THE UNIVERSITY of EDINBURGH
School of Physics and Astronomy

Shrinkage Estimation of Covariance Matrices in Cosmology

Marnix Looijmans

Supervisor: Dr. Mike Wang

Introduction

The covariance matrix is an indispensable quantity for characterizing the uncertainties in correlated multi-dimensional datasets. Since the covariance matrix is usually not known a priori in cosmology, and since only a single data realization is available, it must be determined from mock data realizations. Usually this is done by computing the sample covariance matrix estimate. Shrinkage estimation is a class of techniques which aims to find estimates of the covariance matrix using fewer mock data realization while retaining a precision that is at least as good. Reducing the number of mocks required therefore reduces the computational costs of the data analysis. Two types of shrinkage estimators were studied in this project: **linear shrinkage estimation**[3] and **non-linear shrinkage (NERCOME) estimation**[2]. The differences are that linear shrinkage requires a guess of the shape of the true covariance matrix whereas NERCOME is computationally slightly more expensive.

Methods

Parameter fitting was performed using MCMC likelihood analysis of power spectrum measurements from the BOSS survey[1]. The power spectrum is a summary statistic that describes the distributions of galaxies in the universe and contains information about cosmological parameters. In this analysis, the Hubble parameter h, the bias parameter b, and the growth rate f were fitted. The likelihood function for the power spectrum is given by

$$\mathcal{L} \propto \exp\left[-\frac{1}{2}(\mathbf{P}^{\rm conv} - \mathbf{WMP}^{\rm true,flat-sky})^T \mathbf{C}^{-1}_{\rm conv}(\mathbf{P}^{\rm conv} - \mathbf{WMP}^{\rm true,flat-sky})\right]$$

The vector $\mathbf{P}^{\rm true,flat-sky}$ is a model of the power spectrum which is a function of the parameters being fitted. As can be seen from the equation above, the likelihood function depends on (the inverse of) the covariance matrix. In this project, different estimates of the covariance matrix were used while all other quantities in the likelihood analysis were kept the same. There were a total of 2048 mocks available. Linear shrinkage estimates, NERCOME estimates, and sample estimates of the covariance matrix were computed from different numbers of mocks. These were compared to the sample estimate that was computed from all 2048 mocks.

Results

Posterior constraints of the parameter distributions obtained from the MCMC analyses are shown in figure 1 for 20 and 40 mocks.

