avoid sample variance) of such mocks that are accurate (they reproduce the N-point statistics with limited biases), it is possible to directly compute the covariance matrix. When using a large number of mocks (often in the thousands for major surveys [109]), it can be computationally expensive both to generate the mocks and then to run calculations using them. This is why often large numbers of less precise mocks have to be used for this method. Given measurements of the two-point correlation function in n_b separation bins from n_m mocks, the mock covariance matrix is defined as

$$C_{ij} = \frac{1}{n_m - 1} \sum_{k=1}^{n_m} \left[\xi_i^{[k]} - \langle \xi_i \rangle\right] \left[\xi_j^{[k]} - \langle \xi_j \rangle\right] \tag{2.8}$$

where $\xi_i^{[k]}$ is the i^{th} bin of the correlation function of the k^{th} mock and $\langle \xi_i \rangle$ is the mean of the i^{th} bin over all the mocks.

Finally, there are 'internal' methods. These rely on resampling the data, split into various sub-samples, using chosen schemes to create many observations. For example, bootstrapping relies on creating alternative realizations of the data by sampling with replacement from the original set. Jackknifing, on the other hand, relies on a 'leave one out' approach where the data is split into subsamples and alternate realizations are created by leaving a different subsample out for each of them. These methods do not make assumptions about the underlying model of gravity and are less susceptible to unknown physics. Conversely, these methods have been shown to bias measurements of the covariance [43, 40], but methods have been proposed to overcome this as we will soon see. In this dissertation, we will particularly be concerned with using jackknife covariances [71]. In this case, if the dataset has been divided into n_{jk} subsets, the covariance matrix becomes

$$C_{ij} = \frac{n_{\text{jk}} - 1}{n_{\text{jk}}} \sum_{k=1}^{n_{\text{jk}}} \left[\xi_i^{[k]} - \langle \xi_i \rangle\right] \left[\xi_j^{[k]} - \langle \xi_j \rangle\right] \tag{2.9}$$

where $\xi_i^{[k]}$ is the i^{th} bin of the correlation function of the k^{th} jackknife subset and $\langle \xi_i \rangle$ is the mean of the i^{th} bin over all regions. The pre-factor in Equation 2.9 is larger than the one in Equation 2.8 to account for the decrease in covariance due to the overlap between subsamples.

2.2.1 Jackknife Covariances

Let AA_k be the auto-counts of pairs of galaxies in the k^{th} survey area. These survey areas are the ones that the dataset has been split into and which are used to form the jackknife subsamples. Also let CC_k be the cross-counts between galaxies in this area and those in the subsample formed by removing this area. Finally, let TT_{tot} and TT_k be

the total counts in the full area and the k^{th} subsample, respectively. Then

$$TT_k = TT_{\text{tot}} - AA_k - CC_k$$

Following the prescriptions in [69] and [97], we compute the covariance matrix as

$$\begin{aligned}\text{cov}(TT, TT) &= \text{cov}(AA, AA) + \text{cov}(CC, CC) + 2\text{cov}(AA, CC) \\ &= \frac{n_{\text{jk}} - 1}{n_{\text{jk}}} \sum_{k=1}^{n_{\text{jk}}} (\theta_{a,k} - \bar{\theta}_a)^2 + \frac{(n_{\text{jk}} - 2)^2}{2n_{\text{jk}}(n_{\text{jk}} - 1)} \sum_{k=1}^{n_{\text{jk}}} (\theta_{c,k} - \bar{\theta}_c)^2 \\ &+ \frac{\sqrt{2}(n_{\text{jk}} - 2)}{n_{\text{jk}}} \sum_{k=1}^{n_{\text{jk}}} (\theta_{c,k} - \bar{\theta}_c)^2 (\theta_{a,k} - \bar{\theta}_a)^2\end{aligned} \tag{2.10}$$

where

$$\theta_{a,k} = \frac{1}{n_{\text{jk}} - 1}(n_{\text{jk}}\overline{AA} - AA_k)$$

and

$$\theta_{c,k} = \frac{2}{n_{\text{jk}} - 2}\Big(\frac{n_{\text{jk}}}{2}\overline{CC} - CC_k\Big)$$

are estimators of the normalized auto-counts and normalized cross-counts, such that $\bar{\theta}_a = \overline{AA}$ and $\bar{\theta}_c = \overline{CC}$. One observes that all three terms in Equation 2.10 scale differently with n_{jk}. It has been argued that this difference in scaling between terms is the source of bias in the typical jackknife methodology [69].

2.2.1.1 Correcting the Bias — Mohammad-Percival Correction and FitCov

It is suggested in [69] to introduce a weight α that scales the cross-counts in Equation 2.10 to account for the scaling difference. Using this, the covariance of the weighted

cross-counts becomes

$$\mathrm{cov}(\alpha CC, \alpha CC) = \frac{(n_{\mathrm{jk}} - 2\alpha)^2}{2\alpha^2 n_{\mathrm{jk}}(n_{\mathrm{jk}} - 1)} \sum_{k=1}^{n_{\mathrm{jk}}} (\theta_{c,k} - \bar{\theta}_c)^2$$

Note that with the appropriate value for $\alpha = n_{\mathrm{jk}}/[2 + \sqrt{2}(n_{\mathrm{jk}} - 1)]$, the cross-counts term can be made to scale as the auto-counts term in Equation 2.10. This approach is what we call the Mohammad-Percival correction. A demonstration of its improvement over the 'no correction' approach is presented in Figure 2.2.

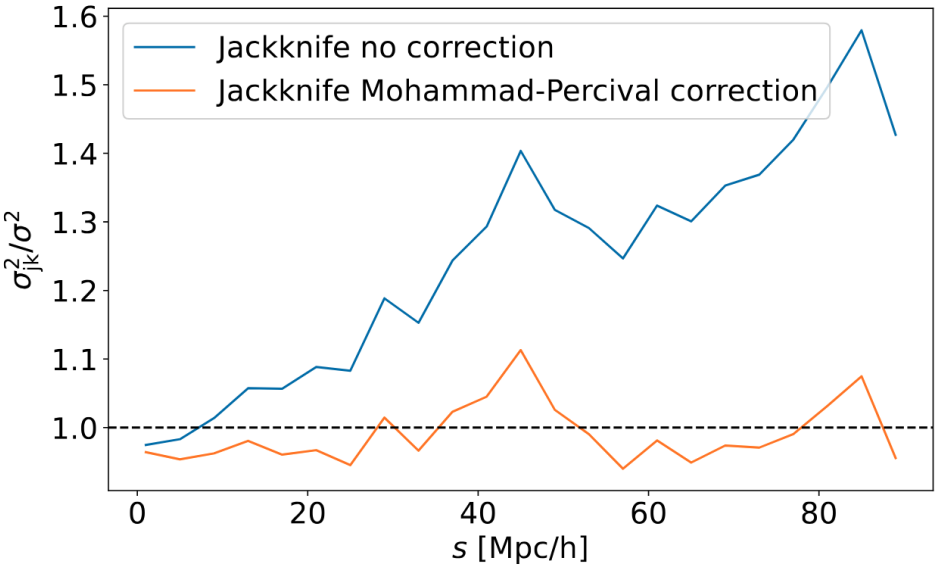


Figure 2.2: A comparison of the diagonals of the estimated covariance matrix of the real-space correlation function. These were obtained from 1000 Poisson random cubic mocks. The quantity on the vertical axis is the ratio of the mean of the diagonals of the covariance obtained via jackknifing to the diagonals of a directly calculated mock covariance. The Mohammad-Percival correction takes care of the scale-dependent bias that is noticeable in the case of the 'no correction' approach. Note that these are Poisson random catalogs and the $\mathrm{cov}(CC, AA)$ term in Equation 2.10 is zero. Figure reproduced from [97].

With the aforementioned correction, if $\mathrm{cov}(CC, AA)$ vanishes, the terms in the covariance equation would indeed scale equally. But for real galaxies, there are physical correlations and this assumption is not valid. Therefore, we have to account for the

correction to the third term in Equation 2.10 as well:

$$\mathrm{cov}(\alpha CC, AA) = \frac{n_{\mathrm{jk}} - 2\alpha}{\sqrt{2}\alpha^2 n_{\mathrm{jk}}} \sum_{k=1}^{n_{\mathrm{jk}}} (\theta_{c,k} - \bar{\theta}_c)(\theta_{a,k} - \bar{\theta}_a)$$

With this adjustment, any fixed value of α would leave the scaling of terms in the covariance equation different. It is proposed in [97] to treat α as a free parameter that can be fit for by fitting the estimated covariance from a limited set of mocks. The scheme they put forward is summarized in Figure 2.3. This is implemented in the code package FitCov and we will use 'FitCov' and 'fitted jackknife' interchangeably as the names for this approach.[1] Using their notation, assume we have a set of N_m mocks $S = \{S_1, ..., S_{N_m}\}$. Let S/S_k be the set of mocks with S_k removed and define $C_{ij}[S/S_k]$ to be the mock covariance of this collection. Further define $C_{ij}[S_k](\alpha)$ to be the jackknife covariance from a mock S_k with a selected α weight.

Then the uncertainty on the diagonal of the covariance $\Xi_{i,j}(\mathrm{diag}(C))$ can be estimated. Resampling our original set of mocks gives us N_m covariances $C_{ij}[S/S_k]$, from which $\Xi_{i,j}(\mathrm{diag}(C))$ can be calculated, where we are confined to the diagonal elements because otherwise we do not have enough degrees of freedom to create a non-singular covariance of matrices [101].

$$\begin{aligned}\Xi_{i,j}(\mathrm{diag}(C)) &= \mathrm{cov}(C_{ii}, C_{jj}) \\ &= \frac{N_m - 1}{N_m} \sum_{k=1}^{N_m} (C_{ii}[S/S_k] - C_{ii}[S])(C_{jj}[S/S_k] - C_{jj}[S])\end{aligned}$$

If N_m is not greater than the number of elements in the fitted part of the covariance, that is, in the case of a small number of mocks, this prescription can be restricted to just the diagonal elements of Ξ_{ij}. We write the estimator of the covariance based on the

[1] https://github.com/theonefromnowhere/FitCov

mean of N_m jackknife covariances:

$$C_{ij}(\alpha) = \frac{1}{N_m}\sum_{k=1}^{N_m} C_{ij}[S_k](\alpha)$$

Then χ^2 of the estimator describing $C_{ii}[S]$ is

$$\chi^2_C(\alpha) = \sum_{ij}(C_{ii}(\alpha) - C_{ii}[S])\left(\Xi^{-1}\right)_{ij}(C_{jj}(\alpha) - C_{jj}[S])$$

This can be minimized to obtain α such that $\chi^2_C(\alpha_{\min}) = \min(\chi^2_C(\alpha))$. So the estimator proposed by [97] is

$$C^{\text{fit}}_{ij} = C_{ij}(\alpha_{\min}) = \frac{1}{N_m}\sum_{k=1}^{N_m} C_{ij}[S_k](\alpha_{\min}) \tag{2.11}$$

To summarize, this fitted jackknife approach is used to adjust the weight α to achieve a 'balanced' scaling between the terms in the pre-weighted covariance, given in Equation 2.10, by using a limited set of mocks. For detailed tests of this approach, we refer the reader to [97] but we present Figure 2.4 from said paper to show how this approach takes care of biases. For this, the authors of this approach use the following definition of relative bias:

$$\frac{\Delta\sigma^2(\xi_l)}{\sigma(\sigma^2_{\text{mock}})} = \frac{\sigma^2(\xi_l) - \sigma^2_{\text{mock}}(\xi_l)}{\sigma(\sigma^2_{\text{mock}}(\xi_l))} \tag{2.12}$$

where $\sigma(\xi_l)$ is the variance of the multipole l from a jackknife approach, $\sigma_{\text{mock}}(\xi_l)$ is the corresponding variance obtained from 1500 lognormal mocks, and $\sigma(\sigma^2_{\text{mock}})$ is the uncertainty on the mock-based error bar, computed by applying the classical jackknife mock estimator to the set of mocks from which the covariance is estimated. This shows that with increasing number density $\bar{n}$, the underestimation of the 'pure' Mohammad-Percival method becomes more significant compared to the FitCov approach. This is

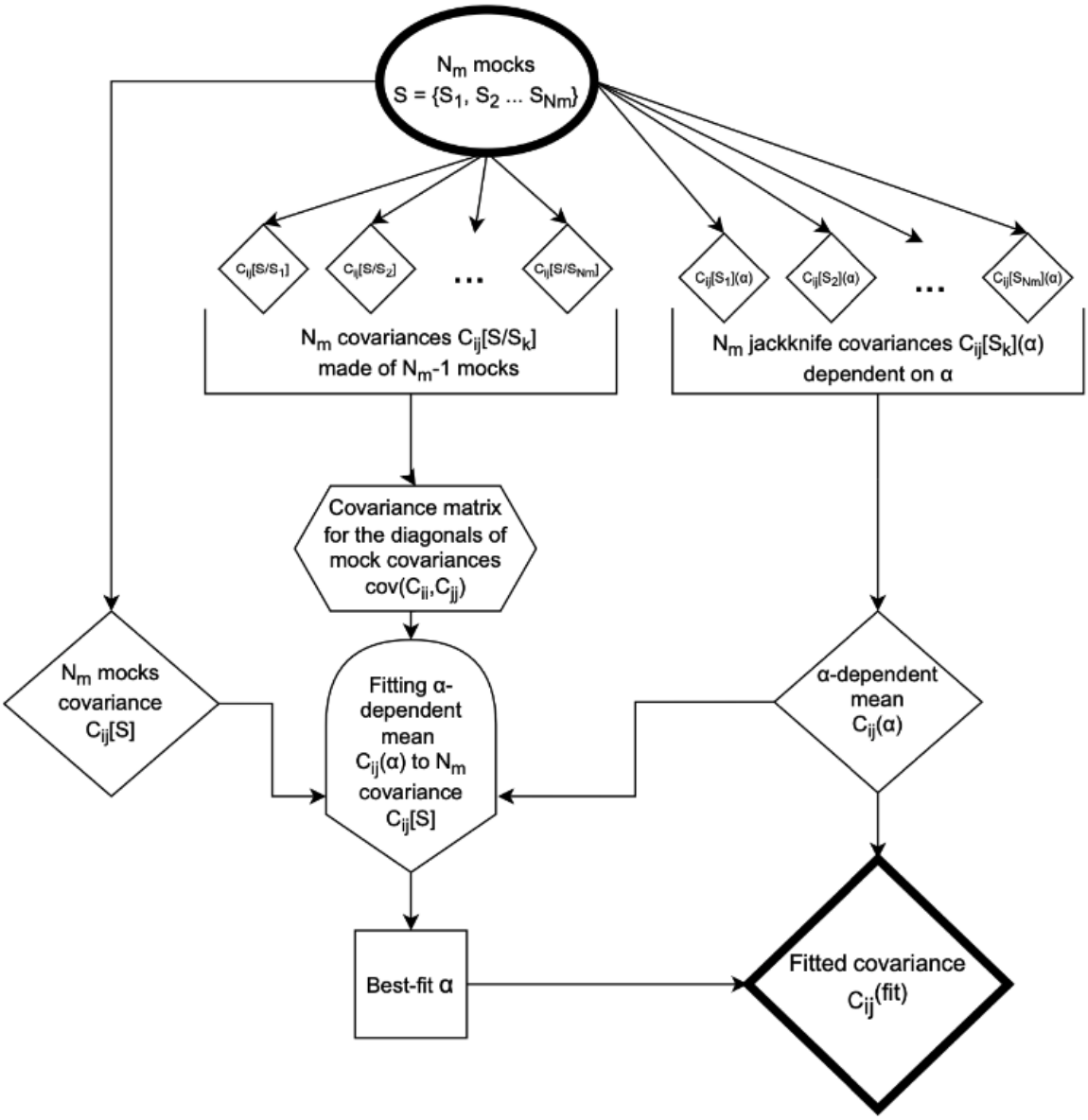


Figure 2.3: Schematic showing the fitting approach utilized by FitCov. Figure reproduced from [97].

because the increasing clustering of the data contributes to the cross-term between CC and AA becoming more significant. Moreover, the FitCov approach is not very sensitive to the number of jackknife samples, as shown in the same figure. This leaves us free to select the number of jackknife subsets we will use in our analysis later in this dissertation.

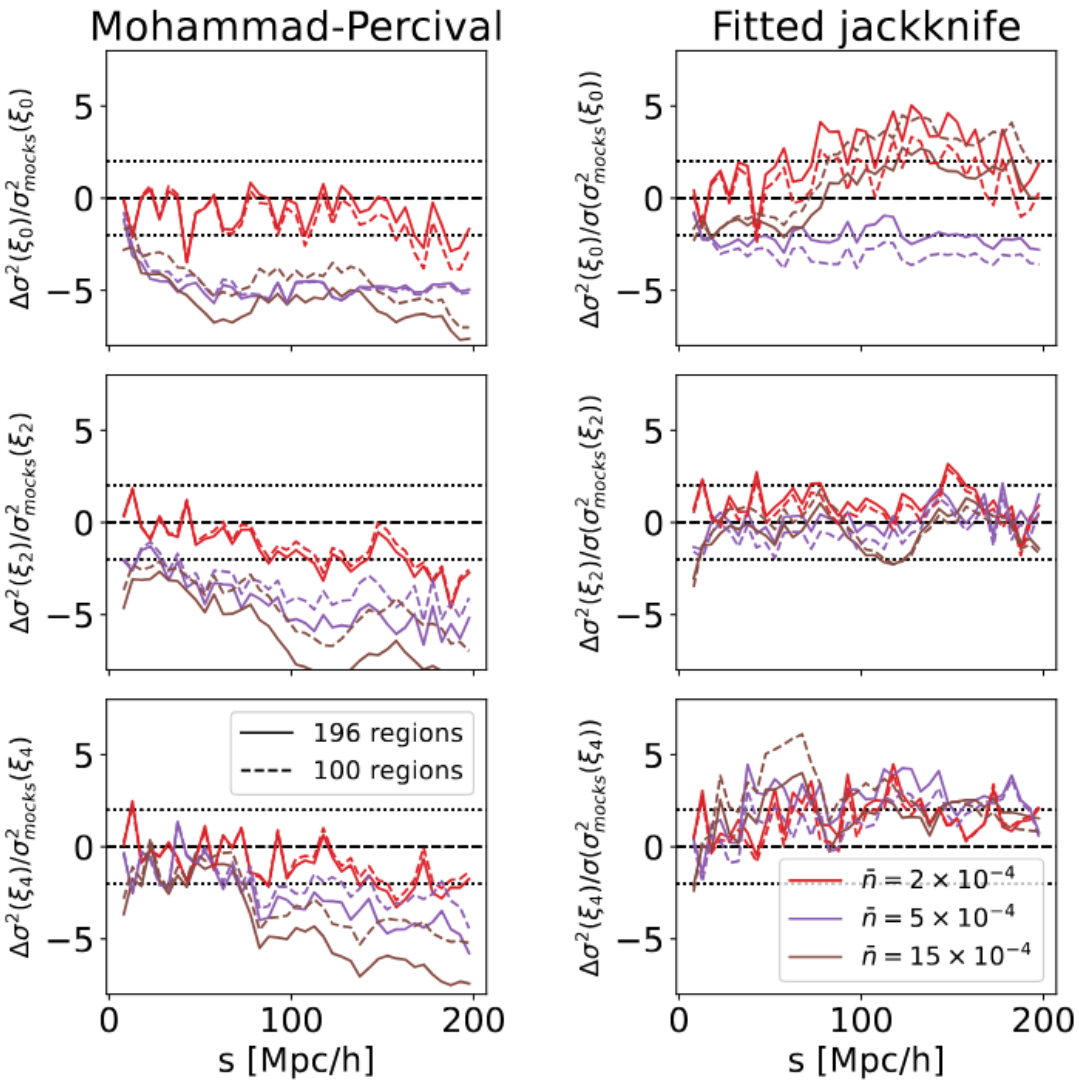


Figure 2.4: The average of the relative bias defined in Equation 2.12 plotted as a function of separation s for different number densities. Figure reproduced from [97].

CHAPTER 3

Fiber Assignment and Fast Fiber Assignment

In Chapter 2, this dissertation introduced the topic of fiber assignment (FA). Before proceeding with this in detail, I note that a lot of the results and observations in this chapter have been published before in [11], a paper on which I am the second author. The work due to others is appropriately cited. In particular, when the 'fast fiber assignment' algorithm is introduced, the results that are quoted from the aforementioned paper are ones that I actively took part in creating.

To return to fiber assignment, the physical geometry of a survey instrument, coupled with the geometry of structure in the sky, imposes certain limitations on the survey strategy. In the case of DESI, there are two main constraints: First, each of DESI's fiber positioners (see Figure 1.11) is confined to a patrol radius, which covers a limited fraction of the focal plane. Despite the fact that adjacent patrol regions overlap, in areas of the sky where the galaxy density is high, there may not be enough fibers to observe every target. Second, the finite size of positioners means that there is a minimum separation closer than which two fibers cannot be placed. This corresponds to a minimum angular separation on the sky, preventing the observation of all closely spaced galaxy pairs with a single observation. All these effects are often referred to conventionally as 'fiber collisions' though it should be noted that it is not always a matter of physical collisions between fibers.

Fiber assignment is therefore the process of selecting which targets on the sky are actually observed by the instrument, especially when fiber collisions are present. These collisions are a notable source of missing observations. This introduces fluctuations in the galaxy density that must be accounted for when analyzing clustering. Their net effect is a complex pattern of missing observations that are unevenly distributed across the survey but are concentrated in regions of high galaxy density. To counter this, different measures have been put forward that can broadly be split into two groups.

First, there are what we call 'model-level' approaches. These try to account for the missing observations through modifications of the underlying theory model itself. For example, they may remove the angular modes that are affected by fiber assignment [17, 75]. This has the drawback of introducing a loss of large-scale information. The model-level approach DESI adopts for the DESI DR1 full shape analysis is often call 'theta-cut' [79]. This involves truncating small angular scales that would be affected by fiber assignment and then modifying the clustering model to account for this. This accurately captures the nature of DESI fiber assignment and has the advantage of being applicable to both configuration and Fourier spaces.

Secondly, we have what we call 'estimator-level' methods. These involve corrections to the clustering measurements themselves. For example, galaxy pairs can be weighted appropriately when calculating the correlation function to account for missing observations. [11, 76] introduced an unbiased estimator-level correction. They show that the 'real' clustering can be inferred via Pairwise Inverse Probability (PIP) weights. This involves weighting galaxy pairs by the inverse probability of them being observed.

3.1 Inverse Probability Weights

This technique is introduced in [11]. Let us suppose that we have a full sample of galaxies and a selection algorithm to select a subset. We define p_i to be the probability of the i^{th} galaxy in the full sample of being selected. Then the probability that the pair

formed by the m^{th} and n^{th} galaxies is selected is

$$p_{mn} = p_m p_n (1 + c_{mn}) \tag{3.1}$$

where

$$c_{mn} = \frac{p_{mn}}{p_m p_n} - 1$$

c_{mn} can be understood as the selection correlation associated with the pair. The weight of each pair is then defined to be

$$w_{mn} = \frac{1}{p_{mn}} \tag{3.2}$$

This weighting scheme is called 'inverse probability weights' or, specifically, 'pairwise inverse probability' (PIP) weights as each galaxy pair requires its own weight. Note from Equation 3.1 that if the selection correlation becomes very small, then we can express the total weight as a product of the two individual weights. This can be seen clearly if we recast Equation 3.2 as

$$w_{mn} = w_m w_n w_{mn}^{(c)} \tag{3.3}$$

where $w_i = 1/p_i$ and $w_{mn}^{(c)} = 1/(1 + c_{mn})$. So if we can find a characteristic angular separation scale θ_{cs} such that for separations greater than this, the probability of a galaxy being observed is independent of the observed status of its neighbors, we can work with 'individual inverse probability' (IIP) weights. This makes the weight of the pair

$$w_{mn} = w_m^{IIP} w_n^{IIP} \tag{3.4}$$

where we use the superscript IIP to make it clear that this is only true under the

aforementioned assumption. In both cases, the data-data pair counts in the Landy-Szalay estimator become

$$DD(\mathbf{r}) = \sum_{\mathbf{x}_m - \mathbf{x}_n \approx \mathbf{r}} w_{mn} \tag{3.5}$$

where $\mathbf{x}_m - \mathbf{x}_n$ is the galaxy separation and '$\approx$' denotes that we are summing over a separation $\mathbf{r}$ bin.

To computationally estimate the PIP weights, within DESI, the following definition is used:

$$w_{ij}^{PIP} = w_i^{eff} w_j^{eff} \frac{w_{ij}^{eff}}{g_K(c_i, c_j)} \tag{3.6}$$

where $g_K(c_i, c_j)$ is the expected value of w_{ij}, the estimator of the individual inverse weights. The fractional term encodes the selection correlation information. This approach is fundamentally the same as the one proposed by [12]. According to this, the selection probability and its inverse can be found by running a targeting algorithm K times, where a target gets selected c times. To be computationally efficient, it is vital that a rapidly converging (with respect to K) estimator be used. [12] thus proposes the 'efficient estimator':

$$w^{eff} = \frac{K+1}{c+1} \tag{3.7}$$

These weights can thus be used to mitigate the effect of fiber assignment. This is important because, aside from needing accurate clustering of real data, we need to be able to accurately study the clustering of mocks to be able to extract good covariances.

3.2 Simulating Fiber Assignment

The capacity to reproduce the fiber assignment process accurately is an important problem. As mentioned in the preceding section, it is crucial for creating realistic mocks that are used to estimate covariances. Moreover, by having many realizations of the targeting process, the fiber assignment process itself can be studied and used to drive the creation of appropriate methods to mitigate its effects. Specifically, we can assign a bitwise array to each galaxy, where each bit corresponds to a simulated realization [11]. These bitweights can be used to efficiently estimate the assignment probability for a combination of galaxies. DESI uses two primary methods of simulating fiber assignment: a more accurate but computationally expensive approach and a much faster, approximate one.

3.2.1 Alternative Merged Target Ledger (altMTL)

We introduced merged target ledgers (MTLs) in Chapter 2. MTLs are used to track the observational state of the targets that DESI may observe. The overarching purpose of these ledgers is to impose decisions for targets that span multiple target classes and so could have competing observing requirements. For more details, we refer the reader to [89]. Targets in MTL files have `PRIORITY` entries that encode their observation priority — this is used to decide which targets are assigned a fiber. When multiple targets have conflicting `PRIORITY`, a randomly generated `SUBPRIORITY` entry controls the assignment. Thus by creating alternate random `SUBPRIORITY` values, realistic alternative MTLs (or altMTLs) can be generated. The assignment probability of targets can be studied by rerunning FiberAssign (the full DESI fiber assignment algorithm) on these altMTLs.

For a more thorough description, we refer the reader to [56]. To emphasize, the advantage of this approach is that the real DESI targeting code is run on mocks generated by this approach. This creates more realistic mock catalogs that contain the same information as real data. On the other hand, this approach is computationally intensive

and therefore cannot be used to create large numbers of mocks. For that, we require a different solution.

3.2.2 Fast Fiber Assignment (FFA)

The need to be able to create large numbers of mock catalogs with simulated fiber assignment motivated the development of an emulator — an algorithm that learns from the input-output datasets of a reference algorithm (in this case, the full FiberAssign code). The input (or training sample) that this emulator learns from is a realistic dataset (mock or real data) that has had the real fiber assignment algorithm run on it. The corresponding output label encodes whether or not a given galaxy has been assigned. We adopt a shallow learning method as there is a good understanding of the mean behavior of the fiber assignment algorithm. This keeps the model computationally simple with few parameters and leaves it interpretable. We call this process fast fiber assignment (FFA) and the code is called Fast Fiber Assign.

This emulator is based on the simplification that we can model the fiber assignment process as a function of how many neighbors a galaxy has and how many tiles it appears on. Recall that 'tile' refers to a specific pointing of the instrument. Therefore, the more tiles a target appears on, the more likely it is that it will be observed. Similarly, the more close neighbors a galaxy has, the more the competition for a fiber, reducing the probability of observation of the given galaxy. This algorithm was conceived by Davide Bianchi and was computationally implemented by him and me together. We summarize the algorithm below:

- To start, we require catalogs of all targets that were 'potential assignments' and values of $n_{\rm tile}$ for each target. From the FiberAssign code, the properties of each scenario where a DESI target is potentially assigned are determined. Concatenating these gives all the tiles where a target could have been assigned, thus giving us $n_{\rm tile}$.

- A friends-of-friends (FOF) algorithm is run on the potential assignments. The linking length (or rather, angle) θ_{LL} is, intuitively, the scale beyond which probabilities of assignment are independent. For galaxies that are closer than this scale, the assignment probabilities are correlated.

- For each galaxy, the number of 'friends' within angular separation $\theta < \theta_{LL}$ is counted. This is denoted by n_{CFC} and is called the 'close friend count' (CFC). n_{CFC} is a measure of how many other galaxies are competing with the current galaxy of interest for the same fiber.

- In the training process, for each combination of n_{CFC} and $n_{\rm tile}$, the ratio of the number of galaxies that got assigned to the total number of galaxies in the full sample is calculated. This is the learned function $\mathscr{F}(n_{CFC}, n_{\rm tile})$, an estimate of the individual probability of assignment for a galaxy.

- Given a test (usually mock) sample, we run the catalog through the same initial process to get $n_{\rm tile}$ and n_{CFC}. We choose the number $N_{\rm bits}$ of realizations to create. Then we randomly sample each galaxy in the test sample $N_{\rm bits}$ times with its own individual probability estimated from $\mathscr{F}$.

- Optionally, a chosen amount of correlation (or correspondingly, anticorrelation) can be injected on small scales to tune the output catalogs.

Two galaxies compete for the same fiber only if they are both inside the patrol radius of said fiber. But because the patrol radii of different fiber positioners overlap, there are actually series of galaxies that are 'linked' together in this competition. These groups of galaxies are encoded in the FOF information that the pre-processing stage of the emulator executes. This fiber assignment anticorrelation thus cannot extend beyond scales larger than a FOF group, allowing each group to be analyzed separately.

To account for this anticorrelation, we can tune the final outputs. Each galaxy has $N_{\rm bits}$ assignment realizations and these are encoded in bitwise arrays. Any operation that

does not change the numbers of zeros and ones in an array (even if it permutes them) does not alter the individual assignment properties but can change collective properties such as the assignment correlation. This idea is at the heart of the anticorrelation algorithm we use. We start by setting a group of galaxies linked via FOF into a state of maximum anticorrelation. This means that the bitweights are such that the corresponding pair count, when averaged over all realizations, is the lowest compatible with the number of bits per bitweights (that is, the given individual probabilities). A schematic of this is shown in Figure 3.1.

To summarize this maximum anticorrelation process, starting with a group n_F ordered galaxies in a FOF group, we begin with galaxy 1. For this, we count the bits in each realization. Then we advance to galaxy 2. For this, we permute its bits randomly, giving priority to the realizations where there are no galaxies yet. These are the zeros from galaxy 1. This process is repeated for all galaxies, giving the state of maximum anticorrelation. Finally, we can adjust how much anticorrelation we want to keep by randomly swapping ones with zeros inside the bitwise arrays. This is essentially reversing some of the anticorrelation that we injected by modifying a free parameter $f_{\rm anti}$).

We also have two free parameters that can be used to modify $\mathscr{F}$: β_0 adjusts its amplitude linearly along the n_{CFC} direction and β_1 tunes the slope to move the kernel. These can be used to modify the learned function to get the right number density in mocks that go through the FFA process when needed. Finally, once a catalog has been run through this process, we can obtain individual weights for each galaxy using Equation 3.7 with $K = N_{\rm bits}$.

3.3 Mock Catalogs

Within DESI, and especially in the context of FFA, we are mainly concerned with two kinds of mock samples: `AbacusSummit` and `EZmocks`. `AbacusSummit` mocks are a set of high-fidelity N-body simulations [45, 64]. They feature density field and dark matter

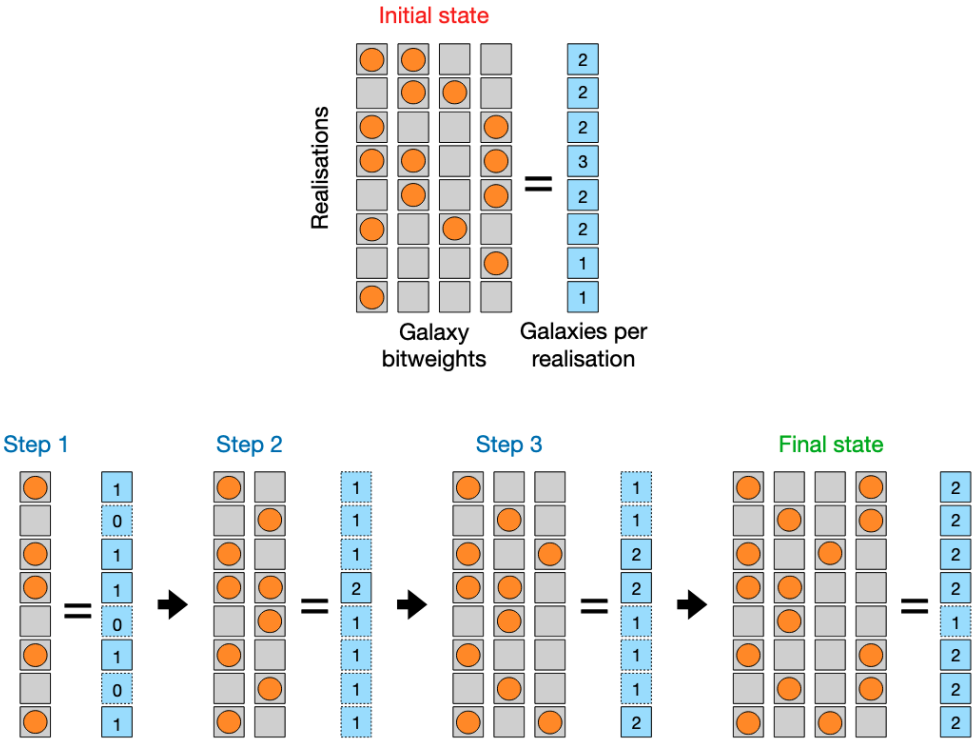


Figure 3.1: Schematic for visualizing the maximum anticorrelation algorithm for bitweights. Each of the columns under the 'step' heading corresponds to a galaxy. For each galaxy, the rows indicate whether it was assigned (orange circle) or not (blank box) on any given realization. Figure reproduced from [10].

halos in cubic boxes with 97 different cosmologies. DESI uses the 25 base simulations with the Planck2018 ΛCDM cosmology [5]. Each contains 6912^3 particles in a cubic box of side length 2 h^{-1} Gpc, corresponding to a particle mass of 2.1×10^9 $M_\odot/h$. For each tracer type, mocks are built at different primary redshifts following a halo occupation distribution (HOD) framework tuned to different datasets, depending on the version [86, 105]. For the mocks that we use in this dissertation, the HOD has been tuned to the DESI Early Data Release clustering. These are called the second generation `Abacus` mocks. The snapshots used for each tracer are:

- For LRGs, the snapshots used are $z = 0.5, 0.8$ and 1.1 for the intervals $z < 0.6$, $0.6 < z < 0.8$ and $z > 0.8$, respectively.
- For ELGs, the snapshots used are $z = 0.95$ and 1.325 for the intervals $z < 1.1$ and $z > 1.1$, respectively.
- For QSOs, the snapshot used is $z = 1.4$ for all redshifts.

These cubic mocks are converted to CutSky mocks by converting Cartesian coordinates into sky coordinates and incorporating redshift-space distortions. This results in galaxy catalogs that are already in redshift space. These are resampled to match the $n(z)$ density distribution and footprint of the DESI data. `Abacus` mock catalogs thus represent high-fidelity, accurate mocks that are used to test the DESI analysis pipeline at all stages. Their detail means they can only be generated in limited quantities and thus for the purpose of calculating clustering covariance matrices, we require a different approach.

`EZmocks` ('effective Zel'dovich' mocks) are approximate mocks generated using a dark matter density field based on the Zel'dovich approximation [26, 107]. For the 'dark time' tracers (LRGs, ELGs and QSOs), DESI uses 2000 cubic boxes of side length 6 h^{-1} Gpc to create 1000 mocks for each of the North and South Galactic Caps. This large number of mocks is useful for calculating the mock-based covariance, as described in Chapter

2. Each box contains 1728^3 dark matter particles from independent Gaussian initial conditions. The galaxy fields are generated from the bias model tuned to two-point clustering of the DESI Abacus HOD mocks. The different snapshots are concatenated into a single pseudo-lightcone into which the DESI Y1 footprint and FFA are applied for realism. For more details, we encourage the reader to refer to [108].

These `EZmocks` are much less expensive to generate than `Abacus` simulations and thus thousands of them are used for the purposes of estimating covariance matrices of the two-point correlation function and the power spectrum.

3.4 Adapting FFA for Mock Catalogs

A major part of the work I did was developing an efficient pipeline for the application of FFA on DESI mock catalogs. This means both the `Abacus` and `EZmocks` mock sets. The FFA algorithm described previously in this chapter is first presented in [10]. I worked with the lead author of that paper, Davide Bianchi, to refine the algorithm to produce a pipeline that would apply this process to DESI simulations to increase realism in the mock catalogs.

At the start, I commenced testing FFA by training it on real DESI data so that it could learn the two-dimensional function dependent on CFC and n_{tile} that would be used to assign galaxies in test datasets. Before any mocks could be run through the emulator, they needed to be processed so they had all the information needed. Specifically, I had to apply hardware vetoes and add n_{tile} information.

`Abacus` mocks for LRGs, QSOs and ELGs were then run through this emulator and their pre and post-assignment pair counts were compared in plots similar to Figure 3.7. What I had found was that the collision window was not accurately captured by FFA and this led to Bianchi and I coding up the anti-correlation algorithm in Figure 3.1 to better model the collision window, which gave us the outputs we needed — post-assignment target counts that matched real data and better looking collision windows.

Specifically, for BGS targets, I found that FFA was assigning too many galaxies compared to real data. This was because the pre-assignment mocks were less dense than real data so I needed to suppress the high n_{CFC} part of the kernel to avoid assigning too many galaxies. The details of this kind of modification of the learned kernel are discussed in detail in Chapter 4.

After these tests, once the emulator was ready to process all mocks, I processed 25 `Abacus` mocks and 1000 `EZmocks` through it, for each of the dark time target classes. I also ran code to calculate the two-point correlation function for these to be able to see that clustering was accurately simulated before DESI could use these further down the line analysis pipeline to extract mock-based covariance matrices for clustering.

3.5 Comparing altMTL and FFA Mock Catalogs

Figure 3.3 shows the angular assignment completeness for the dark-time galaxies (LRGs, ELGs and QSOs) obtained via the altMTL method. This is in keeping with the results from actual data, shown in Figure 3.2. Figure 3.3 also shows the relative angular density ($\rho_{\mathrm{FFA}}/\rho_{\mathrm{altMTL}} - 1$) comparing the two fiber assignment simulation approaches, where ρ_i is the density of method i. The angular distributions are compatible among themselves. The regions with lower completeness reveal larger statistical fluctuations but no clear signs of systematic trends. In Figure 3.4, the redshift distributions of different tracers are presented. There is good agreement between the altMTL and FFA results. Rescaling the post-assignment redshift distributions to match the amplitude of the pre-assignment ones gives very good agreement. Fiber assignment does not noticeably change the spread of redshifts then. The only mismatch happens for ELGs at redshifts above 1.5. This is from intentionally incorporated redshift failures that are inserted after the fiber assignment process. This is to account for how, in real data, our ability to accurately measure redshifts above 1.5 is limited by sky-line contamination. Figure 3.5 shows the distributions of fiducial weights obtained via altMTL and FFA.

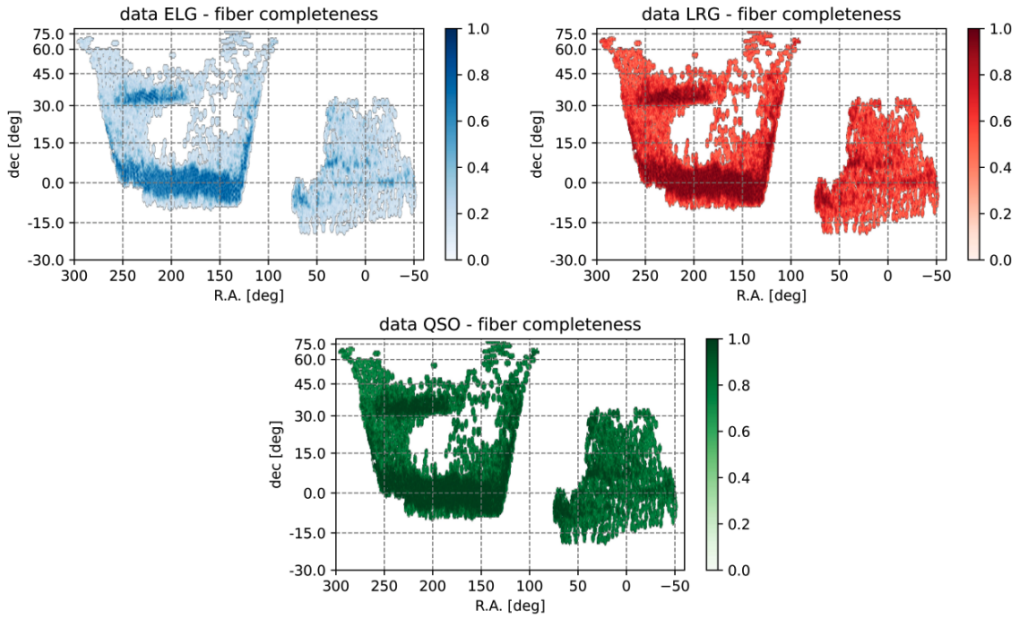


Figure 3.2: The angular assignment completeness of the dark-time galaxies observed by DESI in its first year. This is defined as the number of tracers assigned to a fiber divided by the total number of tracers in the corresponding full sample. Figure from my work in [10].

The two weight definitions are similar but distinct: The fiducial weights for altMTL are derived from completeness considerations while for FFA, they are the individual inverse probability (IIP) weights. The FFA weights are not constrained to take discrete integer values so to give a better sense in the comparison, we also plot the rounded versions of the FFA weights. We see satisfactory agreement, given the different definitions. In the ELG case, the altMTLs showcase a shorter tail at high weights compared to FFA. These differences may arise from variations in how fiber assignment is performed. This is evidenced in [10], where they show great agreement between the altMTL and real data weights, showing that keeping the fiber assignment strategy fixed between catalogs shows better consistency. Figures 3.6 and 3.7 show the collision windows from the altMTL and FFA mocks. The altMTLs show results that are completely compatible with those from real data [10]. Meanwhile, FFA shows a less marked collision window. There is a less steep transition and less overall depth, especially for QSOs. There is

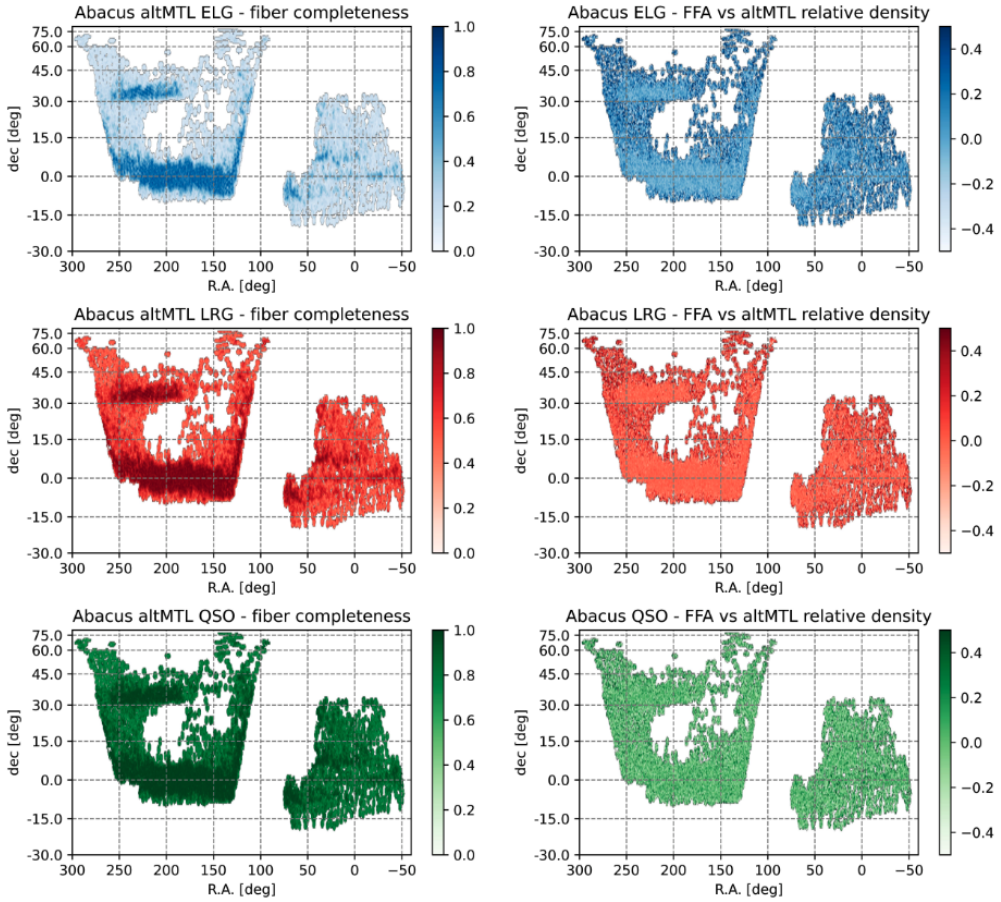


Figure 3.3: The left column of this figure shows the angular assignment completeness of the three dark-time tracers, measured from altMTL mocks. The right column shows the relative behavior of FFA and altMTL. Figure from my work in [10].

qualitative agreement between the two, especially in the window's angular size. Note that the FFA emulator can be modified to adjust the assignments in such a way that we get better agreement in the depths of the collision windows, as will be explored in Chapter 4.

3.6 Impact

Fast fiber assignment and the FFA emulator have been very good successes. The previous section summarized some results that show how various outputs from FFA

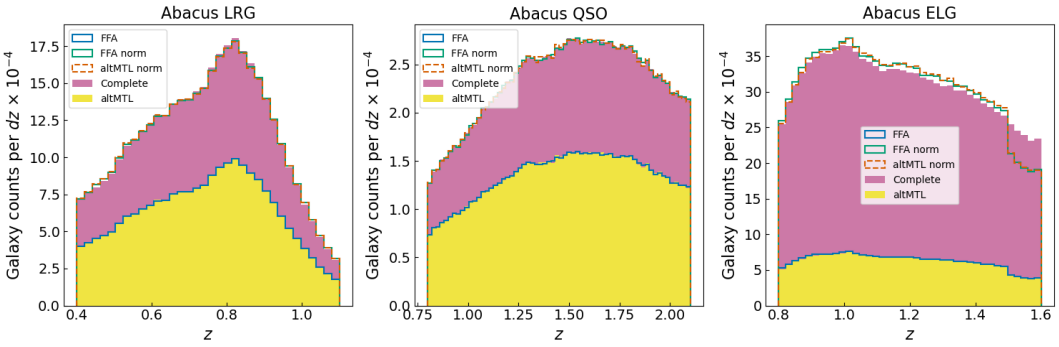


Figure 3.4: A comparison of the redshift distributions ($dz = 0.02$) from various versions of an Abacus mock, with and without fiber assignment run. The 'norm' versions are the simulated altMTL and FFA histograms rescaled by $N_{\text{complete}}/N_{\text{assigned}}$ to match the amplitude of the complete sample. Figure adapted from my work in [10].

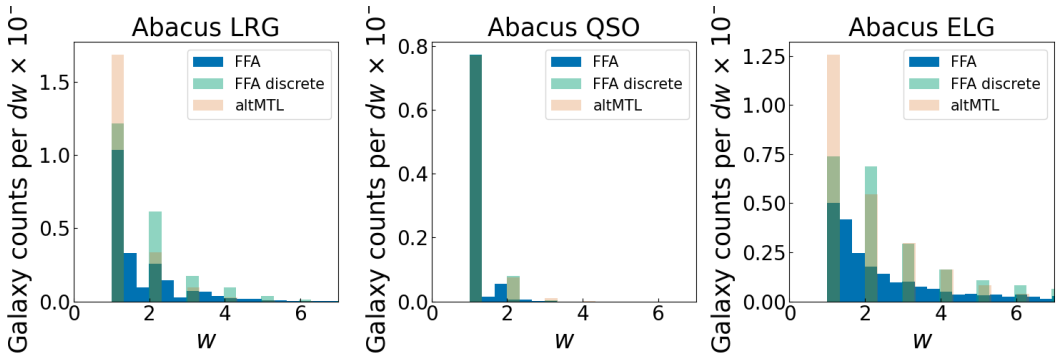


Figure 3.5: Spread of the fiducial weights from the altMTL and FFA mocks, using $dw = 0.33$. The 'discrete' weights are the fiducial FFA weights rounded to the nearest integer. Figure adapted from my work in [10].

		altMTL	FFA	Complete
ELG	Galaxies Assigned	2.41×10^6	2.44×10^6	12.09×10^6
	Fraction Assigned	0.20	0.20	1
LRG	Galaxies Assigned	2.16×10^6	2.15×10^6	3.88×10^6
	Fraction Assigned	0.56	0.56	1
QSO	Galaxies Assigned	0.86×10^6	0.86×10^6	1.49×10^6
	Fraction Assigned	0.58	0.58	1

Table 3.1: The total number of galaxies measured from a sample Abacus mock. The 'complete' sample is before fiber assignment. The 'fraction assigned' refers to the ratio between the counts of galaxies assigned and the total counts of galaxies in the complete sample. Table from [10].

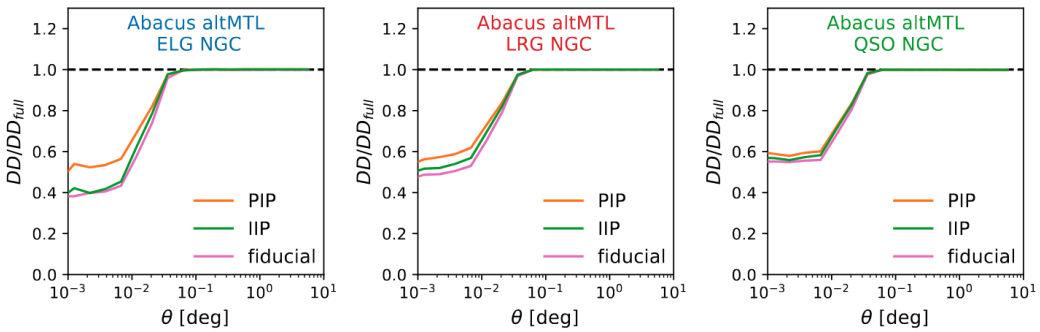


Figure 3.6: Normalized pair counts from altMTL mocks with different kinds of weights divided by normalized pair counts from the full sample of targets. Figure from [10].

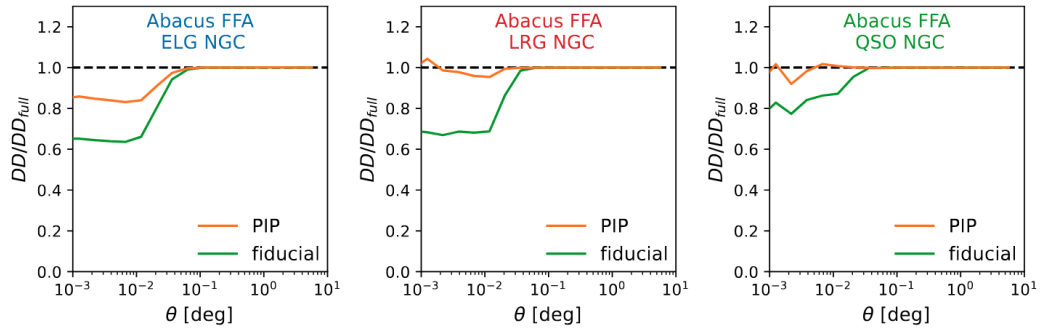


Figure 3.7: Normalized pair counts from FFA mocks with different kinds of weights divided by normalized pair counts from the complete sample of targets. Figure from [10].

compare to realistic altMTL catalogs. FFA demonstrated a huge time improvement over running a complex algorithm to apply fiber assignment to mocks. Whereas previously, a single catalog of galaxies may take up to a day to be processed, FFA is able to simulate fiber assignment effects in ten to twenty minutes, depending on the size of the catalogs, QSOs being the fastest and ELGs the slowest.

The `Abacus` mocks that were processed via FFA have been used by the DESI collaboration for the purpose of testing the analysis pipeline. The FFA `EZmocks` have made a direct impact to the final DESI cosmology results [3, 4]. In configuration space, for the Y1 analysis, DESI used analytical covariance matrices [82, 83]. These were validated using the mock-based covariance from FFA `EZmocks`. In Fourier space, the FFA

`EZmocks` covariance matrix was directly used for cosmological analysis, leading to constraints down the line on cosmological parameters, which will be reviewed in Chapter 5.

To capture the effect of the work I have done in contributing to cosmology constraints, I briefly summarize the process of how FFA covariances contribute to the final cosmology results:

1. I created FFA `EZmocks` catalogs from which the clustering statistics (power spectrum and two-point correlation function) were calculated.

2. The mock-based covariance of the power spectrum was calculated. This covariance matrix was used to fit the power spectrum of real DESI data to extract the dilation parameters in Equations 2.1 or 2.2. The quality of these fits (or the error bars on them) are directly affected by the quality of the mock-based covariance mentioned before.

3. From these dilation parameters, the cosmological distances in their definitions were extracted and Markov Chain Monte Carlo (MCMC) methods were used to calculate the final cosmological quantities such as the Hubble constant or the matter density parameter Ω_m. The uncertainty from the clustering covariance matrix therefore propagates all the way down here.

4. Figure 5.3 is an example of the kinds of plots that are used to display cosmological constraints. The error on the measurement is captured by the size and shape of the ellipses.

This summarizes how the development of FFA directly contributed to the final results for DESI Y1 cosmology analysis.

CHAPTER 4

Fast Fiber Assign: Further Investigations

To start, I note that all the tests and results presented in this chapter were performed by me and have not been published before. Recall from Chapter 3 that the Fast Fiber Assign (FFA) emulator has a set of parameters that govern its behavior. The major ones are listed below. The names in parentheses are the names that the respective parameter is given in the code. We refer to the learned function by $\mathscr{F}(n_{CFC}, n_{\text{tile}})$.

1. The linking angle (`linklen`), interchangeably referred to as the linking length, is the scale that defines the maximum separation two galaxies can have to be considered part of the 'close friends' groups introduced in the preceding chapter.

2. β_0 is used to fine-tune $\mathscr{F}$ by scaling its amplitude linearly along the n_{CFC} direction.

3. β_1 is used to fine-tune $\mathscr{F}$ by scaling its slope.

4. The anti-correlation parameter f_{anti} (`deco`) is used to implement the anti-correlation explained in Chapter 3. To summarize, we use an algorithm (Figure 3.1) to put the bitweights from FFA into a state of maximum anticorrelation. Then f_{anti} controls what fraction of bits are randomly swapped, to effectively undo part of the imposed anticorrelation.

For the first production run of the FFA emulator that was used for the DESI first year data analysis, each of these parameters had a fiducial set of values chosen. Because

of time constraints, the main factor in picking the numbers used was whether they reproduced the same number counts of galaxies assigned as the real data. There were some tests that established that the `deco` parameter was needed but no detailed studies of the effects of varying it on the outputs were conducted. This was especially true for the dark time tracers. The fiducial choices made were $\{\beta_0, \beta_1, f_{\rm anti}\} = \{1.0, 0.0, 0.1\}$.

Later application of the FFA method to BGS targets required modifying the FFA kernel using the parameters mentioned before to match the assigned mock number counts to the data. This was necessary because the pre-assignment mocks were less dense than real data so we needed to suppress the high $n_{\rm CFC}$ part of the kernel to avoid assigning too many galaxies. Moreover, mock-based covariances from `EZmocks` have been found to underestimate errors by around 10% even though the clustering between the data and mocks matches [1]. These have provided further impetus for exploring how the FFA emulator behaves under different parameter constraints for different tracers.

4.1 Methods

To study the aforementioned effects, I use a set of 25 high-fidelity `Abacus` mock catalogs, as described in Chapter 2. These are indexed from 0 to 24. I run this set through the FFA pipeline many times, each with a varied set of emulator parameters to generate a large suite of fiber-assigned catalogs. This is done for the LRG and QSO tracers. I do not work with ELGs for this analysis because of the much larger size of those catalogs which would make this analysis much more expensive in terms of time and computational cost.