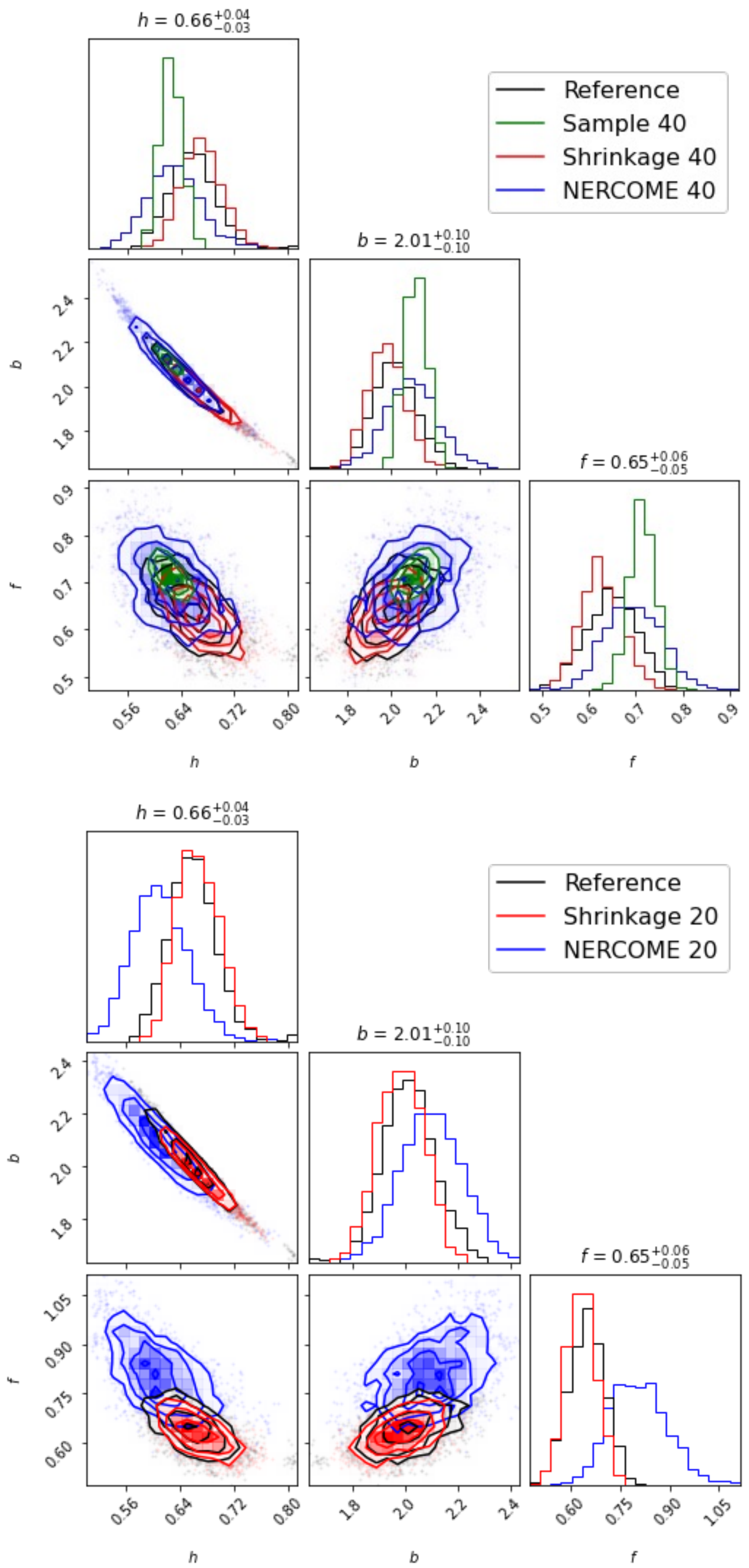

Figure 1: Posterior constraints of fitting distributions of h, b, and f where different estimates of the covariance matrix were used. The subplot titles are the means of the reference distributions.

Conclusions

The purpose of this research project was to explore if shrinkage estimation techniques can be used to find estimates of covariance matrices which outperform sample estimates. It can be seen in figure 1 that such estimates can indeed be found using fewer mock data realizations. When used in likelihood analysis, these estimates produced parameter fitting distributions that are similar to distributions obtained when all 2048 mock data realizations were used to find the sample estimate. The sample covariance matrix calculated from 40 mocks underestimates the uncertainties which can be seen in figure 1 since the parameter distributions are too narrow. Linear shrinkage and NERCOME estimates from 40 mocks yield distributions that are reasonably similar to the reference. The sample estimate computed from 20 mocks was singular and hence unstable upon inversion and is therefore not shown in figure 1. The NERCOME estimate from 20 mocks is biased in comparison to the reference, but the linear shrinkage estimate from 20 mocks yield distributions which are good approximations of the reference.

References

1. Beutler, F., & McDonald, P. (2021). Unified galaxy power spectrum measurements from 6dFGS, BOSS, and eBOSS. *Journal of Cosmology and Astroparticle Physics, 2021(11)*, 031.
2. Joachimi, B. (2017). Non-linear shrinkage estimation of large-scale structure covariance. *Monthly Notices of the Royal Astronomical Society: Letters*, *466*(1), L83–L87.
3. Pope, A. C., & Szapudi, I. (2008). Shrinkage Estimation of the Power Spectrum Covariance Matrix. *Monthly Notices of the Royal Astronomical Society, 389(2)*, 766–774.

Acknowledgements

This project received financial support from the School of Physics & Astronomy at the University of Edinburgh through the award of a Career Development Summer Scholarship. I would like to express my sincere gratitude to my supervisor, Mike, for offering me the opportunity to do this project and for his guidance and advice during the course of it. Finally I would like to thank Florian Beutler for helpful discussions.