Henceforth, I will refer to each set of 25 mocks that has undergone the FFA procedure with a certain set of parameters an 'emulated run'. The first check of realism is seeing which emulated runs result in the right number counts of targets getting assigned. I also analyze the collision window in each of the emulated runs to select the runs that match those of the real data (and some that don't, to see how those change the final

results). These selected emulated runs then go through the clustering pipeline, giving us the configuration space two-point correlation function. I utilize the FitCov procedure for calculating jackknife covariances as explained in 2.2.1.1. This gives the estimated covariance matrices for each emulated run.

To compare the effects of different emulation iterations on the covariance and to connect our results to cosmology, I utilize the covariance matrices to perform cosmological fits. In particular, I use the ShapeFit technique (see 2.1.4). To briefly recall, this is a form of 'full shape' analysis, where a template power spectrum or correlation function evaluated at a template cosmology is used to fit data. The 'full shape' part implies that there is sensitivity to the broadband shape of the clustering statistic, as opposed to in the classic BAO analysis.

I picked full shape analysis as my mode of fitting so that I was able to study the effect of the covariances on a procedure broader than just analyzing the BAO peak. I specifically use ShapeFit because the fits are computationally more economical to run than a direct fit approach, as explained in Chapter 2. Still, the computation of the perturbation theory integrals required to predict the clustering statistic is not entirely computationally cheap and thus I use an emulator to speed this process up. Specifically, I use one based on a Taylor series expansion that is used to improve the calculation of the clustering multipoles [23].

For my analysis, for each emulated run I am analyzing, I use the FitCov covariance matrix to fit 250 `EZmocks` and then report the compression parameters $\alpha_{\rm iso}$ and $\alpha_{\rm AP}$ (Equation 2.2). I also report $df = f/f^{\rm ref}$, where 'ref' is a reference cosmology (see Equations 2.5 and 2.7).

4.2 Results and Discussion

4.2.1 Effect on Kernel

Initially, I ran some emulated runs with a wide range of values for the parameters mentioned in the preceding section. I especially looked to see how the shape of the emulator's learned kernel was affected by different values of β_0 and β_1. See Figure 4.1 for a visualization of the effects. The leftmost plot shows the function as learned from real DESI data for LRGs. The brighter the colors, the higher the selection probability for a target with that combination. Note that there are some scattered regions of non-zero probability for combinations of high NTILE and high CFC values. This is to be expected because while a greater CFC value means that a given target has more competition from its neighbors, a higher NTILE value means that there are more passes of the instrument at that target's location, increasing the odds that it will be observed.

The middle plot shows what happens when I set $\beta_1 = -0.1$. This tilt parameter changes the slope of the kernel along the CFC side. Values farther out on the NTILE scale are more affected by the tilt, making the overall kernel 'smoother' by 'pulling in' the few scattered non-zero selection probability regions corresponding to high NTILE and high CFC targets. The rightmost plot shows what happens when I set $\beta_0 = 1.2$ to rescale the original kernel in the leftmost plot. This stretches the kernel in the vertical CFC direction, increasing the assignment probability for targets across the NTILE range.

Note then that the values of the two β parameters predictably change the number of targets that will be fiber assigned because we can visualize the way the kernel is affected easily. Note also that the f_{anti} (or `deco`) parameter will not affect the assignment counts by construction. Table 4.1 shows the counts of assigned targets for LRGs and QSOs in the data catalogs that were used to train the FFA emulator, and it is these that I wish to match for `Abacus` mocks as well. After many tests that excluded certain ranges of the emulator parameters because of assignment counts very disparate from the data, I

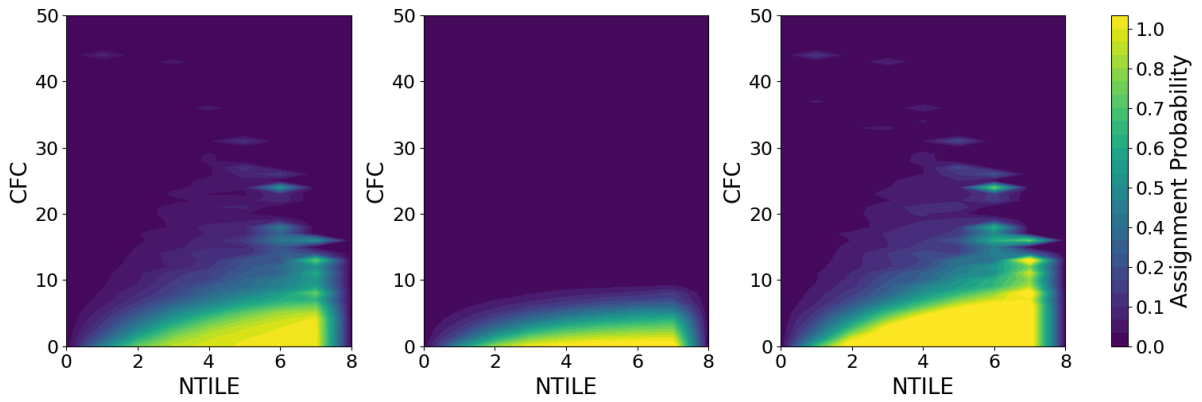


Figure 4.1: This figure contains three subplots that visualize what the kernel that the emulator (trained on real LRG data) learned looks like, with the color representing the probability of assignment. The first figure on the left shows the unmodified version of the kernel. The middle figure shows the same kernel with $\beta_1 = -0.1$ and the right figure shows the first kernel with $\beta_0 = 1.2$.

Tracer	Assigned Counts (Data)
LRG	2 778 582
QSO	2 212 960

Table 4.1: The numbers of assigned counts for different tracers in the data catalogs that were used to train the FFA emulator.

ran the tests I present here on combinations of parameter values within the following ranges: I used β_0 values between 0.8 to 1.2, β_1 values between -0.3 to 0.3 and f_{anti} values between 0.025 to 0.4. Figures 4.2 and 4.3 show how the assignment counts for `Abacus` mock 0 vary depending on the emulator parameters. Observe again that there is no assigned length dependence on f_{anti}, and that larger values of β_0 and more positive values of β_1 both contribute to increasing the assignment counts. For each tracer, I reduced our set of emulated runs to the subset of runs where the assignment counts were within 10% of the counts of the training data. Figures 4.4 and 4.5 show the ratios of the number counts of assigned pairs to the number counts of pairs in the complete sample as a function of the separation angle for LRGs and QSOs, respectively. I wished to select only the runs where the pair count ratios within the collision window agree with that of the altMTL mocks (Figure 3.6). This is because the altMTL mocks represent a more expensive but accurate simulation of fiber-assignment that I would like FFA results to

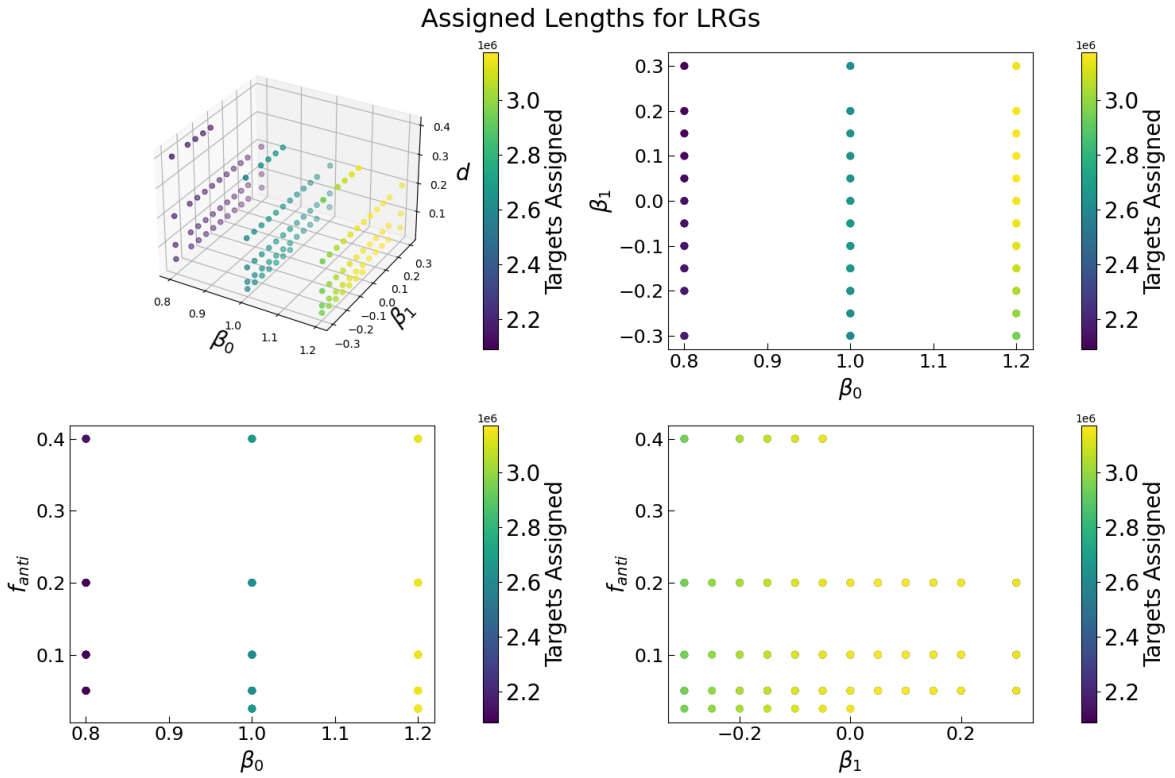


Figure 4.2: This figure shows the distribution of the number of assigned LRG targets for a single `Abacus` mock catalog when run through the emulation pipeline with different parameters. Note that some of the points are super-imposed with multiple data points, but the dependence on different parameters is clear.

get closer to.

For each tracer, I selected a subset of five emulated runs that match the aforementioned criteria, along with what I call the 'fiducial' set. This latter set is the emulated run with the same parameters as used for the DESI Y1 analysis, that is, where $\beta_0 = 1$, $\beta_1 = 0$ and $f_{\rm anti} = 0.1$. Tables 4.2 and 4.3 list the selected parameter sets that I use for the rest of my analysis. Note that there are indeed more runs that fit these criteria but the choices of emulator parameters for each of them are not very different from at least one of the selected combinations.

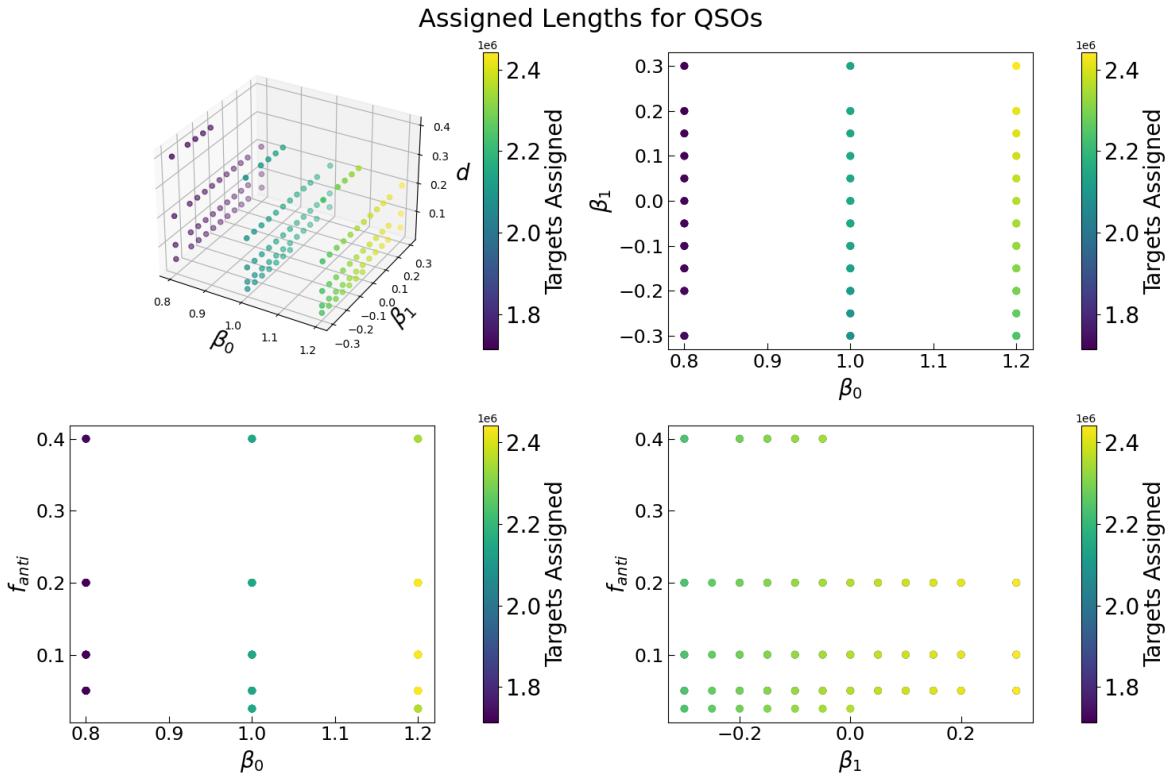


Figure 4.3: This figure shows the distribution of the number of assigned QSO targets for a single `Abacus` mock catalog when run through the emulation pipeline with different parameters. Note that some of the points are super-imposed with multiple data points, but the dependence on different parameters is clear.

β_0	β_1	f_{anti}
1.0	-0.3	0.1
1.0	-0.2	0.05
1.2	-0.3	0.1
1.0	-0.05	0.025
1.0	-0.1	0.4

Table 4.2: The parameter combinations for LRG emulated runs used in our analysis.

β_0	β_1	f_{anti}
1.0	-0.3	0.025
1.0	-0.3	0.1
1.0	-0.25	0.05
1.0	-0.2	0.1
1.2	-0.3	0.025

Table 4.3: The parameter combinations for QSO emulated runs used in our analysis.

4.2.2 Clustering Results

Figures 4.6, 4.7 and 4.8 show the clustering of LRGs across my selected emulated runs for `Abacus` mock 0. Figure 4.9 shows the same for QSOs. Specifically, I plot the monopole and quadrupole of the configuration space two-point correlation function as a function of the separation s. The bottom panel of each subplot shows the difference between each of the functions and the corresponding values for the fiducial run, the latter of which is plotted on the bottom-right of each figure. This is in units of the standard deviation of the fiducial run. The error bars are the standard deviation calculated from 50 jackknife realizations for each mock.

In general, there is good agreement between the clustering, both by eye and by looking at the lower insets within each subplot. For LRGs, the first three panels show a decrease in clustering relative to the fiducial catalog on small scales. This is consistent with how the first three catalogs were from groups where the numbers of pair counts within the collision window were lower than that of the fiducial catalog. I see similarly predictable agreement between the clustering for QSOs. On small scales, all the non-fiducial emulations have clustering suppressed relative to the fiducial choice because the pair counts on small scales for the latter were on the higher side relative to data and altMTL mocks.

4.2.3 Normed Trace Comparisons

In Tables 4.4 and 4.5, I present the normed trace comparisons between different covariance matrices. This was utilized in [82] for quick comparisons between covariance matrices. For two matrices, Cov_A and Cov_B, this quantity, denoted by $\mathrm{tr}(\mathrm{Cov}_A, \mathrm{Cov}_B)$ is computed as

$$\mathrm{tr}(\mathrm{Cov}_A, \mathrm{Cov}_B) = \frac{\mathrm{Trace}(\mathrm{Cov}_A \mathrm{Cov}_B^{-1})}{N} \tag{4.1}$$

where N is the length along one dimension of the matrices, used as a normalizing factor. This value allows for an assessment of how similar or dissimilar the covariance structures are under different configurations by seeing how they deviate from 1.

Using a normed trace comparison test, I found that the fiducial FFA FitCov covariance matrix had roughly twice the covariance than the corresponding `EZmocks` mock-based covariance matrix from the fiducial FFA mocks. I attribute this to the fact that the mock-based covariance matrix is calculated from 1000 mocks while the FitCov covariances I use for my tests are approximate covariances from 25 `Abacus` mocks. Still, the motivation for my analysis was to see how different emulation parameters affect the covariances and hence the quality of the final fits and I was able to get insight into this by studying the relative differences in deviation between different runs regardless.

This is why, in Tables 4.4 and 4.5, I present the 'FFA-fid' comparison between each FitCov covariance matrix under study with the FitCov covariance matrix from the fiducial choice. This is to keep the kind of covariance matrix being studied consistent. For interpretation, a larger than 1 in the 'FFA-fid' comparison would imply larger errors in cosmological fits from the covariance matrix being studied, and vice versa. This is visible when the normed traces are matched to the spread of errors in Table 4.9. This is further explored in the next subsection.

4.2.4 Comparing Fit Parameters

Figures 4.10 to 4.12 show histograms of the compression parameters, $\alpha_{\rm iso}$ and $\alpha_{\rm AP}$, and df from using each of the LRG FitCov covariance matrices I generated to fit 250 `EZmocks`. In each figure, the histogram corresponding to the fiducial emulated run is shown in solid purple. The unfilled histograms with red borders are the corresponding fits done using a semi-analytic covariance from altMTL mocks. I will refer to this as the analytic case for brevity. Figures 4.13 to 4.15 show the corresponding error distributions for each parameter. Tables 4.6 and 4.8 present the mean deviation of the LRG fit

$\{\beta_0, \beta_1, f_{\rm anti}\}$	Comparison	$z \in (0.4, 0.6)$	$z \in (0.6, 0.8)$	$z \in (0.8, 1.1)$
—	altMTL-Data	1.06	1.08	1.02
	altMTL-EZ	1.08	1.12	1.16
{1.0, -0.3, 0.1}	FFA-altMTL	2.92	3.66	2.08
	FFA-fid	2.14	2.65	2.29
$\{1.0, -0.2, 0.05\}$	FFA-altMTL	1.91	2.02	1.25
	FFA-fid	1.34	1.42	1.27
$\{1.2, -0.3, 0.1\}$	FFA-altMTL	2.46	2.69	2.07
	FFA-fid	1.79	1.94	2.19
$\{1.0, -0.05, 0.025\}$	FFA-altMTL	1.55	1.60	1.02
	FFA-fid	1.06	1.09	0.97
$\{1.0, -0.1, 0.4\}$	FFA-altMTL	1.92	1.77	1.31
	FFA-fid	1.31	1.22	1.26
$\{1.0, 0.0, 0.1\}$	FFA-altMTL	1.61	1.60	1.15
	FFA-fid	1.00	1.00	1.00

Table 4.4: Normed trace comparisons between different LRG covariance matrices. 'altMTL' and 'Data' refer to analytical covariance matrices for Abacus altMTL mocks and data, respectively. 'EZ' refers to the mock-based covariance from 1000 `EZmocks` and FFA is the FitCov covariance for each emulation run specified in the first column. 'FFA-fid' refers to the comparison between each FFA FitCov covariance matrix with the FitCov covariance matrix from the fiducial FFA choice. Note this is why the last row is composed of ones. The top row serves as a standard to compare the rest of the covariances to.

parameters and errors for each fit from the analytical version, respectively. For details on this analytical covariance, I refer the reader to [83] and [82]. This deviation ΔP is a mean percent difference for a parameter P and is defined as

$$\Delta P = 100 \left\langle \frac{P_M - P_A}{P_A} \right\rangle$$

where P_M and P_A are the mock and analytical versions of the parameter, respectively. Analogous results for QSOs are also included. I note here then that, generally, when I use the words 'better' or 'worse' as descriptions for results, this is relative to the constraints from using an analytical covariance. 'Better' implies a closer match to the analytical

$\{\beta_0, \beta_1, f_{\rm anti}\}$	Comparison	$z \in (0.8, 2.1)$
—	altMTL-Data	1.02
	altMTL-EZ	1.09
$\{1.0, -0.3, 0.025\}$	FFA-altMTL	1.45
	FFA-fid	1.05
$\{1.0, -0.3, 0.1\}$	FFA-altMTL	2.10
	FFA-fid	1.45
$\{1.0, -0.25, 0.05\}$	FFA-altMTL	2.35
	FFA-fid	1.58
$\{1.0, -0.2, 0.1\}$	FFA-altMTL	1.69
	FFA-fid	1.16
$\{1.2, -0.3, 0.025\}$	FFA-altMTL	1.84
	FFA-fid	1.25
$\{1.0, 0.0, 0.1\}$	FFA-altMTL	1.62
	FFA-fid	1.00

Table 4.5: Normed trace comparisons between different QSO covariance matrices. 'altMTL' and 'Data' refer to analytical covariance matrices for Abacus altMTL mocks and data, respectively. 'EZ' refers to the mock-based covariance from 1000 `EZmocks` and FFA is the FitCov covariance for each emulation run specified in the first column. 'FFA-fid' refers to the comparison between each FFA FitCov covariance matrix with the FitCov covariance matrix from the fiducial FFA choice. Note this is why the last row is composed of one(s). The top row serves as a standard to compare the rest of the covariances to.

covariance matrix fits. I note that the error deviations in Tables 4.8 and 4.10 are large numbers. This is in contrast to Tables 4.6 and 4.7 where the parameter value deviations are in general below a percent. I ascribe this to my use of approximate covariance matrices from a very small number of mocks, as referenced in the previous subsection as well. I especially note that $\Delta\sigma_{\alpha_{\rm AP}}$ tends to have much larger deviations compared to the other dilation parameter. Especially, for example, for the set $\{1.0, -0.3, 0.1\}$, there is a 170% difference in the lowest redshift bin for LRGs. This appears to be a result of the effect we see in Figures 4.11 and 4.14. Note that the distribution of $\alpha_{\rm AP}$ is not Gaussian. In fact, there are many cases where it pushes up against the prior bounds -0.8 and 1.2, creating peaks there. This also translates into the bumps in the tails of

$\{\beta_0, \beta_1, f_{\rm anti}\}$	Parameter	$z \in (0.4, 0.6)$	$z \in (0.6, 0.8)$	$z \in (0.8, 1.1)$
{1.0, -0.3, 0.1}	$\Delta\alpha_{\rm iso}$	-0.02	-0.12	-0.10
	$\Delta\alpha_{\rm AP}$	0.61	0.00	0.29
	Δdf	13.65	1.94	0.43
$\{1.0, -0.2, 0.05\}$	$\Delta\alpha_{\rm iso}$	-0.16	-0.37	-0.26
	$\Delta\alpha_{\rm AP}$	-0.50	0.72	0.10
	Δdf	2.20	-0.92	0.27
$\{1.2, -0.3, 0.1\}$	$\Delta\alpha_{\rm iso}$	-0.11	-0.13	-0.33
	$\Delta\alpha_{\rm AP}$	-0.97	-0.09	0.15
	Δdf	4.18	1.97	-0.48
$\{1.0, -0.05, 0.025\}$	$\Delta\alpha_{\rm iso}$	0.12	-0.18	-0.19
	$\Delta\alpha_{\rm AP}$	-0.35	-0.01	0.68
	Δdf	1.96	1.54	0.75
$\{1.0, -0.1, 0.4\}$	$\Delta\alpha_{\rm iso}$	-0.37	-0.41	-0.07
	$\Delta\alpha_{\rm AP}$	-0.30	0.90	0.45
	Δdf	0.53	0.76	0.49
$\{1.0, 0.0, 0.1\}$	$\Delta\alpha_{\rm iso}$	-0.21	-0.30	-0.29
	$\Delta\alpha_{\rm AP}$	-0.44	1.01	0.13
	Δdf	1.36	-0.06	0.65

Table 4.6: Mean deviation of LRG fit parameters from 250 `EZmocks`, using different FitCov covariance matrices.

the corresponding error histograms. To test this, I removed the fits where $\alpha_{\rm AP}$ was at the boundary to present the results in Table 4.11. I see that both the parameter and error deviations do indeed get smaller. Other works have also noticed similar behavior in the lowest redshift bin for LRGs, where its statistics differ compared to those of the other bins when calculating covariance matrices [8]. So it may be that this is a feature of the `EZmocks` themselves.

Moreover, the $\Delta\sigma_{\alpha_{\rm AP}}$ tends to be larger than $\Delta\sigma_{\alpha_{\rm iso}}$. As I run all fits on the same set of 250 `EZmocks` each time, any differences should be traceable back to the covariance matrices themselves and thus to the clustering from which they are calculated. This is evident in the connection between the normed traces and the error deviations — larger differences from 1 for the traces correspond to worse constraints, as expected.

$\{\beta_0, \beta_1, f_{\text{anti}}\}$	Parameter	$z \in (0.8, 2.1)$
$\{1.0, -0.3, 0.025\}$	$\Delta\alpha_{\text{iso}}$	0.07
	$\Delta\alpha_{\text{AP}}$	0.42
	Δdf	-0.33
$\{1.0, -0.3, 0.1\}$	$\Delta\alpha_{\text{iso}}$	-0.33
	$\Delta\alpha_{\text{AP}}$	0.65
	Δdf	0.66
$\{1.0, -0.25, 0.05\}$	$\Delta\alpha_{\text{iso}}$	-0.42
	$\Delta\alpha_{\text{AP}}$	0.34
	Δdf	0.07
$\{1.0, -0.2, 0.1\}$	$\Delta\alpha_{\text{iso}}$	-0.14
	$\Delta\alpha_{\text{AP}}$	0.14
	Δdf	0.30
$\{1.2, -0.3, 0.025\}$	$\Delta\alpha_{\text{iso}}$	-0.10
	$\Delta\alpha_{\text{AP}}$	0.03
	Δdf	0.47
$\{1.0, 0.0, 0.1\}$	$\Delta\alpha_{\text{iso}}$	0.05
	$\Delta\alpha_{\text{AP}}$	0.80
	Δdf	0.23

Table 4.7: Mean deviation of QSO fit parameters from 250 `EZmocks`, using different FitCov covariance matrices.

Recall the concept of redshift-space distortions (RSD). In the linear regime, peculiar velocities produce a coherent in-fall towards overdensities. As a result, the clustering multipoles gain scale-dependent modulations that enhance or suppress clustering amplitudes depending on the angular projection. On small scales, peculiar velocities due to gravitational effects from nearby structures cause elongation of structures along the line-of-sight. This is consequently visible in the quadrupole of the two-point correlation function on scales less than where the BAO peak is located. It manifests as a 'turnover' where the decreasing quadrupole signal turns around, showing a bump.

For example, in Figure 4.6, we can see how the turnover scale in the first plot (a little to the right of 50 h^{-1} Mpc) is different than in the fiducial clustering. This is also visible in the bottom panels of each subplot in the figure, where the difference between

each respective correlation function and the fiducial choice is plotted. At small scales, we see large deviations. I believe it is this effect that is being captured when we see the relative magnitudes of $\Delta\sigma_{\alpha_{\rm AP}}$ and $\Delta\sigma_{\alpha_{\rm iso}}$, as $\Delta\sigma_{\alpha_{\rm AP}}$ is more strongly influenced by RSD since this is where the anisotropy is recorded.

We can also see the behavior of σ_{df} by looking at the clustering. f, the growth rate of structure, affects the amplitude of clustering. Larger $\Delta\sigma_{df}$ values correspond to runs where the clustering amplitudes in each bin deviate more from the fiducial run over the full shape of the correlation function, once again visible in the lower panels of the clustering subplots.

In trying to draw connections between the emulator parameter sets I used and the clustering I get, in general, larger (more negative) values of β_1 lead to worse constraints. This is most evident in comparing the $\{1.0, 0.0, 0.1\}$ and $\{1.0, -0.3, 0.1\}$ runs for QSOs. Also, $f_{\rm anti}$ exerts very strong effects on the final outcomes. Smaller values of $f_{\rm anti}$ give much better constraints overall. For QSOs, when I go from the $\{1.0, -0.3, 0.1\}$ case to $\{1.0, -0.3, 0.025\}$, there is a significant improvement in the trace (roughly a 45% increase). This is also visible for LRGs when going from the fiducial case to $\{1.0, -0.05, -0.025\}$. For QSOs, the $\{1.0, -0.3, 0.025\}$ case actually improves upon the constraints relative to the fiducial case, despite having a larger β_1 value. The much reduced $f_{\rm anti}$ appears to dominate the effect.

The detrimental effect of larger β_1 values in general should stem from the strong warping they can introduce to the learned kernel (see Figure 4.1). The effect of large values of β_1 can be compensated for by smaller values of $f_{\rm anti}$, that is, by re-injecting less correlation on small scales. This makes up for the narrowing of the kernel due to warping that would otherwise induce more correlation at those scales. QSOs are less dense than LRGs. This leaves more leeway to adjust β_1 to more negative values for QSOs because doing so suppresses the high $n_{\rm CFC}$ tail of the kernel, which affects the denser LRGs more. I also observe that very small values of β_1 can actually be beneficial

to the analysis. The strongest example of this is the $\{1.0, -0.05, 0.025\}$ set for LRGs. My view is that this comes from the effect we see in Figure 4.1. The unmodified kernel has islands of higher probability in the high CFC and high NTILE regions. I believe this serves as noise when emulating new catalogs. Small values of β_1 smooth the kernel without warping it too much.

The numbers of galaxies assigned for both tracers peak at small values of β_1 and fall off on either side. I confirmed that having the numbers of assigned galaxies be closer to the data counts (Table 4.1) indeed gave better constraints.

Ultimately, I find that for LRGs, the $\{1.0, -0.05, 0.025\}$ set gave the best constraints on average . This is also the set whose small scale collision window matched the data best and, from the normed trace comparison, is a ten percent larger covariance matrix than the fiducial choice. The fiducial set had more pairs assigned within the collision window than the former. This is in keeping with my observation that smaller values of f_{anti} give the best constraints — we do not need to inject too much correlation on small scales when running the emulator.

For QSOs, I find that the $\{1.0, -0.3, 0.025\}$ set gives constraints much better than the fiducial case. This was a surprising result because the large negative value of β_1 warps the kernel noticeably. As mentioned before, there is certainly more leeway for doing so in the case of QSOs as they are less densely distributed than LRGs. From the normed trace comparison, this set gives a covariance matrix roughly five percent larger than the fiducial choice. Once again, the small value of f_{anti} looks to have the most pull. This set, too, corresponds to the best match to the data in terms of assigned pair counts within the collision window.

I recommend for future analyses of DESI data that use the fast fiber assignment method to use these parameter choices for the emulator.

$\{\beta_0, \beta_1, f_{\rm anti}\}$	Parameter	$z \in (0.4, 0.6)$	$z \in (0.6, 0.8)$	$z \in (0.8, 1.1)$
{1.0, -0.3, 0.1}	$\Delta\sigma_{\alpha_{\rm iso}}$	59.42	62.45	38.60
	$\Delta\sigma_{\alpha_{\rm AP}}$	169.82	75.91	38.71
	$\Delta\sigma_{df}$	52.77	41.76	37.86
$\{1.0, -0.2, 0.05\}$	$\Delta\sigma_{\alpha_{\rm iso}}$	29.18	36.94	19.42
	$\Delta\sigma_{\alpha_{\rm AP}}$	43.85	33.44	22.88
	$\Delta\sigma_{df}$	32.40	20.48	29.73
$\{1.2, -0.3, 0.1\}$	$\Delta\sigma_{\alpha_{\rm iso}}$	38.85	35.44	36.52
	$\Delta\sigma_{\alpha_{\rm AP}}$	61.94	45.79	32.22
	$\Delta\sigma_{df}$	39.58	34.42	41.38
$\{1.0, -0.05, 0.025\}$	$\Delta\sigma_{\alpha_{\rm iso}}$	22.88	20.54	10.49
	$\Delta\sigma_{\alpha_{\rm AP}}$	65.02	23.62	13.94
	$\Delta\sigma_{df}$	29.49	26.31	27.36
$\{1.0, -0.1, 0.4\}$	$\Delta\sigma_{\alpha_{\rm iso}}$	37.89	26.85	19.40
	$\Delta\sigma_{\alpha_{\rm AP}}$	63.88	34.02	28.17
	$\Delta\sigma_{df}$	40.16	26.63	38.81
$\{1.0, 0.0, 0.1\}$	$\Delta\sigma_{\alpha_{\rm iso}}$	26.65	32.66	13.33
	$\Delta\sigma_{\alpha_{\rm AP}}$	39.88	26.30	18.40
	$\Delta\sigma_{df}$	33.27	21.59	38.54

Table 4.8: Mean deviation of LRG fit errors from 250 `EZmocks`, using different FitCov covariance matrices.

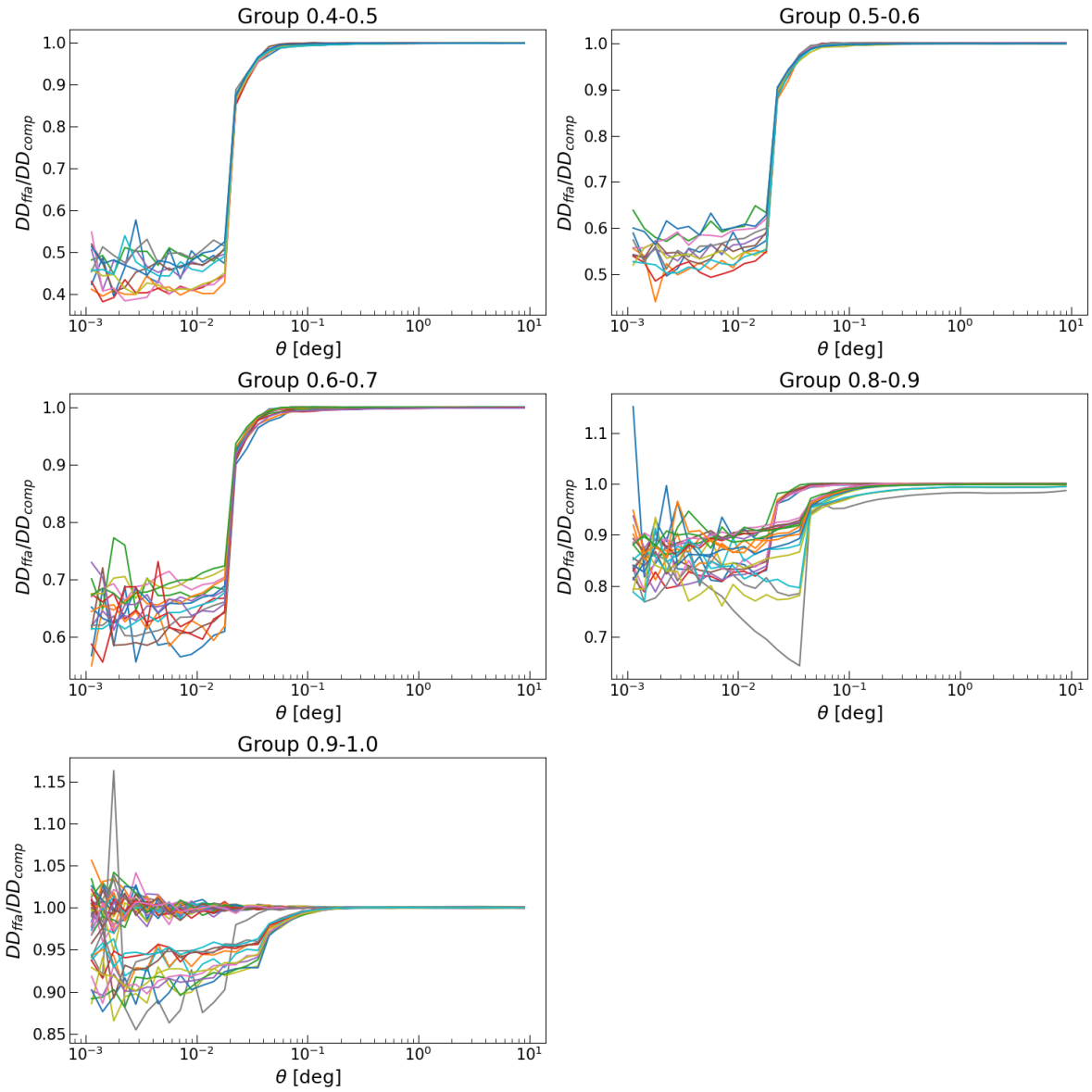


Figure 4.4: Figure showing the ratio of the number counts of assigned pairs to the number counts of pairs in the complete sample of LRG targets for `Abacus` mock 0 for different emulator runs. The groups in each subplot title refer to what the mean ratio of counts is inside the collision window for each graph.

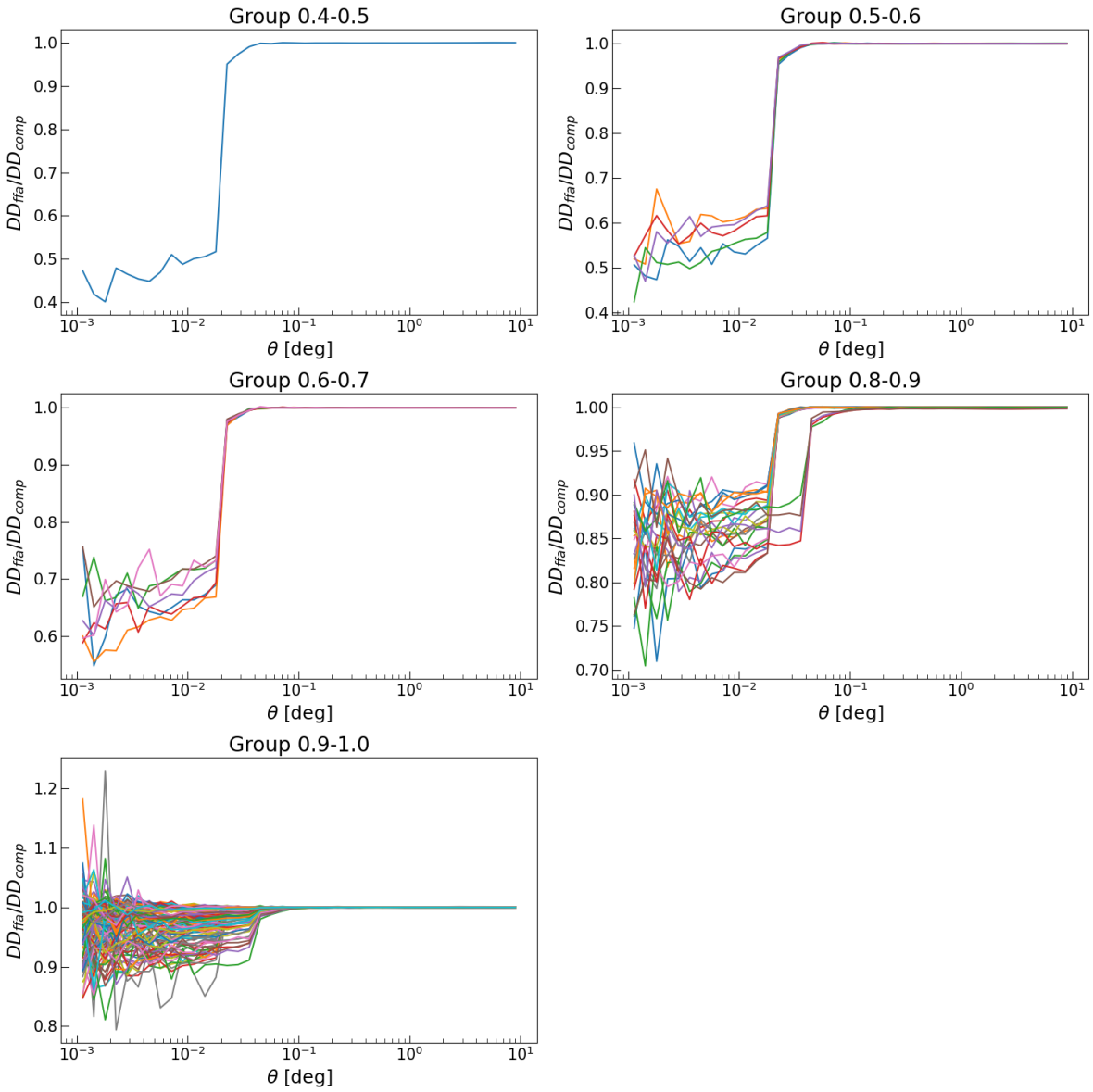


Figure 4.5: Figure showing the ratio of the number counts of assigned pairs to the number counts of pairs in the complete sample of QSO targets for `Abacus` mock 0 for different emulator runs. The groups in each subplot title refer to what the mean ratio of counts is inside the collision window for each graph.

$\{\beta_0, \beta_1, f_{\rm anti}\}$	Parameter	$z \in (0.4, 0.6)$	$z \in (0.6, 0.8)$	$z \in (0.8, 1.1)$
{1.0, -0.3, 0.1}	$\Delta\sigma_{\alpha_{\rm iso}}$	36.46	33.98	27.01
	$\Delta\sigma_{\alpha_{\rm AP}}$	146.65	60.17	35.31
	$\Delta\sigma_{df}$	27.24	24.99	4.33
$\{1.0, -0.2, 0.05\}$	$\Delta\sigma_{\alpha_{\rm iso}}$	11.72	11.22	8.47
	$\Delta\sigma_{\alpha_{\rm AP}}$	56.69	18.66	14.24
	$\Delta\sigma_{df}$	13.24	6.55	-1.94
$\{1.2, -0.3, 0.1\}$	$\Delta\sigma_{\alpha_{\rm iso}}$	19.11	18.51	25.32
	$\Delta\sigma_{\alpha_{\rm AP}}$	58.90	29.14	27.31
	$\Delta\sigma_{df}$	16.61	20.20	8.04
$\{1.0, -0.05, 0.025\}$	$\Delta\sigma_{\alpha_{\rm iso}}$	3.89	4.12	2.95
	$\Delta\sigma_{\alpha_{\rm AP}}$	40.05	13.63	4.22
	$\Delta\sigma_{df}$	5.51	9.71	-4.65
$\{1.0, -0.1, 0.4\}$	$\Delta\sigma_{\alpha_{\rm iso}}$	18.48	7.02	8.80
	$\Delta\sigma_{\alpha_{\rm AP}}$	44.09	21.20	20.84
	$\Delta\sigma_{df}$	16.28	10.64	4.04
$\{1.0, 0.0, 0.1\}$	$\Delta\sigma_{\alpha_{\rm iso}}$	0.00	0.00	0.00
	$\Delta\sigma_{\alpha_{\rm AP}}$	0.00	0.00	0.00
	$\Delta\sigma_{df}$	0.00	0.00	0.00

Table 4.9: Mean deviation of LRG fit errors from 250 `EZmocks`, using different FitCov covariance matrices, relative to the FitCov covariance matrix obtained from the fiducial FFA choice. Note this is why all the entries for the last (fiducial) iteration are zero.

$\{\beta_0, \beta_1, f_{\rm anti}\}$	Parameter	$z \in (0.8, 2.1)$
$\{1.0, -0.3, 0.025\}$	$\Delta\sigma_{\alpha_{\rm iso}}$	15.69
	$\Delta\sigma_{\alpha_{\rm AP}}$	15.56
	$\Delta\sigma_{df}$	10.80
$\{1.0, -0.3, 0.1\}$	$\Delta\sigma_{\alpha_{\rm iso}}$	40.40
	$\Delta\sigma_{\alpha_{\rm AP}}$	58.44
	$\Delta\sigma_{df}$	46.25
$\{1.0, -0.25, 0.05\}$	$\Delta\sigma_{\alpha_{\rm iso}}$	45.73
	$\Delta\sigma_{\alpha_{\rm AP}}$	48.09
	$\Delta\sigma_{df}$	46.13
$\{1.0, -0.2, 0.1\}$	$\Delta\sigma_{\alpha_{\rm iso}}$	21.83
	$\Delta\sigma_{\alpha_{\rm AP}}$	23.90
	$\Delta\sigma_{df}$	23.70
$\{1.2, -0.3, 0.025\}$	$\Delta\sigma_{\alpha_{\rm iso}}$	25.17
	$\Delta\sigma_{\alpha_{\rm AP}}$	31.33
	$\Delta\sigma_{df}$	30.15
$\{1.0, 0.0, 0.1\}$	$\Delta\sigma_{\alpha_{\rm iso}}$	23.86
	$\Delta\sigma_{\alpha_{\rm AP}}$	28.19
	$\Delta\sigma_{df}$	19.60

Table 4.10: Mean deviation of QSO fit errors from 250 `EZmocks`, using different FitCov covariance matrices.

$\{\beta_0, \beta_1, f_{\rm anti}\}$	Parameter	$z \in (0.4, 0.6)$	$z \in (0.6, 0.8)$	$z \in (0.8, 1.1)$
{1.0, -0.3, 0.1}	$\Delta\alpha_{\rm AP}$	-0.38	0.04	0.42
	$\Delta\sigma_{\alpha_{\rm AP}}$	90.50	65.66	39.50
$\{1.0, -0.2, 0.05\}$	$\Delta\alpha_{\rm AP}$	-0.20	0.56	0.01
	$\Delta\sigma_{\alpha_{\rm AP}}$	28.52	21.19	19.70
$\{1.2, -0.3, 0.1\}$	$\Delta\alpha_{\rm AP}$	-0.82	0.07	0.18
	$\Delta\sigma_{\alpha_{\rm AP}}$	53.51	43.69	29.91
$\{1.0, -0.05, 0.025\}$	$\Delta\alpha_{\rm AP}$	-0.43	-0.22	0.36
	$\Delta\sigma_{\alpha_{\rm AP}}$	49.00	22.10	15.15
$\{1.0, -0.1, 0.4\}$	$\Delta\alpha_{\rm AP}$	0.06	0.25	0.52
	$\Delta\sigma_{\alpha_{\rm AP}}$	41.14	24.01	25.32
$\{1.0, 0.0, 0.1\}$	$\Delta\alpha_{\rm AP}$	-0.40	0.41	-0.39
	$\Delta\sigma_{\alpha_{\rm AP}}$	22.51	21.88	11.92

Table 4.11: Mean deviation of LRG fits to α_{AP} and the corresponding errors on it from 250 `EZmocks`, using different FitCov covariance matrices.

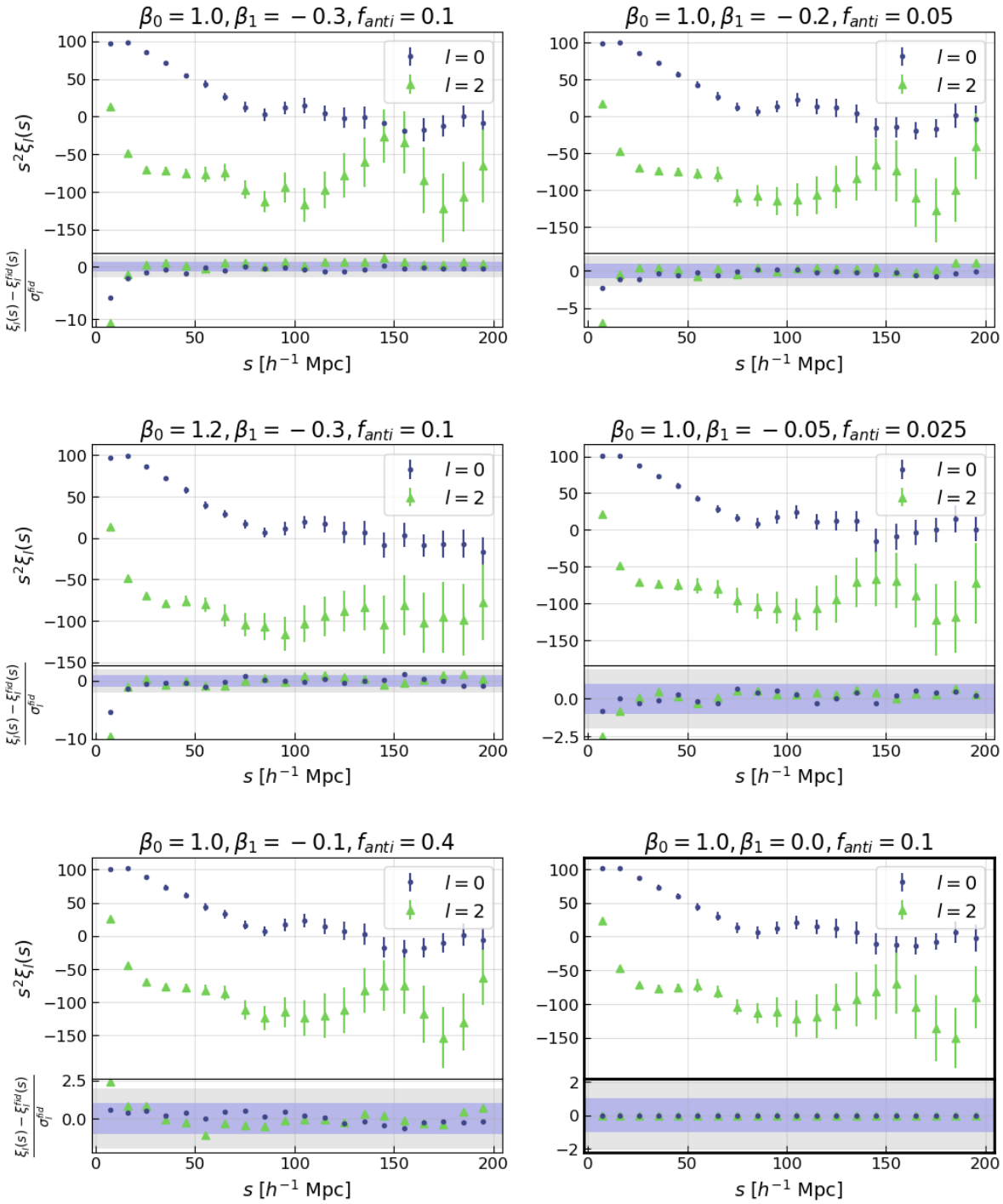


Figure 4.6: This figure shows the monopole ($l = 0$) and quadrupole ($l = 2$) of the configuration space two-point correlation function for LRGs in the $0.4 < z < 0.6$ redshift range for various emulated runs. The deviation of each clustering statistic relative to a 'fiducial' one (bottom-right panel with bold border) is also plotted. The blue and gray bands are the 1σ and 2σ intervals, respectively.

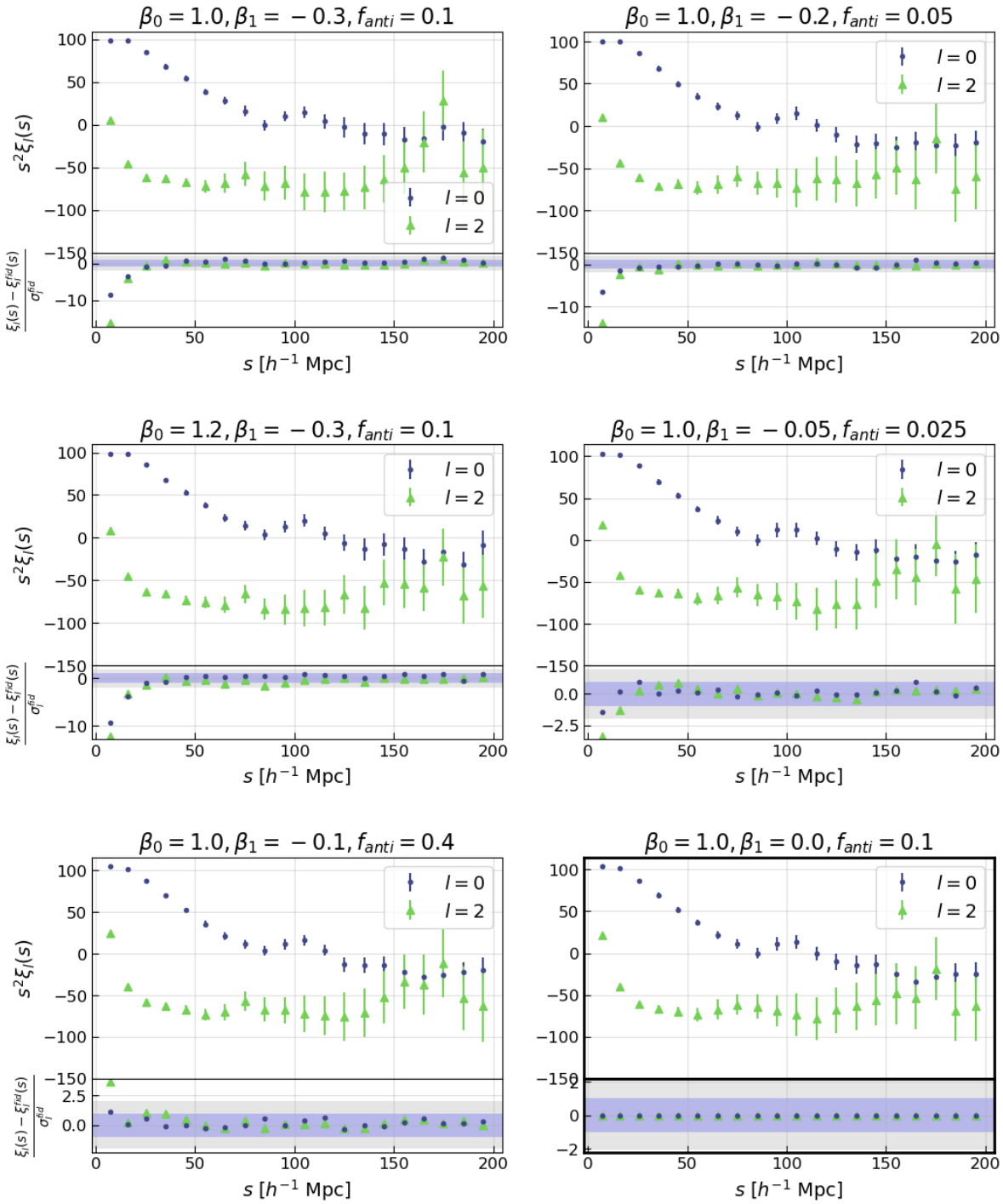


Figure 4.7: This figure shows the monopole ($l = 0$) and quadrupole ($l = 2$) of the configuration space two-point correlation function for LRGs in the $0.6 < z < 0.8$ redshift range for various emulated runs. The deviation of each clustering statistic relative to a 'fiducial' one (bottom-right panel with bold border) is also plotted. The blue and gray bands are the 1σ and 2σ intervals, respectively.

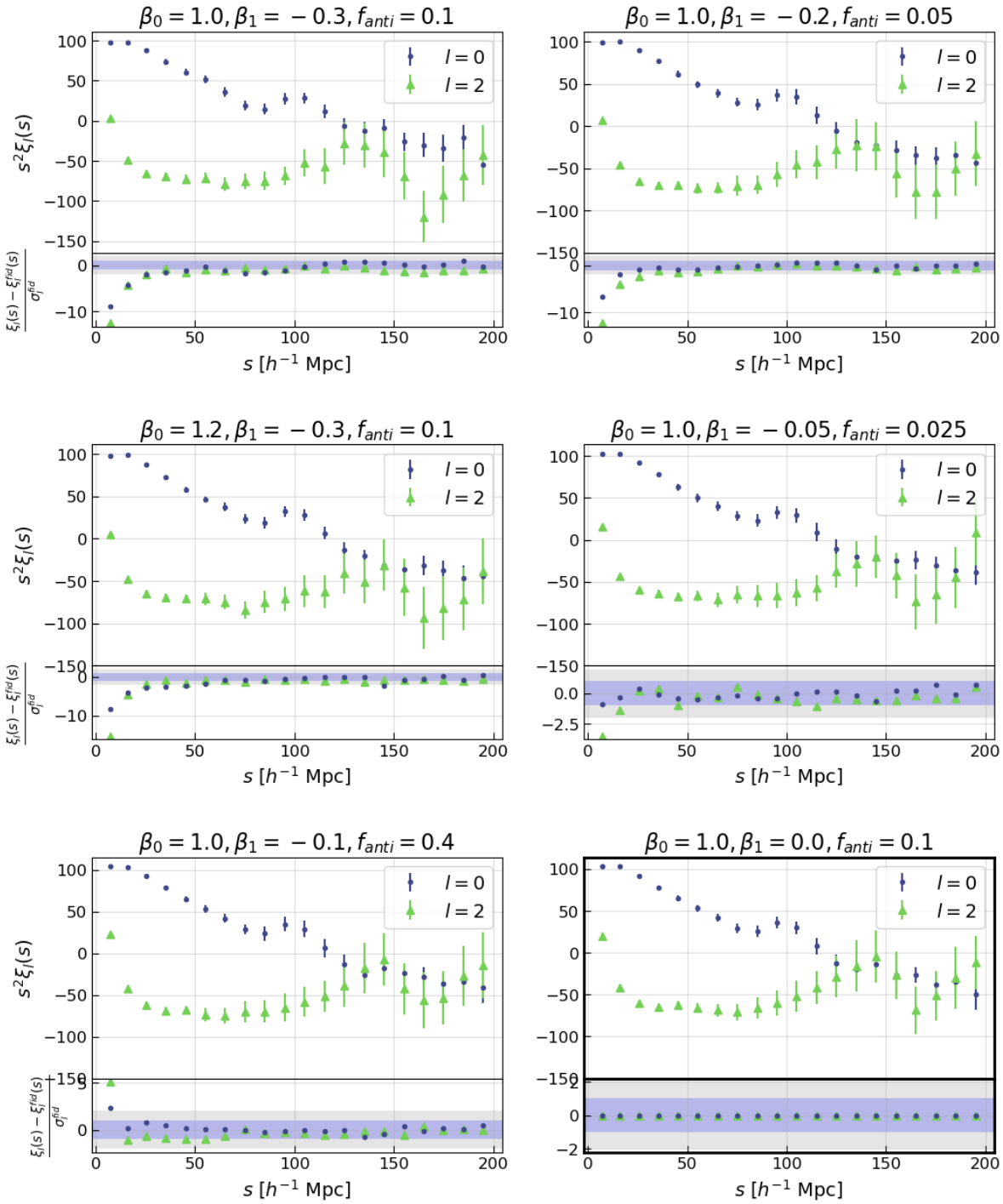


Figure 4.8: This figure shows the monopole ($l = 0$) and quadrupole ($l = 2$) of the configuration space two-point correlation function for LRGs in the $0.8 < z < 1.1$ redshift range for various emulated runs. The deviation of each clustering statistic relative to a 'fiducial' one (bottom-right panel with bold border) is also plotted. The blue and gray bands are the 1σ and 2σ intervals, respectively.

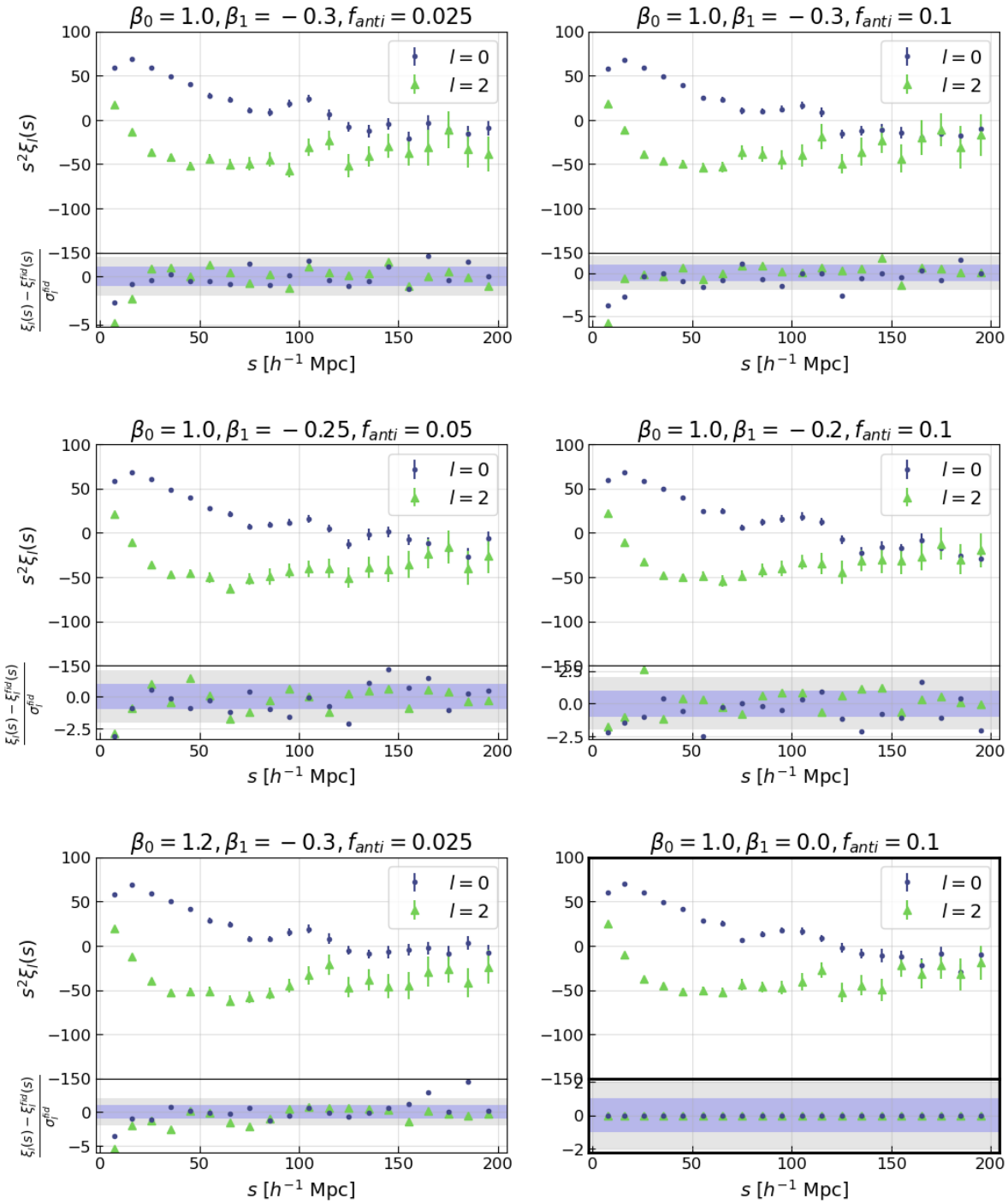


Figure 4.9: This figure shows the monopole ($l = 0$) and quadrupole ($l = 2$) of the configuration space two-point correlation function for QSOs in the $0.8 < z < 2.1$ redshift range for various emulated runs. The deviation of each clustering statistic relative to a 'fiducial' one (bottom-right panel with bold border) is also plotted. The blue and gray bands are the 1σ and 2σ intervals, respectively.

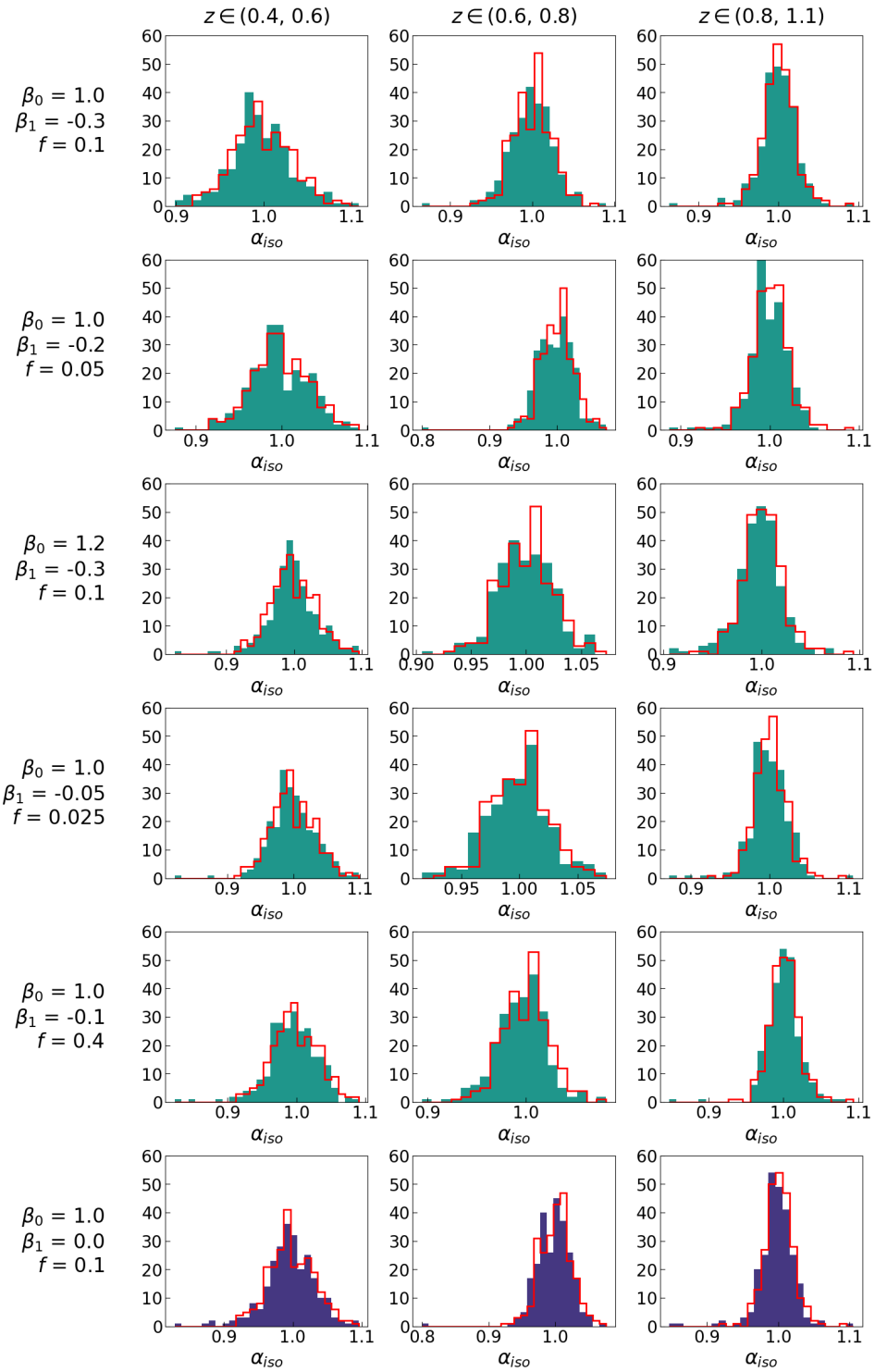


Figure 4.10: The solid histograms are the distribution of $\alpha_{\rm iso}$ from fitting 250 LRG `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix.

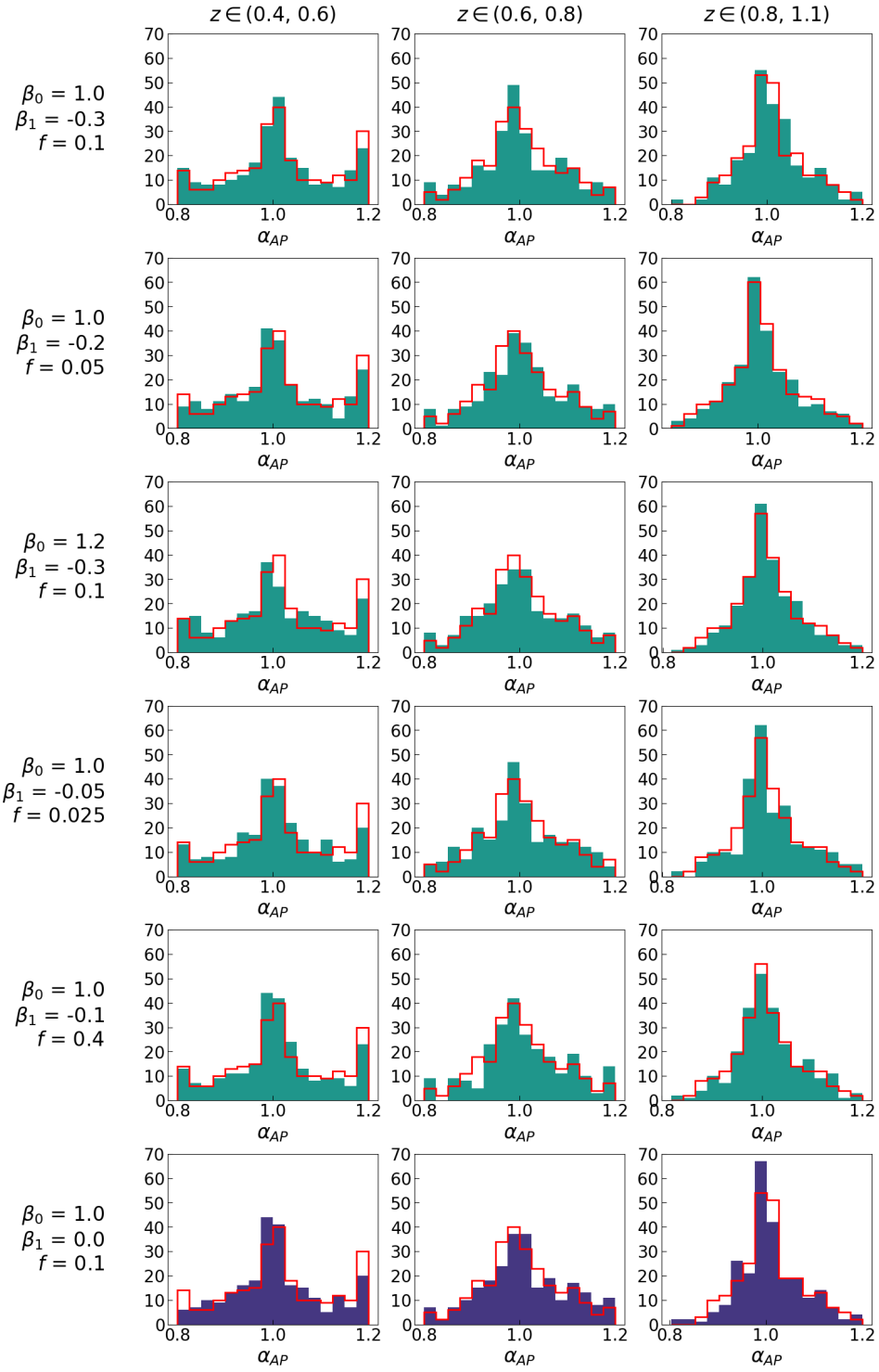


Figure 4.11: The solid histograms are the distribution of $\alpha_{\rm AP}$ from fitting 250 LRG `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix.

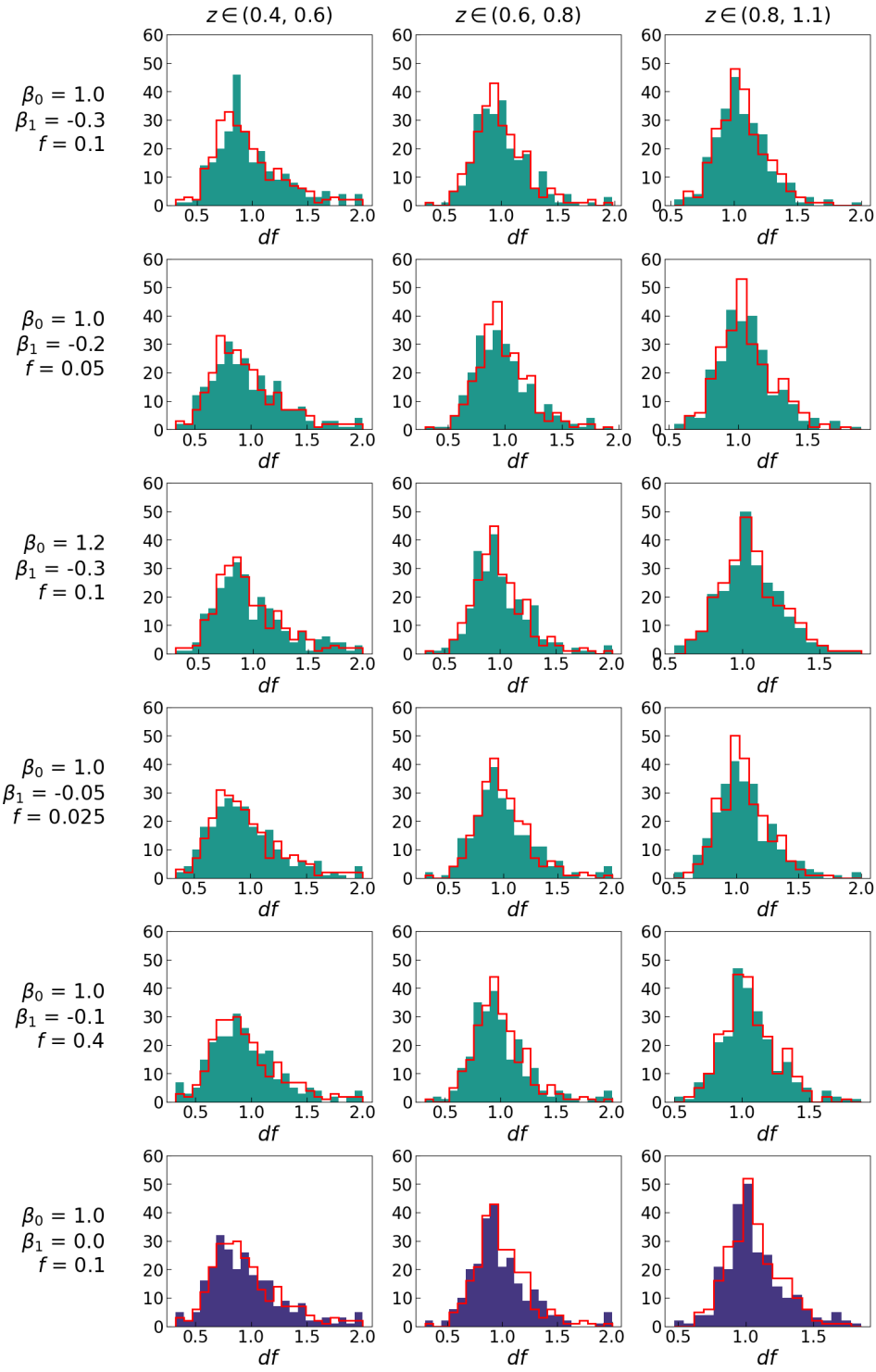


Figure 4.12: The solid histograms are the distribution of *df* from fitting 250 LRG `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix.

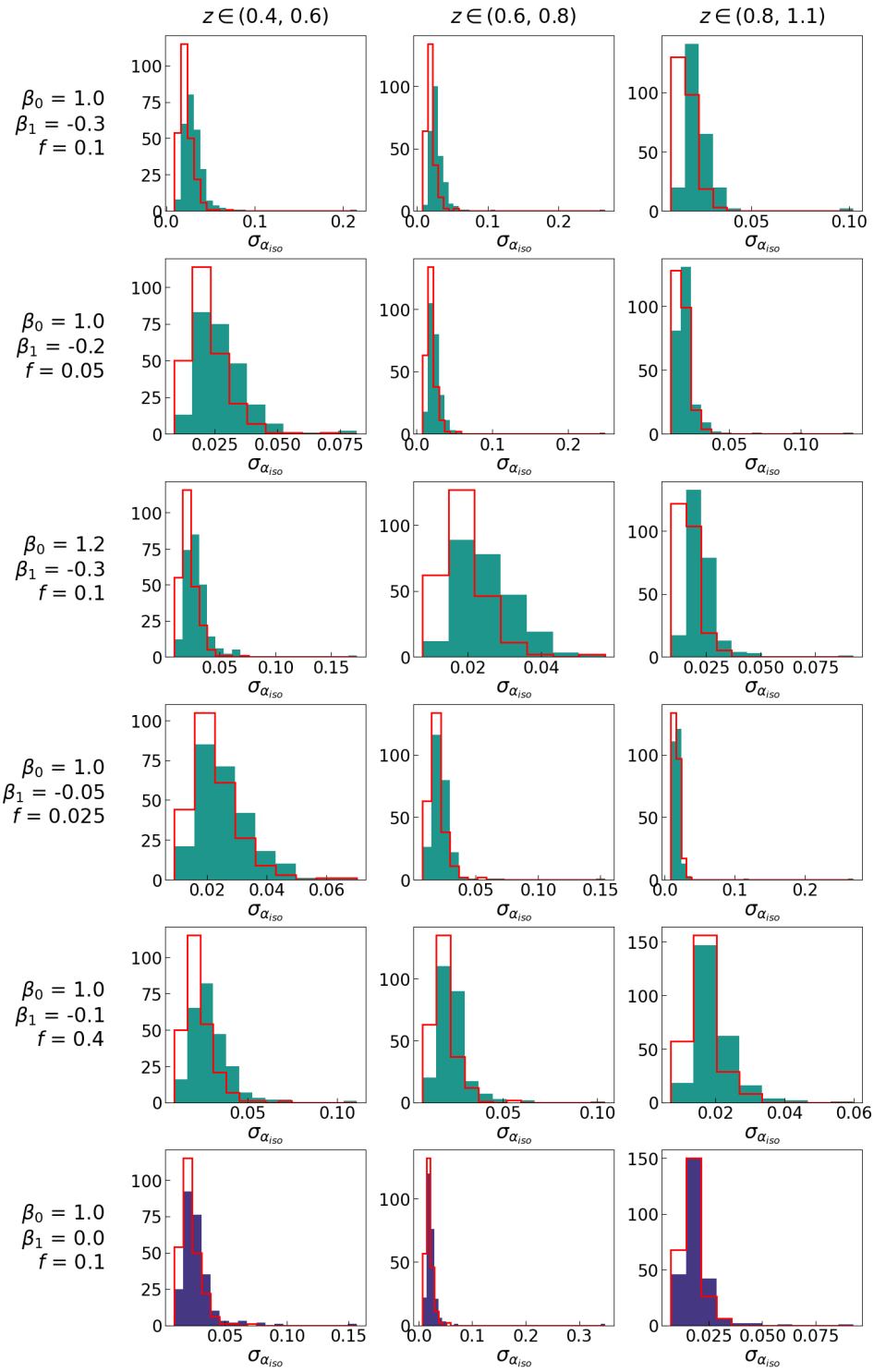


Figure 4.13: The solid histograms are the distribution of $\sigma_{\alpha_{\rm iso}}$ from fitting 250 LRG `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix.

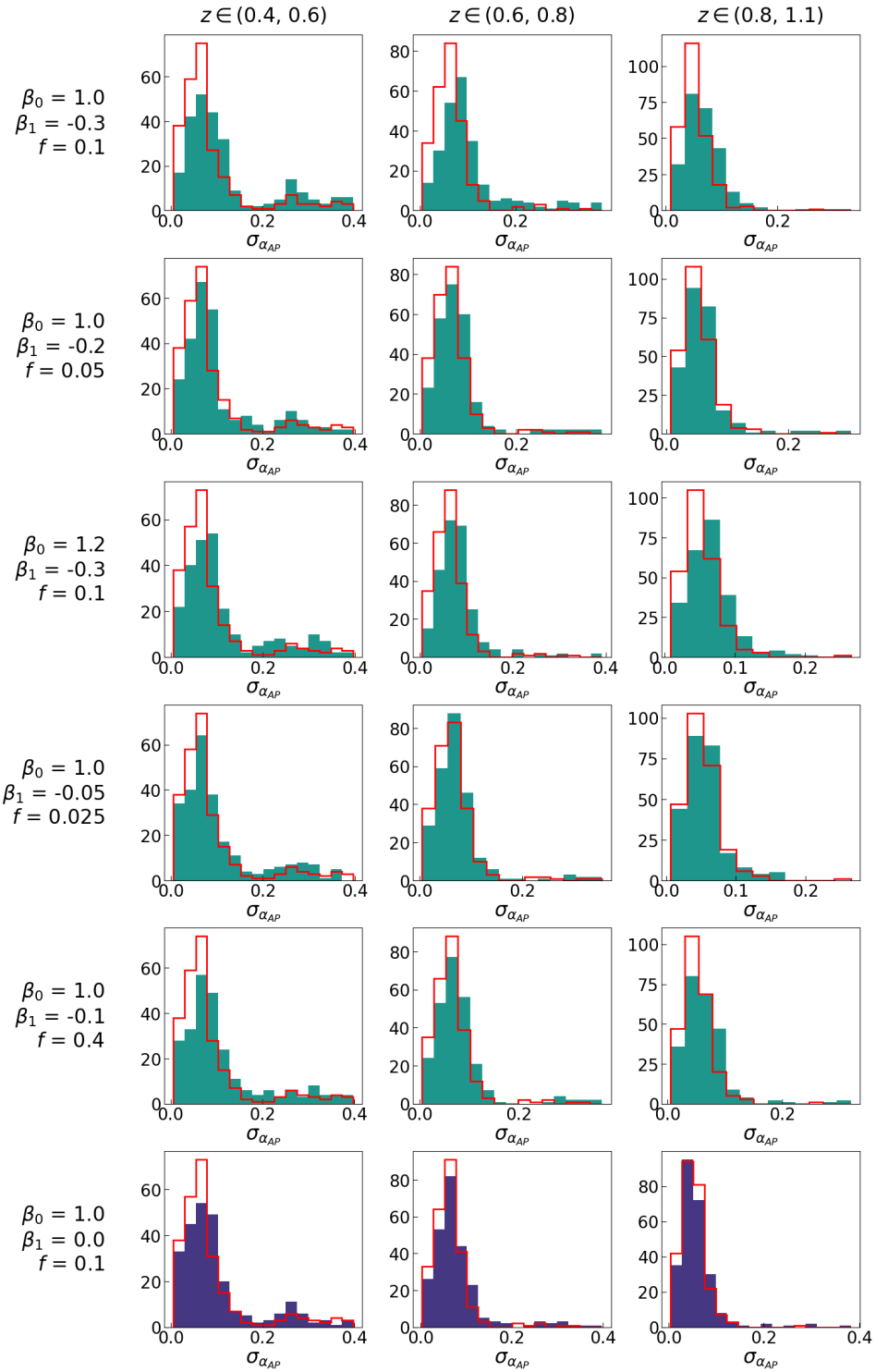


Figure 4.14: The solid histograms are the distribution of $\sigma_{\alpha_{\rm AP}}$ from fitting 250 LRG `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix.

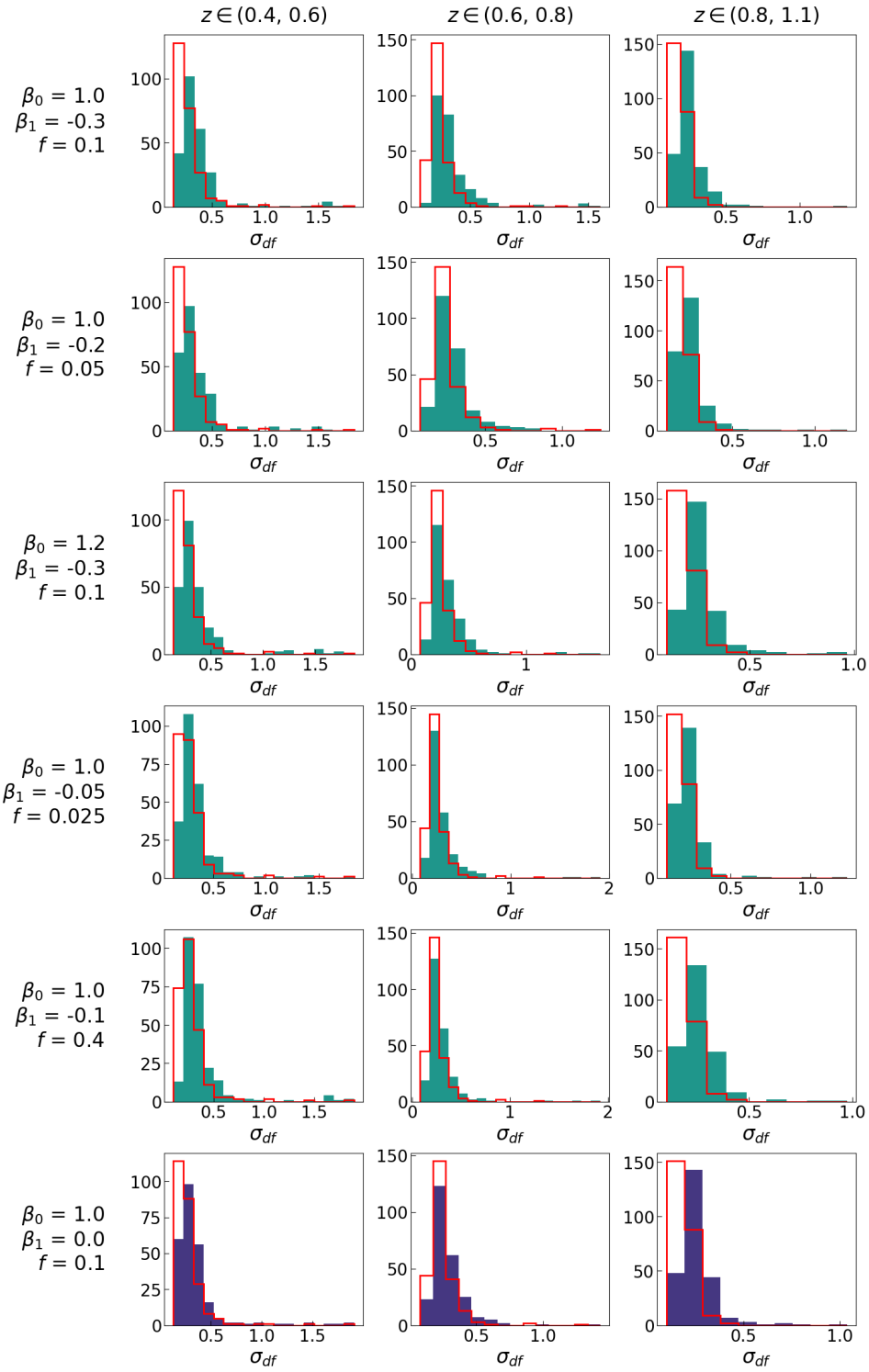


Figure 4.15: The solid histograms are the distribution of σ_{df} from fitting 250 LRG `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix.

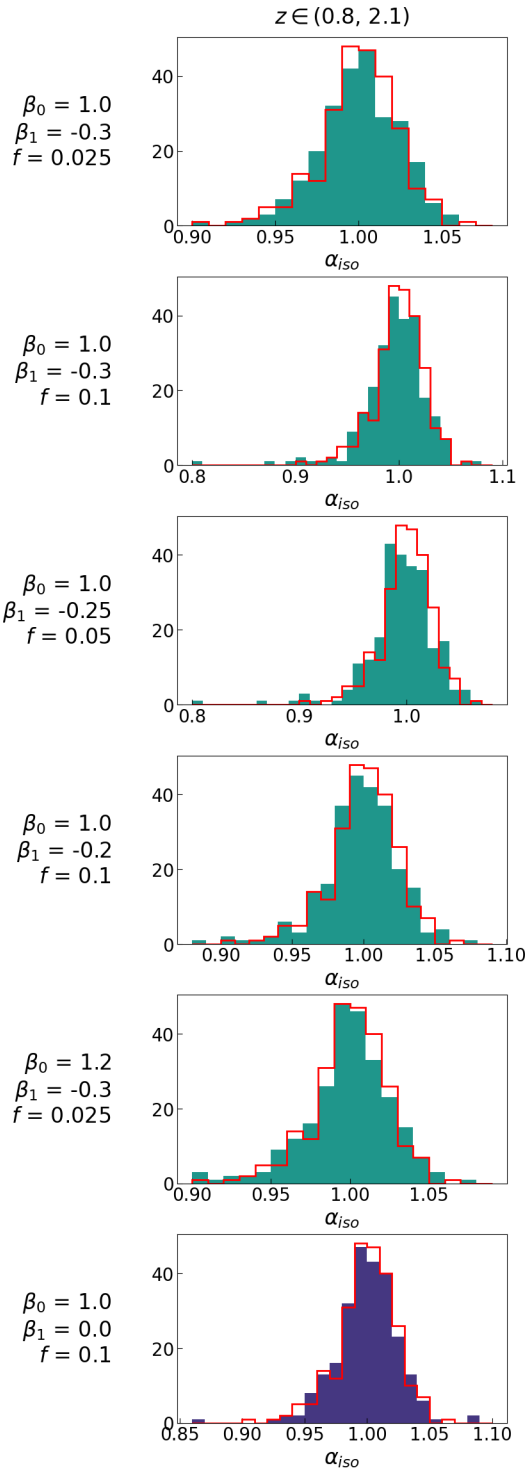


Figure 4.16: The solid histograms are the distribution of $\alpha_{\rm iso}$ from fitting 250 QSO `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix.

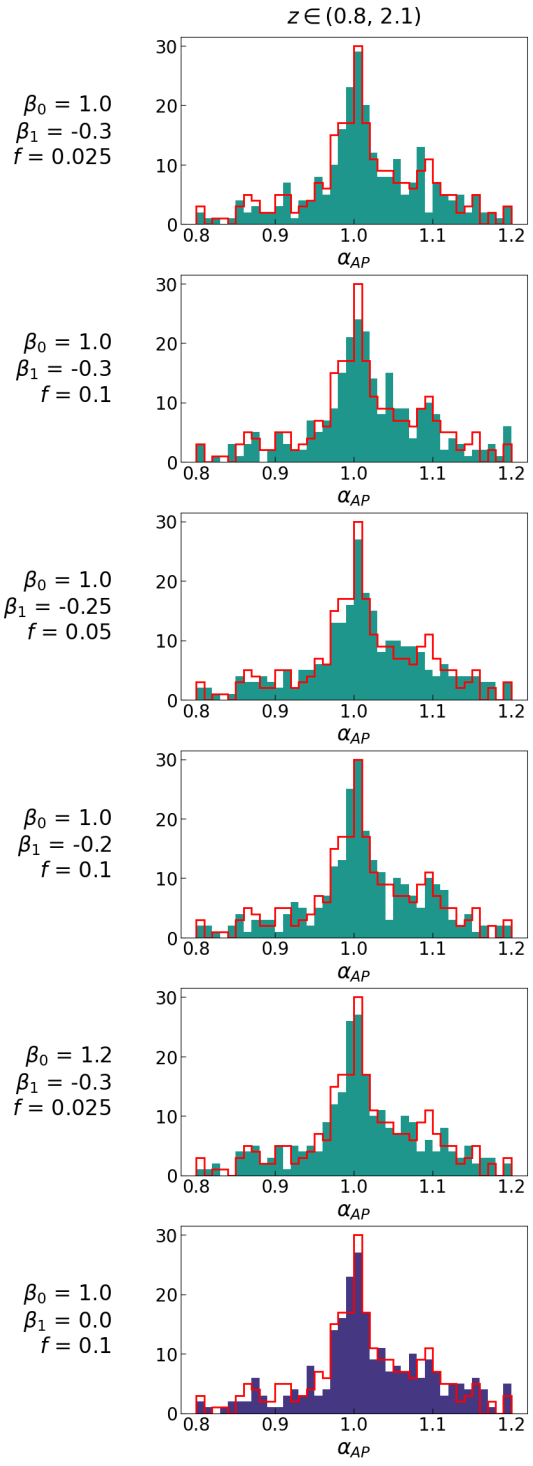


Figure 4.17: The solid histograms are the distribution of $\alpha_{\rm AP}$ from fitting 250 QSO `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix.

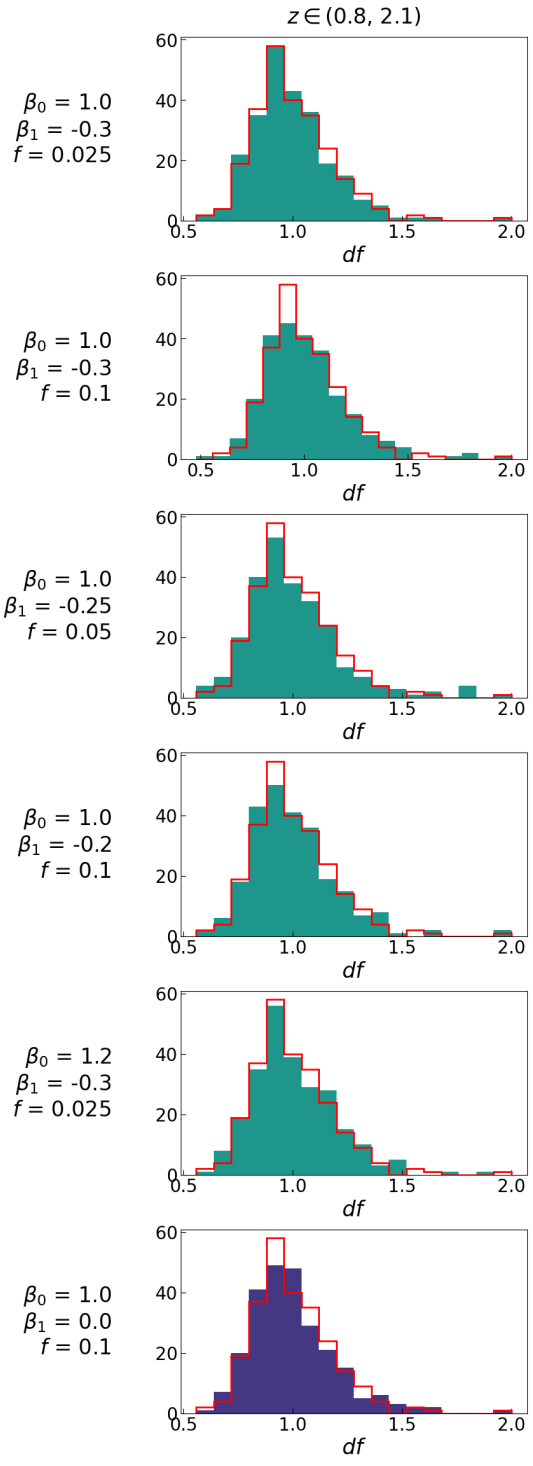


Figure 4.18: The solid histograms are the distribution of *df* from fitting 250 QSO `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix.

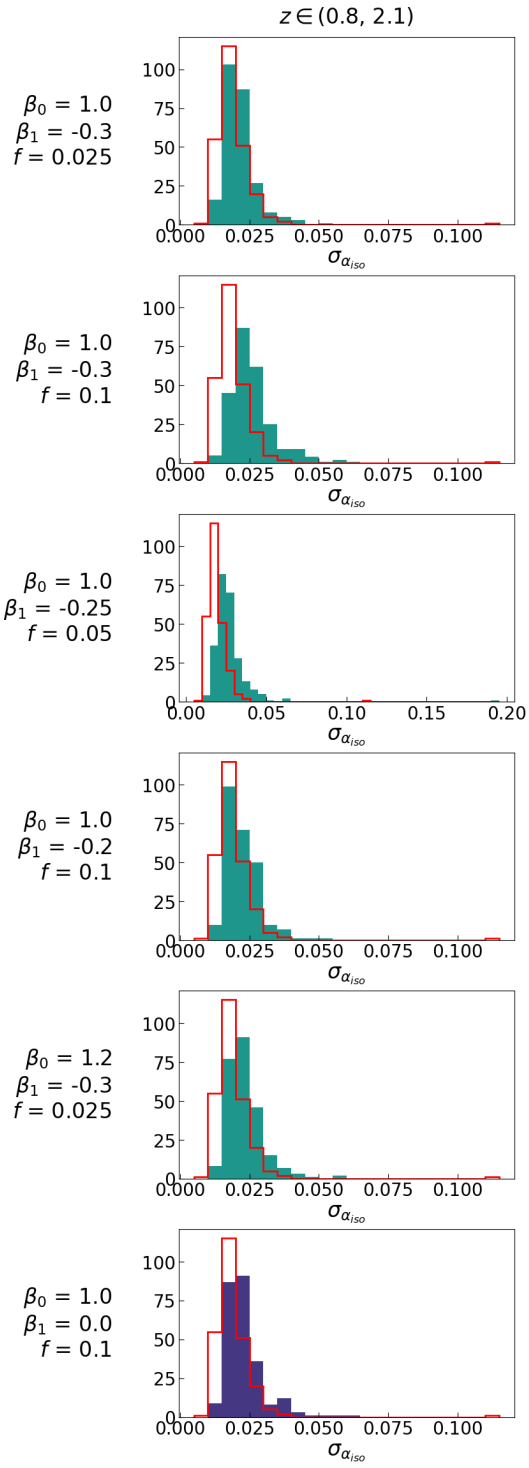


Figure 4.19: The solid histograms are the distribution of $\sigma_{\alpha_{\rm iso}}$ from fitting 250 QSO `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix.

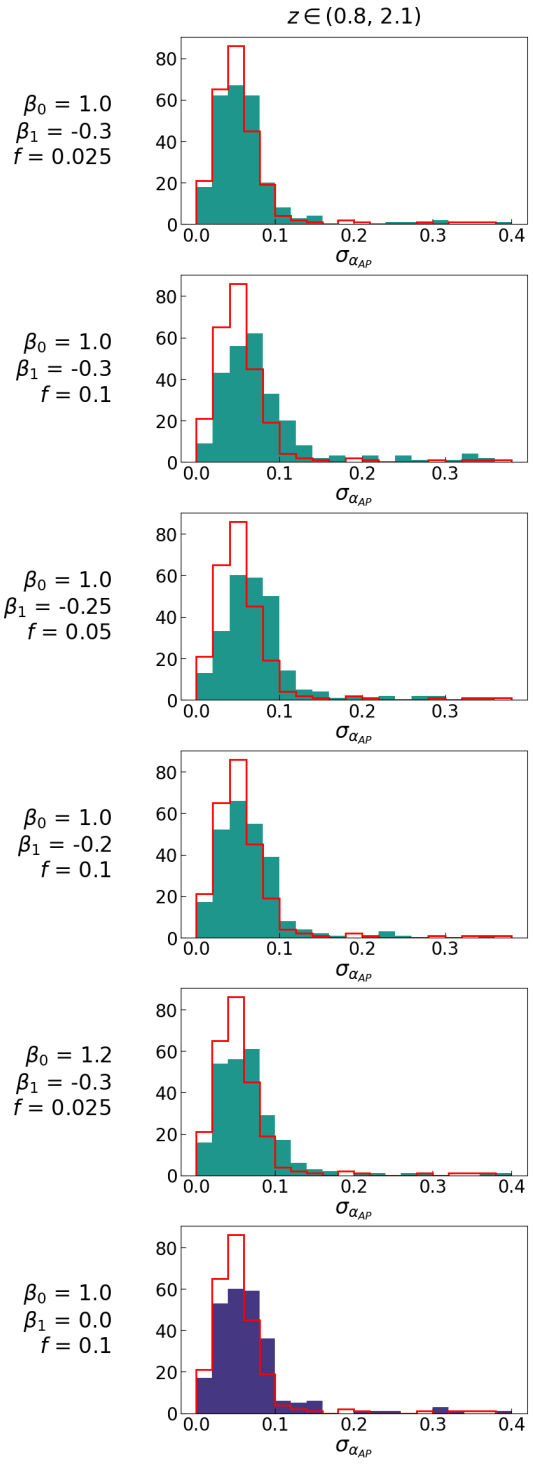


Figure 4.20: The solid histograms are the distribution of $\sigma_{\alpha_{\rm AP}}$ from fitting 250 QSO `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix.

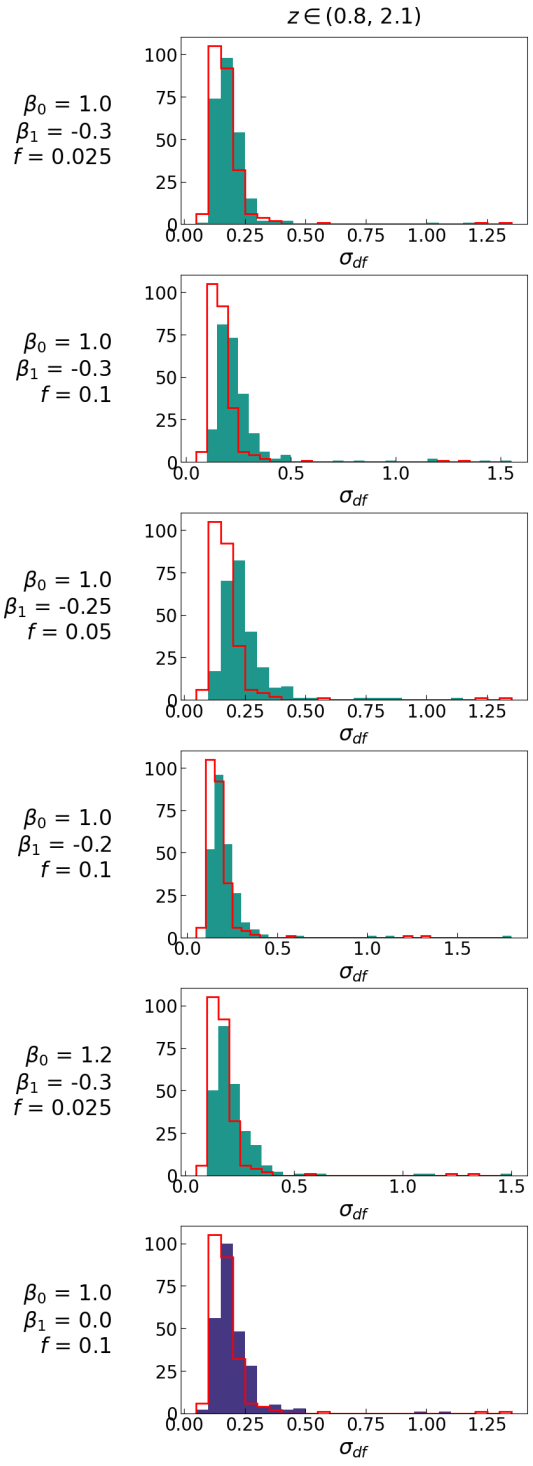


Figure 4.21: The solid histograms are the distribution of σ_{df} from fitting 250 QSO `EZmocks` with various FitCov covariance matrices. The 'fiducial' choice is shown in purple. The red line is the corresponding distribution using an analytical covariance matrix.

CHAPTER 5

DESI Y1 Cosmology

I will now present a brief review of DESI Y1 cosmology results, in pursuit of which fast fiber assignment and the pipeline for getting covariance matrices of clustering were developed. In particular, I will focus on cosmology from baryon acoustic oscillations and full shape analyses.

The Y1 analysis used results from the DR1 (Data Release 1) spectroscopic dataset. This is built from main survey operations from May 14, 2021 to June 14, 2022, comprising 2,744 dark tiles and 2,275 bright ones. The LSS catalogs for the clustering analysis had their two-point measurements obtained via the processes in [1]. The selection function and weights used are designed to accommodate for the effect of imaging anisotropies on the input target samples [87, 19], redshift measurement efficiency [104, 53] and fiber assignment effects [56, 10, 79]. The different targets comprising this dataset are discussed below.

5.1 DESI Data Release 1

DESI defines four different classes of targets, split into two categories. Recall from Chapter 1 that the 'dark time' tracers are comprised of luminous red galaxies (LRGs), emission line galaxies (ELGs) and quasars (QSOs). The 'bright time' targets are called the bright galaxy sample (BGS) and these are observed when observing conditions are

not optimal for dark time viewing, for example, because of a full moon.

5.1.1 LRG Sample

The LRG target selection [110] uses photometry from the Wide-field Infrared Survey Explorer (WISE) [102] to select red objects. The DESI DR1 LRG clustering sample used for BAO measurements comprises $2,138,600$ targets in the redshift range from 0.4 to 1.1 This is split into three redshift bins: LRG1 ($0.4 < z < 0.6$), LRG2 ($0.6 < z < 0.8$) and LRG3 ($0.8 < z < 1.1$). This third redshift bin overlaps with an ELG bin and is combined with that for the analysis.

5.1.2 ELG Sample

ELG targets are selected with color cuts in the $(g-r)$ versus $(r-z)$ plane to prioritize objects with [OII] emission in the desired redshift range for good redshift measurements [81]. The ELG sample used for clustering contains $2,432,022$ good redshifts in the redshift range from 0.8 to 1.6. This is split into two redshift bins: ELG1 ($0.8 < z < 1.1$) and ELG2 ($1.1 < z < 1.6$). The lower redshift bin is combined with LRG3 as they overlap.

5.1.3 QSO Sample

The DESI QSO sample used for BAO analysis contains $856,652$ targets in the redshift range from 0.8 to 2.1.

5.1.4 BGS Sample

The BGS clustering sample used for DESI BAO analysis comprises $300,017$ redshifts in the redshift range from 0.1 to 0.4.

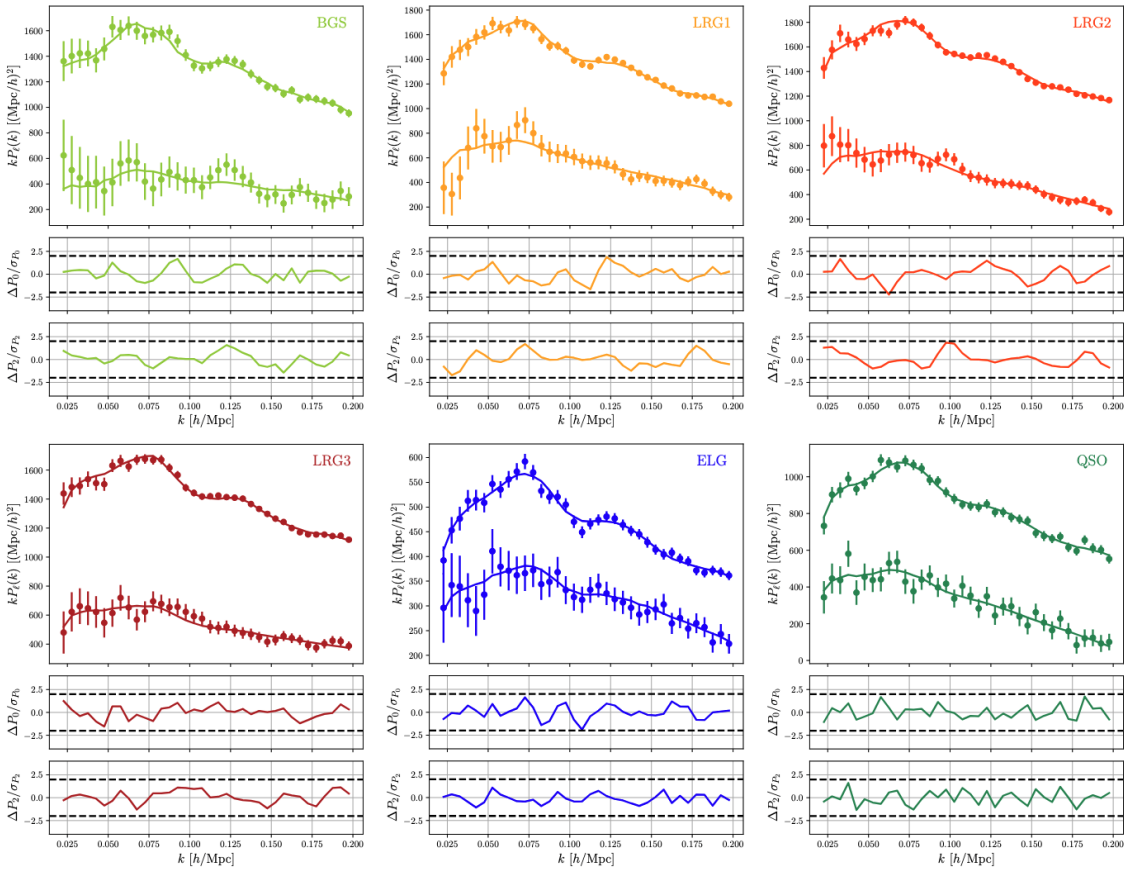


Figure 5.1: Measured and best-fit monopole and quadrupole from DESI DR1 measurements. Symbols display the measurements and solid lines the best-fitting power spectrum model to each of the individual redshift bins according to the DESI full modeling baseline model (see Table 10 in [2]). The errors are 1σ and include both the statistical and systematic error budget. The bottom sub-panels display the residuals between data and model in units of the diagonal error. Caption adapted and figure reproduced from [2].

5.2 BAO Analysis

As was introduced in Chapter 2, measured target redshifts and angular positions have to be converted to comoving coordinates so that the clustering can be measured. This is done by adopting a reference cosmology. The clustering patterns perpendicular and parallel to the line-of-sight are affected in predictable ways between the reference and potential 'true' (or trial) cosmology, encoded by the parameters in Equation 2.1,

reproduced below as a reminder:

$$\alpha_\perp = \frac{D_M(z) r_d^{\rm ref}}{D_M^{\rm ref}(z) r_d}$$
$$\alpha_\parallel = \frac{H^{\rm ref}(z) r_d}{H(z) r_d^{\rm ref}}$$

The theoretical calibration of the BAO scale is encoded by a primordial matter power spectrum generated via a fiducial cosmological model. See Figure 2.1 for examples. This is used to produce a template clustering model, indexed by the sound horizon at the drag epoch. In addition to the parameters in Equations 2.1 and 2.2, models also include nuisance parameters for the full shape of the power spectrum which are marginalized over in the 'classic' BAO analysis.

The non-linear growth of structure in the post early universe dilutes the amplitude of the BAO peak. This is due to the bulk flow motion of matter from its true location in the density field, altering the real separations. Reconstruction techniques are employed to partially reverse this effect [39, 73, 74]. This has been shown to sharpen the BAO peak and improve inference and thus it is these post-reconstruction measurements that are used in the configuration space analysis.

Anisotropic BAO measurements are obtained for LRGs and ELGs but for BGS targets and QSOs, DESI restricted to isotropic fits. This was because of the lower signal-to-noise ratio of the two-point function measurements for the latter tracers. This is reflected in the values quoted in Table 5.2. The uncertainties in these clustering measurements are encoded via covariance matrices. The total systematic error budget comprises of contributions from different areas: theoretical modeling uncertainties [21], uncertainties due to the galaxy-halo connection [68, 44], and uncertainties from observational systematic effects [87, 104]. These are added together in quadrature to calculate the total error budget for the BAO measurements.

For the published results, cosmological inference was done using the code `cobaya`

Parametrization	Parameter	Default	Prior
background-only	Ω_m	—	$\mathcal{U}[0.01, 0.99]$
no r_d calibration	$r_d h$ (Mpc)	—	$\mathcal{U}[10, 1000]$
with r_d calibration	H_0 (km s^{-1} Mpc^{-1})	—	$\mathcal{U}[20, 100]$
	ω_b	—	$\mathcal{U}[0.005, 0.1]$
CMB	$\omega_{\rm cdm}$	—	$\mathcal{U}[0.001, 0.99]$
	$\omega_{\rm b}$	—	$\mathcal{U}[0.005, 0.1]$
	$100\theta_{\rm MC}$	—	$\mathcal{U}[0.5, 10]$
	$\ln(10^{10} A_s)$	—	$\mathcal{U}[1.61, 3.91]$
	n_s	—	$\mathcal{U}[0.8, 1.2]$
	τ	—	$\mathcal{U}[0.01, 0.8]$
extended	$\Omega_{\rm K}$	0	$\mathcal{U}[-0.3, 0.3]$
	w_0 or w	-1	$\mathcal{U}[-3, 1]$
	w_a	0	$\mathcal{U}[-3, 2]$
	$\sum m_\nu$ (eV)	0.06	$\mathcal{U}[0, 5]$
	$\rm N_{eff}$	3.044	$\mathcal{U}[0.05, 10]$

Table 5.1: Parameters and priors used in the DESI Y1 BAO analysis. The 'background-only' parametrization refers to using BAO and SN data only. 'CMB' refers to data from *Planck* and ACT being involved. Table reproduced from [3].

[96]. SN Ia likelihoods for PantheonPlus, Union3 and DESY5 were incorporated along with new DESI BAO likelihoods. The Boltzmann code CAMB was used for theoretical cosmology calculations [60, 49]. Inference was performed via the Metropolis-Hastings MCMC sampler [59, 58]. The parameters that are sampled, along with their priors, are detailed in Table 5.1. In the concordance ΛCDM model, Ω_m and $r_d h$ were sampled in the base cases. When external datasets or BBN priors were involved to break degeneracies, Ω_m, H_0 and ω_b (the physical baryon density) were sampled. Figure 5.2 presents the Hubble diagram that summarizes the DESI BAO results. Table 5.2 shows the measured distances.

5.3 Cosmological Constraints From BAO

The results presented in this section are drawn from the Y1 DESI Key Paper on BAO cosmology [3].

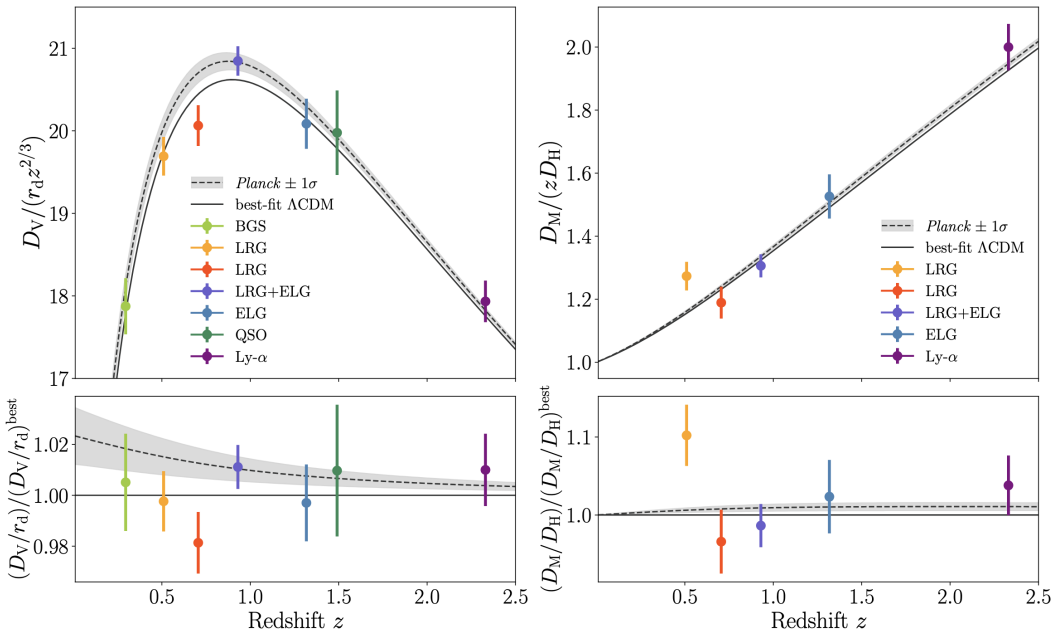


Figure 5.2: Top row: DESI measurements of the BAO distance scales at different redshifts, parametrized as (left) the ratio of the angle-averaged distance $D_V \equiv (zD_M^2 D_H)^{1/3}$ to the sound horizon at the baryon drag epoch, r_d, and (right) the ratio of transverse and line-of-sight comoving distances D_M/D_H, from all tracers and redshift bins as labeled. Caption adapted from and figure reproduced from [3].

Tracer	Redshift	N_{tracer}	D_{M}/r_d	D_{H}/r_d	r or D_{V}/r_d
BGS	$0.1 - 0.4$	$300,017$	—	—	7.93 ± 0.15
LRG1	$0.4 - 0.6$	$506,905$	13.62 ± 0.25	20.98 ± 0.61	-0.445
LRG2	$0.6 - 0.8$	$771,875$	16.85 ± 0.32	20.08 ± 0.60	-0.420
LRG3 + ELG1	$0.8 - 1.1$	$1,876,164$	21.71 ± 0.28	17.88 ± 0.35	-0.389
ELG2	$1.1 - 1.6$	$1,415,687$	27.79 ± 0.69	13.82 ± 0.42	-0.444
QSO	$0.8 - 2.1$	$856,652$	—	—	26.07 ± 0.67
Lya QSO	$1.77 - 4.16$	$709,565$	39.71 ± 0.94	8.52 ± 0.17	-0.477

Table 5.2: Statistics for the DESI samples used for the DESI DR1 BAO measurements used for the BAO analysis. For each tracer and redshift range, N_{tracer} is the number of objects. The angular diameter distance D_{M} and the Hubble distance D_{H} are presented in units of the sound horizon at the drag epoch. When these two are measured, they are correlated with coefficient r. Otherwise, D_{V}/r_d is the measured quantity. Table adapted from [3].

5.3.1 ΛCDM

In the concordance cosmology model ΛCDM, using only BAO scales provides constraints on two free parameters: Ω_m, the matter density parameter, and the combination $H_0 r_d$ where H_0 is the Hubble constant and r_d is the sound horizon at the drag epoch. We present below the 68% credible interval for these and Figure 5.3 shows the posterior constraints.

$$\left.\begin{aligned} \Omega_m &= 0.295 \pm 0.015 \\ r_d h &= (101.8 \pm 1.3)\ \text{Mpc} \end{aligned}\right\} \text{DESI BAO} \tag{5.1}$$

where $r_d h \equiv H_0 r_d / (100\ \text{km s}^{-1}\ \text{Mpc}^{-1})$ for simplicity. The degeneracy between the

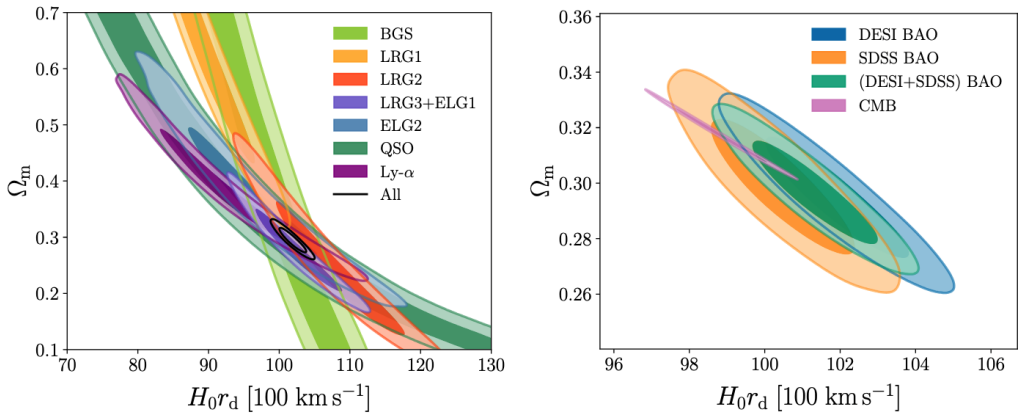


Figure 5.3: *Left panel:* 68% and 95% credible-interval contours for Ω_m and $r_d h$ for a flat ΛCDM cosmology from fits to BAO data from each DESI tracer individually. The results from all tracers are consistent with each other and the change in the degeneracy direction comes from the different effective redshifts of the samples. *Right panel:* The corresponding results for fits to data from the combined tracers from DESI, supplemented by information from external datasets. Caption adapted from and figure reproduced from [3].

Hubble constant and the sound horizon can be broken by calibrating the sound horizon using external data. This calibration of BAO distance scales via the sound horizon at early times is called the 'inverse distance ladder' method [9, 30]. Using CMB and CMB lensing information from [5] to obtain $r_d = 147.09 \pm 0.26$ Mpc, for example, yields $H_0 = 69.29 \pm 0.87$ km s^{-1} Mpc^{-1}. Figure 5.4 summarizes the results for this using

different priors to break the aforementioned degeneracy.

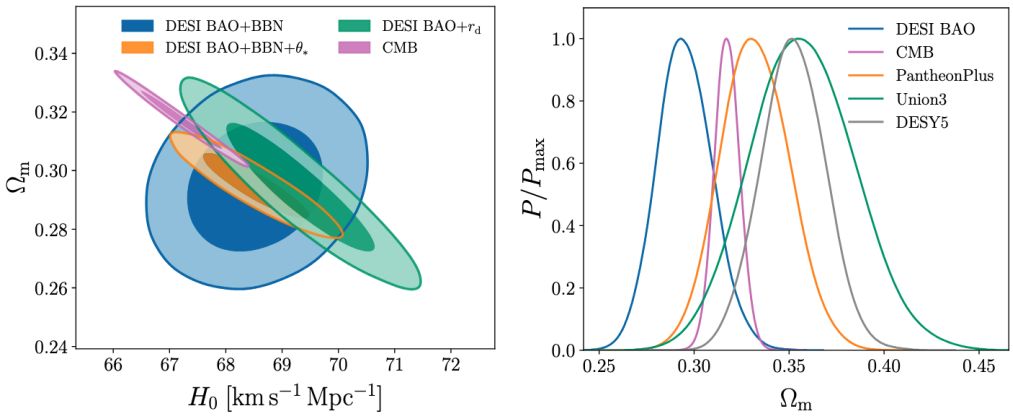


Figure 5.4: *Left panel:* Marginalized posterior constraints on Ω_M and H_0 from combining DESI BAO data with external data used to calibrate r_d in a flat ΛCDM model. *Right panel:* The marginalized 1D posteriors on Ω_m in a flat ΛCDM model from different probes, as labeled. Caption adapted from and figure reproduced from [3].

5.3.2 Dark Energy Constraints

The concordance cosmological model describes a universe with a flat geometry and a matter density that is lower than the critical value. However, researchers continue to study the universe's expansion and structure growth in search of any deviations from this model. Therefore, one of the main objectives of DESI is to test and constrain alternative dark energy models beyond the basic ΛCDM framework.

5.3.2.1 Flat wCDM and w_0w_aCDM

The small observed value of the cosmological constant Λ in contrast to typical particle physics scales has posed theoretical challenges [65]. To study models beyond a constant value of dark energy, one of the models proposed is wCDM, where w refers to the equation of state parameter from Chapter 1. The goal here has been to look for statistically significant deviations from $w = -1$, the value that *also* corresponds to a cosmological constant. Therefore, note that measuring $w = -1$ does not necessarily confirm the

cosmological constant but measuring deviations would be a sign of new physics. DESI BAO data give the following constraints:

$$\left.\begin{aligned} \Omega_m &= 0.293 \pm 0.015 \\ w &= -0.99^{+0.15}_{-0.13} \end{aligned}\right\} \text{DESI BAO} \tag{5.2}$$

Beyond this, accounting for dark energy's physical dynamics, the following evolving parametrization for dark energy has been put forward:

$$w(a) = w_0 + w_a(1 - a) \tag{5.3}$$

This has been shown to match the background evolution of distances arising from the dark energy equations of motion very well, over a range of cosmologies [61, 80]. Refer to Figure 5.5 for the marginalized posteriors from DESI, along with combinations of data from external probes. DESI cannot break the $w0 - wa$ degeneracy by itself due to a lack of constraining power. The addition of CMB and SN Ia data allows for better constraints, breaking the degeneracy. This is also visualized in the right panel of Figure 5.5. DESI presents the following constraints for this model using three different SN datasets. Using PantheonPlus,

$$\left.\begin{aligned} w_0 &= -0.827 \pm 0.0063 \\ w_a &= -0.75^{+0.29}_{-0.25} \end{aligned}\right\} \text{DESI + CMB + PantheonPlus}$$

Using Union3,

$$\left.\begin{aligned} w_0 &= -0.64 \pm 0.11 \\ w_a &= -1.27^{+0.40}_{-0.34} \end{aligned}\right\} \text{DESI + CMB + Union3}$$

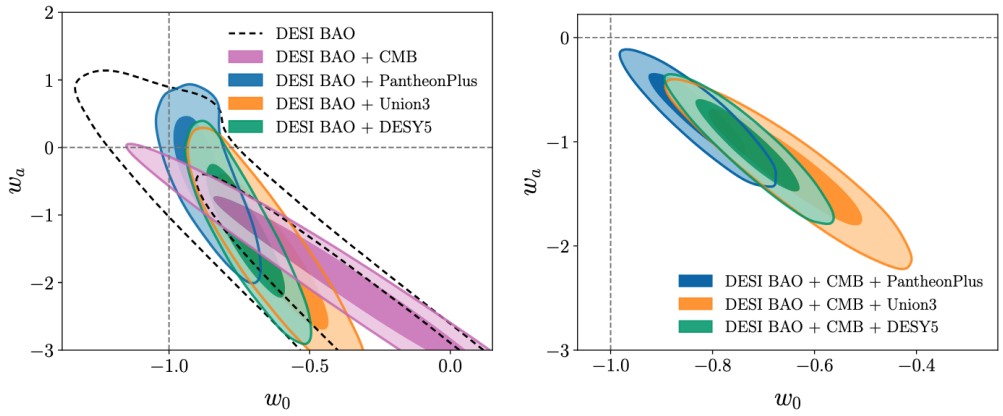


Figure 5.5: *Left panel:* 68% and 95% marginalized posterior constraints in the $w_0 - w_a$ plane for the flat $w_0 w_a$CDM cosmology. *Right panel:* 68% and 95% marginalized posterior constraints from DESI BAO combined with CMB and different SN datasets. The significance of the tension with ΛCDM is estimated to be 2.5σ, 3.5σ and 3.9σ, respectively, with the addition of PantheonPlus, Union3 and DESY5 SN Ia datasets. Caption adapted from and figure reproduced from [3].

Using DES Y5,

$$\left.\begin{aligned} w_0 &= -0.727 \pm 0.0067 \\ w_a &= -1.05^{+0.31}_{-0.27} \end{aligned}\right\} \text{DESI + CMB + DESY5}$$

The authors of [3] caution against over-interpretation of these results since these constraints could be mapped onto many different smooth dark energy models. While these constraints don't strongly favor a specific dark energy model, they do provide tantalizing hints of departures from ΛCDM.

5.4 Full Shape Analysis

DESI's full shape analysis aimed to extract cosmological information from features beyond just the BAO peak. The first few multipoles of the Fourier-space power spectra were measured. These contain information from redshift-space distortions. The power spectrum covariance matrix was estimated from 1000 `EZmocks`, as described in preceding

chapters. For the published analysis, DESI used a perturbation-theory approach to full-shape clustering analysis called Full Modeling (see Chapter 2). To briefly recall, Full Modeling directly fits a model to the full shape of the power spectrum multipoles [31, 51, 22]. The linear matter power spectrum is modeled using a set of cosmological parameters, coupled with nuisance parameters to describe the anisotropic power spectrum in the mildly nonlinear regime.

Full shape techniques are able to extract information from baryon acoustic oscillations, redshift-space distortions and, generally, any broadband shape imprinted in the two-point function. As in the BAO-only analysis, reconstruction techniques are used by DESI to sharpen the BAO peak. The 'Gaussianization' involved in reconstruction results in the pre- and post-reconstruction catalogs containing different information. The two-point statistics of the post-reconstruction catalog hold information from the three and four-point statistics as well [100]. So to use both the post-reconstruction full shape information and the post-reconstruction BAO information, DESI needs to account for the correlation between the two catalogs.

For the results published in [2], DESI opted to combine the post-reconstruction BAO information at the compressed parameter level, $\alpha_\perp$ and $\alpha_\parallel$, with the power spectrum bins in the pre-reconstruction catalog for Full Modeling. This removed the need to compute the post-reconstruction statistics during likelihood evaluation. The full BAO post-reconstruction information was thus compressed into just two parameters that were added to the pre-reconstruction data vector.

5.5 Dark Energy Constraints From Full Shape Analysis

The results in this section are quoted from [4]. This analysis focused on two models for dark energy: ΛCDM and w_0w_aCDM, both of which have been introduced before in this dissertation.

Data or Model	Parameter	Default	Prior
DESI (ΛCDM)	$\Omega_b h^2$	—	$\mathcal{N}(0.02218, 0.00055^2)$
	n_s	—	$\mathcal{N}(0.9649, 0.042^2)$
	$\Omega_{cdm} h^2$	—	$\mathcal{U}[0.001, 0.99]$
	$\ln 10^{10} A_s$	—	$\mathcal{U}[1.61, 3.91]$
CMB (ΛCDM)	$100\theta_{\rm MC}$	—	$\mathcal{U}[0.5, 10]$
	τ	0.0544	$\mathcal{U}[0.01, 0.8]$
	$\Omega_b h^2$	—	$\mathcal{U}[0.005, 0.1]$
	n_s	—	$\mathcal{U}[0.8, 1.2]$
Beyond ΛCDM	w_0	-1	$\mathcal{U}[-3, 1]$
(dynamical DE)	w_a	0	$\mathcal{U}[-3, 2]$
(massive ν)	$\sum m_\nu$ (eV)	0.06	$\mathcal{U}[0, 5]$
	μ_0	0	$\mathcal{U}[-3, 3]$
(modified gravity)	Σ_0	0	$\mathcal{U}[-3, 3]$

Table 5.3: Parameters and priors used in DESI Y1 Full Shape analysis. $\mathcal{U}$ refers to a uniform prior within the given range, and $\mathcal{N}$ refers to a Gaussian normal prior with the given mean and standard deviation squared. Table adapted from [4].

5.5.1 Flat ΛCDM

Figure 5.6 presents the constraints from DESI data alone in two different planes from the combined DESI full shape and BAO analysis, marginalized over all other parameters. Measurements from individual tracers are shown along with the combined constraints. There is excellent mutual agreement between different tracers. The combined measurements, along with a BBN (big bang nucleosynthesis) prior, give

$$\left.\begin{aligned} \Omega_m &= 0.2962 \pm 0.0095 \\ \sigma_8 &= 0.842 \pm 0.034 \\ H_0 &= (68.56 \pm 0.75)\ \text{km s}^{-1}\ \text{Mpc}^{-1} \end{aligned}\right\} \text{DESI (FS+BAO) + BBN + } n_{s10} \qquad (5.4)$$

Note how using information from the broadband shape of the power spectrum improves the constraints on Ω_m significantly, compared to BAO analysis alone. There is additional information gained on σ_8, the amplitude of mass fluctuations (Equation 1.26), which was not available through just the BAO analysis because the baryon acoustic oscillations are

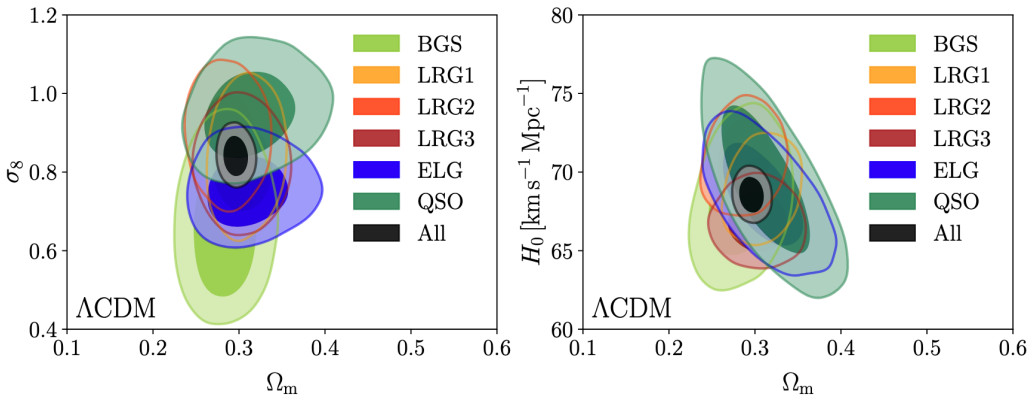


Figure 5.6: 68% and 95% credible intervals from the combined DESI full shape and BAO analysis, assuming ΛCDM. Caption adapted and figure reproduced from [4].

a geometric probe insensitive to this quantity. This can also be probed by CMB. When this CMB information is combined with DESI, the following constraints are yielded:

$$\left.\begin{aligned} \Omega_m &= 0.3056 \pm 0.0049 \\ \sigma_8 &= 0.8121 \pm 0.0053 \\ H_0 &= (68.07 \pm 0.38)\ \text{km s}^{-1}\ \text{Mpc}^{-1} \end{aligned}\right\} \text{DESI (FS+BAO) + CMB} \tag{5.5}$$

These constraints display full compatibility with just the DESI full shape and BAO results.

5.5.2 Flat w_0w_aCDM

For the w_0w_aCDM model, [4] does not provide measurements from DESI full shape and BAO alone because these are noticeably affected by parameter projection effects. Projection effects usually occur in the presence of long degeneracy directions in parameter space, when the posterior exhibits strong non-Gaussianity. This can lead to the mean of the marginalized posterior to be significantly offset from the maximum of the posterior. These show the fact that the statistical constraints on relevant cosmological parameters can be weak due to the presence of degeneracies, which can be created by

the correlation between the cosmological and nuisance parameters. And in the case of this model, there are many more nuisance parameters than there were in the equivalent model for just the BAO analysis.

Thus, to improve constraints, the results are presented with the addition of information from the CMB and different SN Ia datasets. Using PantheonPlus gives the following results:

$$\left.\begin{aligned} w_0 &= -0.858 \pm 0.061 \\ w_a &= -0.68^{+0.27}_{-0.23} \end{aligned}\right\} \text{DESI (FS+BAO) + CMB + PantheonPlus} \tag{5.6}$$

Using Union3 gives the following results:

$$\left.\begin{aligned} w_0 &= -0.742 \pm 0.096 \\ w_a &= -1.02^{+0.36}_{-0.32} \end{aligned}\right\} \text{DESI (FS+BAO) + CMB + Union3} \tag{5.7}$$

Using DESI-SN5YR gives the following results:

$$\left.\begin{aligned} w_0 &= -0.761 \pm 0.065 \\ w_a &= -0.96^{+0.30}_{-0.26} \end{aligned}\right\} \text{DESI (FS+BAO) + CMB + DESI-SN5YR} \tag{5.8}$$

For a summary of results, refer to Figure 5.7.

5.6 Summary

To sum up, the DESI full shape and BAO data, both by themselves and combined, along with information from the CMB and SN Ia, show signs of departures from the concordance ΛCDM model. The extent of this difference depends on which SN Ia dataset is used. This preference has been studied in analyses of BOSS data with similar external datasets as DESI [25]. These departures from ΛCDM are also encouraging to the pursuit of models with even more freedom in dark energy than in w_0w_aCDM [18, 63].

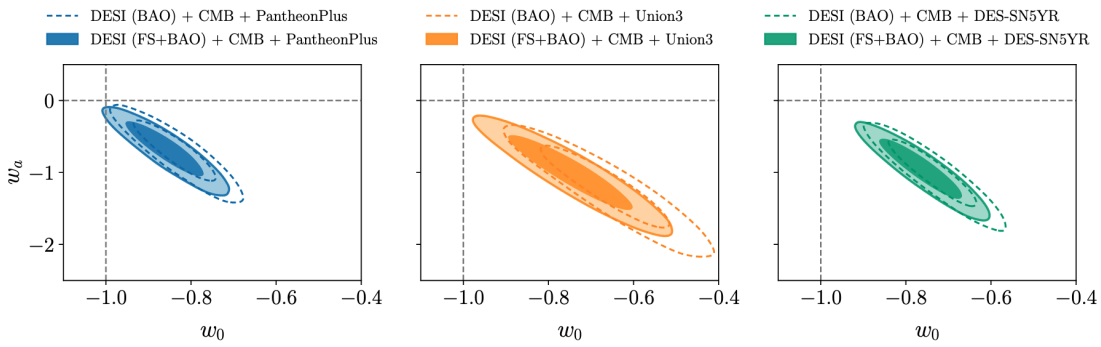


Figure 5.7: Constraints on w_0 and w_a assuming a w_0w_aCDM model with a time-varying dark energy equation of state parametrization. The contours represent the 68% and 95% credible intervals. The solid blue, orange and green contours represent the combination of DESI (FS+BAO) and CMB with three respective SN Ia data sets: PantheonPlus, Union3 and DES-SN5YR. The dashed blue, orange and green contours show the same respective combinations, but with the DESI full shape clustering and BAO replaced by DESI (BAO). The figure shows how the addition of the full shape information to the BAO-only data improves the precision of the constraints. The measurements of these two parameters remain mutually consistent, and prefer $w_0 > -1$ and $w_a < 0$. Caption and figure reproduced from [4].

These results show a preference for an effective dark energy equation of state that is apparently in the phantom regime ($w(z) < -1$) at $z > 1$. Phantom dark energy is a theoretical form of dark energy with the aforementioned equation of state parameter upper bound. This means its energy density grows stronger over time, potentially leading to Big Rip scenarios. Phantom models often violate standard energy conditions, leading to 'ghost' fields with negative kinetic energy, which can introduce theoretical instabilities and conflict with quantum field theory.

But as mentioned before, the authors of [3] caution against over-interpretation of these results as at least some of the models preferred by the DESI analysis could be mapped to quintessence models which reproduce the observables D_M/r_d and D_H/r_d to high precision even though they do not violate the null energy condition at any redshift [80]. Quintessence is a theoretical form of dark energy modeled by a dynamic scalar field that evolves over time. Unlike a simple cosmological constant, quintessence

can have w that varies with time. This changing nature arises from how the scalar field rolls down its potential, allowing quintessence to potentially explain the universe's accelerating expansion in a more flexible way than a constant dark energy term. An effective equation of state like this can also be obtained from modified gravity models.

Together, all of this leaves great room for further exploration when it comes to dark energy models, especially on the theory side. Looking ahead, BAO and full shape analyses from DESI's future observations should provide major new information along with even better constraints on the parameters and models discussed in this section, shining an even brighter light on dark energy.

CHAPTER 6

Concluding Remarks

Modern cosmology, an intensely data-driven field, draws on the observations of large galaxy surveys such as DESI. Such surveys are big collaborations with complex pipelines to accommodate all the processing that is required to get the final cosmological constraints, starting from instrumentation all the way to getting values of, say, the Hubble constant and the matter density in the universe. These constraints are often presented in the form of contour plots with ellipses to show the quality of fit to data, for example in Figure 5.6. The shapes of these ellipses are determined by the uncertainties on measurement.

In particular, extracting such measurements from baryon acoustic oscillations (BAO) and full shape (FS) analyses depends on measuring the clustering of galaxies as measured by DESI in the form of two-point correlation functions or power spectra. A huge limitation upon us is that we live in a single realization of the universe and do not have access to more to have repeated draws of data so that we can quantify the uncertainties on our measurement. Thus, collaborations such as DESI rely on alternative methods to estimate the covariance matrix of the correlation function (note that I use 'power spectrum' and 'correlation function' interchangeably). To do so, DESI uses approximate simulations of the measured universe in the form of large numbers of `EZmocks` to estimate the covariance matrix of the correlation function.

With the correlation function and its covariance matrix in hand, fits to the data can be performed and cosmological quantities are extracted in the form of dilation parameters (see Equation 2.1), which are themselves functions of cosmological quantities. Thus, the uncertainties on these dilation parameters are directly connected to the quality of the clustering covariance matrix. Finally, Markov Chain Monte Carlo (MCMC) chains are run to extract the cosmological parameters such as H_0 and results are presented in the aforementioned contour plots. This description summarizes how the clustering covariance matrix connects to the quality of cosmological fits at the end of the DESI cosmological analysis pipeline.

Instruments such as DESI are subject to physical constraints and systematics. One such very important consideration is the process of fiber assignment, which selects which targets on the sky actually get observed when there is competition between many different ones. The `EZmocks` used to calculate the clustering covariance matrix must reflect survey realism to be an accurate representation of the quality of real data. To generate large numbers of such mock catalogs, an efficient method for simulating fiber assignment was required by DESI.

In this dissertation, in Chapter 1, I started with a survey of cosmological theory, building up to large scale structure (LSS) and how it is probed by studying the clustering of galaxies, which is strongly influenced by effects such as BAO and RSD. I also gave a brief introduction to DESI. In Chapter 2, I summarized different ways of analyzing clustering to connect it to cosmology via BAO and full shape analyses. I also discussed methods for estimating covariance matrices, both for a general overview and to introduce the FitCov method that I would make use of later. In Chapter 3, I discussed the fiber assignment process in more detail and introduced fast fiber assignment and the Fast Fiber Assign (FFA) emulator, along with results of its performance on Abacus mocks. I also described my contributions to and impact on the DESI Y1 analysis. This chapter is based on my work in [10], a paper on which I am the second author.

In Chapter 4, I explored the performance of the emulator in more detail by studying how it performs when different meta-parameters are modified. I especially studied how these effects influence ShapeFit constraints on `EZmocks` through estimated covariance matrices. Finally, in Chapter 5, I presented dark energy results from DESI Y1 to show the final cosmological constraints the collaboration extracted. This serves to provide concrete constraints that my work in fiber assignment emulation and mock development contributed to.

The FFA algorithm was designed not only to reproduce the complex outcomes of fiber assignment but to do so in a manner that is computationally efficient. I showed various plots comparing the FFA emulator's performance on DESI `Abacus` mocks (a set of high-fidelity mocks) to real data and/or altMTL mocks. The altMTL mocks are an alternative set of `Abacus` mocks with a different method to simulate fiber assignment applied to them. This technique is highly accurate but computationally expensive but it provides a baseline set of mocks to compare the performance of FFA to. I noted deficiencies in the performance of FFA, especially focusing on the quality of the 'collision window' and the underestimation of the mock covariance matrix for the DESI Y1 analysis. The collision window is a direct consequence of the fact that DESI's fibers have a finite minimum separation that they can be placed at.

To summarize my results, I found that the amount of correlation re-injected back into the emulated catalogs at the end carries a lot of weight when looking at constraints. Tilting the kernel affects the results as well. In general, we do not want to warp the kernel too much but a small amount of fine-tuning can be beneficial. More dense tracers are more susceptible to detrimental effects from such adjustments. I also found that matching the FFA collision window on small scales to that of the real data or altMTLs improves the constraints. With all of this, I put forward my suggestions for how the emulator should be run for the next full analysis so that these results can be analyzed with more rigorous covariance matrices.

For future work, I recommend a similar analysis with a larger set of emulator parameters. It would especially be of interest to conduct it with more fine-grained small β_1 values. I selected my parameter sets based on how closely the small scale pair counts matched the data but it would be worthwhile to study how the performance of constraints is affected even if the small scale clustering is not as good a match. This is motivated by how, for QSOs, the fiducial set of emulator parameters' constraints are still a good match for the analytic covariance's constraints out of the sets I studied. The small scale pair counts for this fiducial set were not as good a match as some of the other sets I used.

Also, I recommend doing the same tests on ELGs as well. Considering their very high density, it would be interesting to see if they are even more sensitive to modifying the kernel. Both these suggested analyses require more time and computational cost. I also briefly mentioned unusual behavior in the low redshift bin of the `EZmocks` and that other studies have noticed this as well when running different covariance analyses. It would be worthwhile investigating the `EZmocks` themselves to see if there is a feature there causing this.

The process of emulating fiber assignment itself could also be modified. In its current form, FFA works with one tracer at a time. That is, it learns the kernel for one tracer and is then applied to a catalog of the same tracer only. This can be extended to a multi-tracer strategy as well. For the real data, as some ELGs get promoted to the higher priority of LRGs (to assist in the extraction of the LRG-ELG cross-correlation signal), being able to process both tracers simultaneously would be an improvement.

I also believe it would be worthwhile to use a machine-learning approach such as using a neural network to emulate the assignment. This would especially be interesting in the above case of an emulator that is trained on all tracers at the same time to see what kinds of cross-correlations the model can learn and whether it recognizes any other parameters as being important to the fiber assignment procedure beyond the count of

close friends and number of tiles a target appears on. Building on this, I encourage training any future versions of the emulator on combinations of data and altMTL mocks to increase the size of the training sets to see how this affects performance. This training can also solely be done on `Abacus` mocks that have been processed through the altMTL pipeline. Any major differences in performance from this compared to the real data might point to differences in the `Abacus` mocks that are better captured this way. All of these studies should contribute to a better understanding of the fast fiber assignment procedure and thus better estimates of mock-based covariances for any future survey analyses.

As cosmological surveys grow in the size of the data they collect, the importance of creating realistic simulations and mocks will only grow due to the requirements for increased precision. The important of fiber assignment and studying its effects will remain important even if the geometry of instruments differs as there are always physical constraints that contribute to missed observations on the sky. With increasingly large simulated datasets needed, it will also become ever more vital to produce accurate mock catalogs as opposed to running computationally and environmentally expensive N-body simulations. Therefore, machine learning will only grow in significance for applications to cosmology. Fast fiber assignment has been an effective step along this direction with my work and that of my collaborators contributing valuably to the efforts of a major international research collaboration such as DESI.

BIBLIOGRAPHY

[1] A. G. Adame et al. "DESI 2024 II: Sample Definitions, Characteristics, and Two-point Clustering Statistics". *TBD* (Nov. 2024).

[2] A. G. Adame et al. "DESI 2024 V: Full-Shape Galaxy Clustering from Galaxies and Quasars". *TBD* (Nov. 2024).

[3] A. G. Adame et al. "DESI 2024 VI: cosmological constraints from the measurements of baryon acoustic oscillations". *Journal of Cosmology and Astroparticle Physics* 2025.02 (Feb. 2025).
DOI: 10.1088/1475-7516/2025/02/021.

[4] A. G. Adame et al. "DESI 2024 VII: Cosmological Constraints from the Full-Shape Modeling of Clustering Measurements". *TBD* (Nov. 2024).

[5] N. Aghanim et al. "Planck2018 results: VI. Cosmological parameters". *Astronomy & Astrophysics* 641 (Sept. 2020), A6.
DOI: 10.1051/0004-6361/201833910.

[6] S. Alam et al. "Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory". *Physical Review D* 103.8 (Apr. 2021).
DOI: 10.1103/physrevd.103.083533.

[7] S. Alam et al. "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample". *Monthly Notices of the Royal Astronomical Society* 470.3 (Sept. 2017), pp. 2617–2652.
DOI: 10.1093/mnras/stx721.

[8] O. Alves et al. "Analytical covariance matrices of DESI galaxy power spectra". *In preparation* (2025).

[9] É. Aubourg et al. "Cosmological implications of baryon acoustic oscillation measurements". *Phys. Rev. D* 92 (12 Dec. 2015), p. 123516.
DOI: 10.1103/PhysRevD.92.123516.

[10] D. Bianchi et al. "Characterization of DESI fiber assignment incompleteness effect on 2-point clustering and mitigation methods for DR1 analysis". *TBD* (Nov. 2024).

[11] D. Bianchi and W. J. Percival. "Unbiased clustering estimation in the presence of missing observations". *Monthly Notices of the Royal Astronomical Society* 472.1 (Nov. 2017), pp. 1106–1118.
DOI: 10.1093/mnras/stx2053.

[12] D. Bianchi and L. Verde. "Confronting missing observations with probability weights: Fourier space and generalized formalism". *Monthly Notices of the Royal Astronomical Society* 495.1 (May 2020), pp. 1511–1529.
DOI: 10.1093/mnras/staa1267.

[13] C. Blake, P. Carter, and J. Koda. "Power spectrum multipoles on the curved sky: an application to the 6-degree Field Galaxy Survey". *Monthly Notices of the Royal Astronomical Society* 479.4 (Oct. 2018), pp. 5168–5183.
DOI: 10.1093/mnras/sty1814.

[14] D. Blas, J. Lesgourgues, and T. Tram. "The Cosmic Linear Anisotropy Solving System (CLASS). Part II: Approximation schemes". *JCAP* 2011.7, 034 (July 2011), p. 034.
DOI: 10.1088/1475-7516/2011/07/034.

[15] S. Brieden, H. Gil-Marín, and L. Verde. "A tale of two (or more) h's". *Journal of Cosmology and Astroparticle Physics* 2023.04 (Apr. 2023), p. 023.
DOI: 10.1088/1475-7516/2023/04/023.

[16] S. Brieden, H. Gil-Marín, and L. Verde. "ShapeFit: extracting the power spectrum shape information in galaxy surveys beyond BAO and RSD". *Journal of Cosmology and Astroparticle Physics* 2021.12 (Dec. 2021), p. 054.
DOI: 10.1088/1475-7516/2021/12/054.

[17] A. Burden et al. "Mitigating the Impact of the DESI Fiber Assignment on Galaxy Clustering". *JCAP* 03 (Mar. 2017), p. 001.
DOI: 10.1088/1475-7516/2017/03/001.

[18] R. Calderon et al. "DESI 2024: reconstructing dark energy using crossing statistics with DESI DR1 BAO data". *Journal of Cosmology and Astroparticle Physics* 2024.10 (Oct. 2024), p. 048.
DOI: 10.1088/1475-7516/2024/10/048.

[19] E. Chaussidon et al. "Constraining primordial non-Gaussianity with DESI 2024 LRG and QSO samples". *TBD* (Nov. 2024).

[20] E. Chaussidon et al. "Target Selection and Validation of DESI Quasars". *The Astrophysical Journal* 944.1, 107 (Feb. 2023), p. 107.
DOI: 10.3847/1538-4357/acb3c2.

[21] S. .-. Chen et al. "Baryon acoustic oscillation theory and modelling systematics for the DESI 2024 results". *Monthly Notices of the Royal Astronomical Society* 534.1 (Sept. 2024).
DOI: 10.1093/mnras/stae2090.

[22] S.-F. Chen, Z. Vlah, and M. White. "A new analysis of galaxy 2-point functions in the BOSS survey, including full-shape information and post-reconstruction BAO". *Journal of Cosmology and Astroparticle Physics* 2022.02 (Feb. 2022), p. 008.
DOI: 10.1088/1475-7516/2022/02/008.

[23] S.-F. Chen, Z. Vlah, and M. White. "Consistent Modeling of Velocity Statistics and Redshift-Space Distortions in One-Loop Perturbation Theory". *JCAP* 07 (July 2020), p. 062.
DOI: 10.1088/1475-7516/2020/07/062.

[24] S.-F. Chen et al. "Redshift-space distortions in Lagrangian perturbation theory". *Journal of Cosmology and Astroparticle Physics* 2021.03 (Mar. 2021), p. 100.
DOI: 10.1088/1475-7516/2021/03/100.

[25] S.-F. Chen et al. "Suppression without Thawing: Constraining Structure Formation and Dark Energy with Galaxy Clustering". *Phys. Rev. Lett.* 133 (23 Dec. 2024), p. 231001.
DOI: 10.1103/PhysRevLett.133.231001.

[26] C.-H. Chuang et al. "EZmocks: extending the Zel'dovich approximation to generate mock galaxy catalogues with accurate clustering statistics". *Monthly Notices of the Royal Astronomical Society* 446.3 (Nov. 2014), pp. 2621–2628.
DOI: 10.1093/mnras/stu2301.

[27] C.-H. Chuang et al. "nIFTy cosmology: Galaxy/halo mock catalogue comparison project on clustering statistics". *Monthly Notices of the Royal Astronomical Society* 452.1 (Sept. 2015), pp. 686–700.
DOI: 10.1093/mnras/stv1289.

[28] M. Colavincenzo et al. "Comparing approximate methods for mock catalogues and covariance matrices - III: bispectrum". *Monthly Notices of the Royal Astronomical Society* 482.4 (Feb. 2019), pp. 4883–4905.
DOI: 10.1093/mnras/sty2964.

[29] D. Collaboration et al. "Overview of the Instrumentation for the Dark Energy Spectroscopic Instrument". *The Astronomical Journal* 164.5 (Oct. 2022), p. 207.
DOI: 10.3847/1538-3881/ac882b.

[30] A. J. Cuesta et al. "Calibrating the cosmic distance scale ladder: the role of the sound-horizon scale and the local expansion rate as distance anchors". *Monthly Notices of the Royal Astronomical Society* 448.4 (Mar. 2015), pp. 3463–3471.
DOI: 10.1093/mnras/stv261.

[31] G. d'Amico et al. "The cosmological analysis of the SDSS/BOSS data from the Effective Field Theory of Large-Scale Structure". *Journal of Cosmology and Astroparticle Physics* 2020.05 (May 2020), pp. 005–005.
DOI: 10.1088/1475-7516/2020/05/005.

[32] E. Darragh-Ford et al. "Target Selection and Sample Characterization for the DESI LOW-Z Secondary Target Program". *The Astrophysical Journal* 954.2 (Sept. 2023).
DOI: 10.3847/1538-4357/ace902.

[33] DESI Collaboration et al. "The DESI Early Data Release white dwarf catalogue". English. *Monthly Notices of the Royal Astronomical Society* 535.1 (Nov. 2024), pp. 254–289.
DOI: 10.1093/mnras/stae2205.

[34] A. Dey et al. "DESI Observations of the Andromeda Galaxy: Revealing the Immigration History of Our Nearest Neighbor". *The Astrophysical Journal* 944.1 (Feb. 2023).
DOI: 10.3847/1538-4357/aca5f8.

[35] A. Dey et al. "Overview of the DESI Legacy Imaging Surveys". *The Astronomical Journal* 157.5 (Apr. 2019), p. 168.
DOI: 10.3847/1538-3881/ab089d.

[36] A. Einstein. "Cosmological Considerations in the General Theory of Relativity". *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* (Feb. 1917), pp. 142–152.

[37] D. J. Eisenstein and W. Hu. "Baryonic Features in the Matter Transfer Function". *The Astrophysical Journal* 496.2 (Apr. 1998), pp. 605–614.
DOI: 10.1086/305424.

[38] D. J. Eisenstein et al. "Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies". *The Astrophysical Journal* 633.2 (Nov. 2005), pp. 560–574.
DOI: 10.1086/466512.

[39] D. J. Eisenstein et al. "Improving Cosmological Distance Measurements by Reconstruction of the Baryon Acoustic Peak". *The Astrophysical Journal* 664.2 (Aug. 2007), pp. 675–679.
DOI: 10.1086/518712.

[40] G. Favole et al. "Does jackknife scale really matter for accurate large-scale structure covariances?" *Monthly Notices of the Royal Astronomical Society* 505.4 (June 2021), pp. 5833–5845.
DOI: 10.1093/mnras/stab1720.

[41] V. A. Fawcett et al. "A striking relationship between dust extinction and radio detection in DESI QSOs: evidence for a dusty blow-out phase in red QSOs". *Monthly Notices of the Royal Astronomical Society* 525.4 (Nov. 2023), pp. 5575–5596.
DOI: 10.1093/mnras/stad2603.

[42] A. Friedmann. "Über die Krümmung des Raumes". *Zeitschrift fur Physik* 10 (Jan. 1922), pp. 377–386.
DOI: 10.1007/BF01332580.

[43] O. Friedrich et al. "Performance of internal covariance estimators for cosmic shear correlation functions". *Monthly Notices of the Royal Astronomical Society* 456.3 (Dec. 2015).
DOI: 10.1093/mnras/stv2833.

[44] C. Garcia-Quintero et al. "HOD-dependent systematics in Emission Line Galaxies for the DESI 2024 BAO analysis". *Journal of Cosmology and Astroparticle Physics* 2025.01 (Jan. 2025), p. 132.
DOI: 10.1088/1475-7516/2025/01/132.

[45] L. H. Garrison et al. "The abacus cosmological N-body code". *Mon. Not. Roy. Astron. Soc.* 508.1 (Sept. 2021), pp. 575–596.
DOI: 10.1093/mnras/stab2482.

[46] A. H. Guth. "Inflationary universe: A possible solution to the horizon and flatness problems". *Phys. Rev. D* 23.2 (Jan. 1981), pp. 347–356.
DOI: 10.1103/PhysRevD.23.347.

[47] J. Guy et al. "The Spectroscopic Data Processing Pipeline for the Dark Energy Spectroscopic Instrument". *The Astronomical Journal* 165.4 (Mar. 2023), p. 144.
DOI: 10.3847/1538-3881/acb212.

[48] C. Hahn et al. "The DESI Bright Galaxy Survey: Final Target Selection, Design, and Validation". *The Astronomical Journal* 165.6 (May 2023).
DOI: 10.3847/1538-3881/accff8.

[49] C. Howlett et al. "CMB power spectrum parameter degeneracies in the era of precision cosmology". *JCAP* 2012.4, 027 (Apr. 2012), p. 027.
DOI: 10.1088/1475-7516/2012/04/027.

[50] E. Hubble. "A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae". *Proceedings of the National Academy of Science* 15.3 (Mar. 1929), pp. 168–173.
DOI: 10.1073/pnas.15.3.168.

[51] M. M. Ivanov, M. Simonović, and M. Zaldarriaga. "Cosmological parameters from the BOSS galaxy power spectrum". *Journal of Cosmology and Astroparticle Physics* 2020.05 (May 2020), pp. 042–042.
DOI: 10.1088/1475-7516/2020/05/042.

[52] N. Kaiser. "Clustering in real space and in redshift space". *MNRAS* 227 (July 1987), pp. 1–21.
DOI: 10.1093/mnras/227.1.1.

[53] A. Krolewski et al. "Impact and mitigation of spectroscopic systematics on DESI DR1 clustering measurements". *Journal of Cosmology and Astroparticle Physics* 2025.01 (Jan. 2025), p. 147.
DOI: 10.1088/1475-7516/2025/01/147.

[54] T.-W. Lan et al. "The DESI Survey Validation: Results from Visual Inspection of Bright Galaxies, Luminous Red Galaxies, and Emission-line Galaxies". *The Astrophysical Journal* 943.1 (Jan. 2023), p. 68.
DOI: 10.3847/1538-4357/aca5fa.

[55] S. D. Landy and A. S. Szalay. "Bias and Variance of Angular Correlation Functions". *The Astrophysical Journal* 412 (July 1993), p. 64.
DOI: 10.1086/172900.

[56] J. Lasker et al. "Production of alternate realizations of DESI fiber assignment for unbiased clustering measurement in data and simulations". *Journal of Cosmology and Astroparticle Physics* 2025.01 (Jan. 2025), p. 127.
DOI: 10.1088/1475-7516/2025/01/127.

[57] G. Lemaitre. "A Homogeneous Universe of Constant Mass and Growing Radius Accounting for the Radial Velocity of Extragalactic Nebulae". *Annales Soc. Sci. Bruxelles A* 47 (1927), pp. 49–59.
DOI: 10.1007/s10714-013-1548-3.

[58] A. Lewis. "Efficient sampling of fast and slow cosmological parameters". *Phys. Rev. D* 87 (10 May 2013), p. 103529.
DOI: 10.1103/PhysRevD.87.103529.

[59] A. Lewis and S. Bridle. "Cosmological parameters from CMB and other data: A Monte Carlo approach". *Phys. Rev. D* 66 (10 Nov. 2002), p. 103511.
DOI: 10.1103/PhysRevD.66.103511.

[60] A. Lewis, A. Challinor, and A. Lasenby. "Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models". *The Astrophysical Journal* 538.2 (Aug. 2000), pp. 473–476.
DOI: 10.1086/309179.

[61] E. V. Linder. "Exploring the Expansion History of the Universe". *Phys. Rev. Lett.* 90 (9 Mar. 2003), p. 091301.
DOI: 10.1103/PhysRevLett.90.091301.

[62] M. Lippich et al. "Comparing approximate methods for mock catalogues and covariance matrices – I. Correlation function". *Monthly Notices of the Royal Astronomical Society* 482.2 (Oct. 2018), pp. 1786–1806.
DOI: 10.1093/mnras/sty2757.

[63] K. Lodha et al. "DESI 2024: Constraints on physics-focused aspects of dark energy using DESI DR1 BAO data". *Phys. Rev. D* 111 (2 Jan. 2025), p. 023532.
DOI: 10.1103/PhysRevD.111.023532.

[64] N. A. Maksimova et al. "AbacusSummit: a massive set of high-accuracy, high-resolution N-body simulations". *Monthly Notices of the Royal Astronomical Society* 508.3 (Sept. 2021), pp. 4017–4037.
DOI: 10.1093/mnras/stab2484.

[65] J. Martin. "Everything you always wanted to know about the cosmological constant problem (but were afraid to ask)". *Comptes Rendus. Physique* 13.6–7 (July 2012), pp. 566–665.
DOI: 10.1016/j.crhy.2012.04.008.

[66] M. Maus et al. "An analysis of parameter compression and Full-Modeling techniques with Velocileptors for DESI 2024 and beyond". *JCAP* 01 (Jan. 2025), p. 138.
DOI: 10.1088/1475-7516/2025/01/138.

[67] P. McDonald et al. "The Lyα Forest Power Spectrum from the Sloan Digital Sky Survey". *The Astrophysical Journal Supplement Series* 163.1 (Mar. 2006), pp. 80–109.
DOI: 10.1086/444361.

[68] J. Mena-Fernández et al. "HOD-dependent systematics for luminous red galaxies in the DESI 2024 BAO analysis". *Journal of Cosmology and Astroparticle Physics* 2025.01 (Jan. 2025).
DOI: 10.1088/1475-7516/2025/01/133.

[69] F. G. Mohammad and W. J. Percival. "Creating jackknife and bootstrap estimates of the covariance matrix for the two-point correlation function". *Mon. Not. Roy. Astron. Soc.* 514.1 (May 2022), pp. 1289–1301.
DOI: 10.1093/mnras/stac1458.

[70] A. D. Myers et al. "The Target-selection Pipeline for the Dark Energy Spectroscopic Instrument". *Astron. J.* 165.2 (Feb. 2023), p. 50.
DOI: 10.3847/1538-3881/aca5f9.

[71] P. Norberg et al. "Statistical analysis of galaxy surveys – I. Robust error estimation for two-point clustering statistics". *Monthly Notices of the Royal Astronomical Society* 396.1 (June 2009), pp. 19–38.
DOI: 10.1111/j.1365-2966.2009.14389.x.

[72] N. Padmanabhan and M. White. "Constraining anisotropic baryon oscillations". *Physical Review D* 77.12, 123540 (June 2008), p. 123540.
DOI: 10.1103/PhysRevD.77.123540.

[73] N. Padmanabhan et al. "A 2 per cent distance toz= 0.35 by reconstructing baryon acoustic oscillations – I. Methods and application to the Sloan Digital Sky Survey: A 2 per cent distance to z = 0.35". *Monthly Notices of the Royal Astronomical Society* 427.3 (Nov. 2012), pp. 2132–2145.
DOI: 10.1111/j.1365-2966.2012.21888.x.

[74] E. Paillas et al. "Optimal reconstruction of baryon acoustic oscillations for DESI 2024". *Journal of Cosmology and Astroparticle Physics* 2025.01 (Jan. 2025), p. 142.
DOI: 10.1088/1475-7516/2025/01/142.

[75] R. Paviot et al. "Angular systematics-free cosmological analysis of galaxy clustering in configuration space". *Mon. Not. Roy. Astron. Soc.* 512.1 (Mar. 2022), pp. 1341–1356.
DOI: 10.1093/mnras/stac560.

[76] W. J. Percival and D. Bianchi. "Using angular pair upweighting to improve 3D clustering measurements". *MNRAS* 472.1 (Nov. 2017), pp. L40–L44.
DOI: 10.1093/mnrasl/slx135.

[77] A. Pérez-Fernández et al. "Fiducial-cosmology-dependent systematics for the DESI 2024 BAO analysis". *JCAP* 01 (Jan. 2025), p. 144.
DOI: 10.1088/1475-7516/2025/01/144.

[78] S. Perlmutter et al. "Measurements of Omega and Lambda from 42 High-Redshift Supernovae". *The Astrophysical Journal* 517.2 (June 1999), pp. 565–586.
DOI: 10.1086/307221.

[79] M. Pinon et al. "Mitigation of DESI fiber assignment incompleteness effect on two-point clustering with small angular scale truncated estimators". *Journal of Cosmology and Astroparticle Physics* 2025.01 (Jan. 2025), p. 131.
DOI: 10.1088/1475-7516/2025/01/131.

[80] R. de Putter and E. V. Linder. "Calibrating dark energy". *Journal of Cosmology and Astroparticle Physics* 2008.10 (Oct. 2008), p. 042.
DOI: 10.1088/1475-7516/2008/10/042.

[81] A. Raichoor et al. "Target Selection and Validation of DESI Emission Line Galaxies". *The Astronomical Journal* 165.3, 126 (Mar. 2023), p. 126.
DOI: 10.3847/1538-3881/acb213.

[82] M. Rashkovetskyi et al. "Semi-analytical covariance matrices for two-point correlation function for DESI 2024 data". *Journal of Cosmology and Astroparticle Physics* 2025.01 (Jan. 2025).
DOI: 10.1088/1475-7516/2025/01/145.

[83] M. Rashkovetskyi et al. "Validation of semi-analytical, semi-empirical covariance matrices for two-point correlation function for early DESI data". *Monthly Notices of the Royal Astronomical Society* 524.3 (Sept. 2023), pp. 3894–3911.
DOI: 10.1093/mnras/stad2078.

[84] A. G. Riess et al. "Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant". *The Astronomical Journal* 116.3 (Sept. 1998), pp. 1009–1038.
DOI: 10.1086/300499.

[85] H. P. Robertson. "Kinematics and World-Structure". *Astrophys. J.* 82 (1935), pp. 284–301.
DOI: 10.1086/143681.

[86] A. Rocher et al. "The DESI One-Percent survey: exploring the Halo Occupation Distribution of Emission Line Galaxies with AbacusSummit simulations". *Journal of Cosmology and Astroparticle Physics* 2023.10 (Oct. 2023), p. 016. DOI: 10.1088/1475-7516/2023/10/016.

[87] A. J. Rosado-Marín et al. "Mitigating Imaging Systematics for DESI 2024 Emission Line Galaxies and Beyond". *TBD* (Nov. 2024).

[88] C. Saulder et al. "Target selection for the DESI Peculiar Velocity Survey". *Monthly Notices of the Royal Astronomical Society* 525.1 (July 2023). DOI: 10.1093/mnras/stad2200.

[89] E. F. Schlafly et al. "Survey Operations for the Dark Energy Spectroscopic Instrument". *The Astronomical Journal* 166.6 (Nov. 2023), p. 259. DOI: 10.3847/1538-3881/ad0832.

[90] M. Schubnell et al. "The DESI fiber positioner system". Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series 9908, 990892 (Aug. 2016). Ed. by C. J. Evans, L. Simard, and H. Takami, p. 990892. DOI: 10.1117/12.2233370.

[91] J. H. Silber et al. "The Robotic Multiobject Focal Plane System of the Dark Energy Spectroscopic Instrument (DESI)". *Astron. J.* 165.1 (Jan. 2023), p. 9. DOI: 10.3847/1538-3881/ac9ab1.

[92] V. M. Slipher. "The radial velocity of the Andromeda Nebula". *Lowell Observatory Bulletin* 2.8 (Jan. 1913), pp. 56–57.

[93] A. Slosar et al. "The Lyman-alpha forest in three dimensions: measurements of large scale flux correlations from BOSS 1st-year data". *JCAP* 09 (Sept. 2011), p. 001. DOI: 10.1088/1475-7516/2011/09/001.

[94] G. F. Smoot et al. "Structure in the COBE differential microwave radiometer first year maps". *Astrophys. J. Lett.* 396 (Sept. 1992), pp. L1–L5. DOI: 10.1086/186504.

[95] M. T. Soumagnac et al. "The MOST Hosts Survey: Spectroscopic Observation of the Host Galaxies of 40,000 Transients Using DESI". *The Astrophysical Journal. Supplement Series* 275.2 (Nov. 2024). DOI: 10.3847/1538-4365/ad76ae.

[96] J. Torrado and A. Lewis. "Cobaya: code for Bayesian analysis of hierarchical physical models". *Journal of Cosmology and Astroparticle Physics* 2021.05 (May 2021), p. 057. DOI: 10.1088/1475-7516/2021/05/057.

[97] S. Trusov et al. "The two-point correlation function covariance with fewer mocks". *Monthly Notices of the Royal Astronomical Society* 527.3 (Nov. 2023), pp. 9048–9060. DOI: 10.1093/mnras/stad3710.

[98] M. Valluri et al. "GD-1 Stellar Stream and Cocoon in the DESI Early Data Release". *The Astrophysical Journal* 980.1 (Feb. 2025).
DOI: 10.3847/1538-4357/ada690.

[99] A. G. Walker. "On Milne's Theory of World-Structure". *Proc. Lond. Math. Soc. s* 2-42.1 (1937), pp. 90–127.
DOI: 10.1112/plms/s2-42.1.90.

[100] Y. Wang et al. "Extracting high-order cosmological information in galaxy surveys with power spectra". *Communications Physics* 7.1 (Apr. 2024).
DOI: 10.1038/s42005-024-01624-7.

[101] J. Wishart. "The Generalised Product Moment Distribution in Samples from a Normal Multivariate Population". *Biometrika* 20A.1-2 (Dec. 1928), pp. 32–52.
DOI: 10.1093/biomet/20A.1-2.32.

[102] E. L. Wright et al. "The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance". *The Astronomical Journal* 140.6 (Nov. 2010), pp. 1868–1881.
DOI: 10.1088/0004-6256/140/6/1868.

[103] X. Yang et al. "An Extended Halo-based Group/Cluster Finder: Application to the DESI Legacy Imaging Surveys DR8". *The Astrophysical Journal* 909.2 (Mar. 2021).
DOI: 10.3847/1538-4357/abddb2.

[104] J. Yu et al. "ELG spectroscopic systematics analysis of the DESI Data Release 1". *JCAP* 01 (Jan. 2025), p. 126.
DOI: 10.1088/1475-7516/2025/01/126.

[105] S. Yuan et al. "The DESI one-per cent survey: exploring the halo occupation distribution of luminous red galaxies and quasi-stellar objects with ABACUS-SUMMIT". *MNRAS* 530.1 (May 2024), pp. 947–965.
DOI: 10.1093/mnras/stae359.

[106] I. Zehavi et al. "Galaxy Clustering in the Completed SDSS Redshift Survey: The Dependence on Color and Luminosity". *The Astrophysical Journal* 736.1 (July 2011), p. 59.
DOI: 10.1088/0004-637x/736/1/59.

[107] Y. B. Zel'dovich. "Gravitational instability: An approximate theory for large density perturbations." *A&A* 5 (Mar. 1970), pp. 84–89.

[108] C. Zhao et al. "Mock catalogues with survey realism for the DESI DR1". *In preparation* (2025).

[109] C. Zhao et al. "The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: 1000 multi-tracer mock catalogues with redshift evolution and systematics for galaxies and quasars of the final data release". *Monthly Notices of the Royal Astronomical Society* 503.1 (Feb. 2021), pp. 1149–1173.
DOI: 10.1093/mnras/stab510.

[110] R. Zhou et al. “Target Selection and Validation of DESI Luminous Red Galaxies”. *The Astronomical Journal* 165.2 (Jan. 2023).
DOI: 10.3847/1538-3881/aca5fb.

[111] F. Zwicky. “Die Rotverschiebung von extragalaktischen Nebeln”. *Helv. Phys. Acta* 6 (1933), pp. 110–127.
DOI: 10.1007/s10714-008-0707-4.