



## BIROn - Birkbeck Institutional Research Online

---

Enabling Open Access to Birkbeck's Research Degree output

### The bootstrap in state space models

<https://eprints.bbk.ac.uk/id/eprint/52459/>

Version: Full Version

**Citation: Coppin, Kim Michelle (2023) The bootstrap in state space models. [Thesis] (Unpublished)**

© 2020 The Author(s)

---

All material available through BIROn is protected by intellectual property law, including copyright law.

Any use made of the contents should comply with the relevant law.

---

[Deposit Guide](#)  
Contact: [email](#)

# **The Bootstrap In State Space Models**

Kim Coppin

A thesis submitted for the degree of

Doctor of Philosophy

School of Computing and Mathematical Sciences

University of London, Birkbeck College

May 2023

# Acknowledgements

I wish to express my sincere gratitude to Professor Zacharias Psaradakis, my primary supervisor, for his invaluable assistance, counsel, and unwavering patience throughout this eventful journey. Additionally, I extend my thanks to Dr. Swati Chandna, my secondary supervisor, for her guidance and support.

# Dedication

This dissertation is dedicated to Dr. Michal Jezek, Mrs. Colleen A. Coppin and Mr. Lisle E. Coppin.

# Abstract

This dissertation offers some theoretical and practical contributions to the relatively sparse literature on the bootstrap in state space models.

In the first chapter, we present the rationale for our research and offer an overview of the structure of the document.

In Chapter 2, we introduce a gradient-based approach to the bootstrap which can be applied to heteroskedastic state space models. To the best of our knowledge, there is presently only one bootstrap technique available for models in state space form and it is limited to homoskedastic models. We evaluate the effectiveness of our methodology by analysing the coverage probabilities of different bootstrap confidence intervals and we find that they perform better than their asymptotic equivalents. We go on to use our methodology to modify an existing bootstrap model selection criterion so that it can be utilised with heteroskedastic state space models. Moreover, we introduce a bootstrap procedure that is suitable for homoskedastic state space models, providing an alternative to the current approach. Monte Carlo simulations demonstrate that this new method is significantly faster than the existing technique.

In Chapter 3, we apply an approximate bootstrap algorithm to state

space models for the purpose of conducting hypothesis tests. We illustrate the advantages of working within the state space framework by highlighting how Kalman filter output can aid in the construction of certain chi-squared test statistics. We then apply the approximate bootstrap approach to actual data and reveal how, in small to moderately sized samples, bootstrap test statistics can lead researchers to different conclusions than those drawn from using standard asymptotic theory.

Chapter 4 focuses on the transformation of multilevel models into state space form for the purpose of estimation. We show how the output from the Kalman filter recursions can facilitate the parametric and residual bootstraps in these models. In hierarchical settings, the full information maximum likelihood estimates of the variance elements in the covariance matrix of the vector of random effects are commonly biased downward in small samples. To alleviate this problem, researchers generally use restricted maximum likelihood estimation as it yields less biased estimates. However, we present an alternative solution by employing the parametric and residual bootstraps to correct for this bias in the estimates.

# Contents

<b>1</b>	<b>Introduction</b>	<b>12</b>
<b>2</b>	<b>The State Space Model Bootstrap</b>	<b>21</b>
2.1	Introduction . . . . .	21
2.2	Review of the Score Bootstrap . . . . .	24
2.2.1	The Wild Bootstrap . . . . .	25
2.2.2	The Score Bootstrap . . . . .	27
2.3	The Bootstrap of Stoffer and Wall . . . . .	31
2.4	Deriving the State Space Model Bootstrap . . . . .	33
2.4.1	The Wild State Space Model Bootstrap . . . . .	35
2.4.2	The Residual State Space Model Bootstrap . . . . .	37
2.5	Model Selection . . . . .	39
2.6	Simulations . . . . .	42
2.6.1	Simulation I . . . . .	42
2.6.2	Simulation I: Results and Discussion . . . . .	45
2.6.3	Simulation II . . . . .	46
2.6.4	Simulation II: Results and Discussion . . . . .	48
2.6.5	Simulation III . . . . .	50

2.6.6	Simulation III: Results and Discussion . . . . .	51
2.7	Concluding Remarks . . . . .	52
2.8	Tables and Graphs . . . . .	53
<b>3</b>	<b>The <math>k</math>-Step Bootstrap In State Space Models</b>	<b>58</b>
3.1	Introduction . . . . .	58
3.2	Summary of the $k$ -Step Bootstrap . . . . .	60
3.2.1	The Conventional Bootstrap . . . . .	60
3.2.2	The Approximate Bootstrap . . . . .	61
3.3	Constructing Test Statistics Using Kalman Filter Output . . . . .	66
3.4	Simulations . . . . .	69
3.4.1	Simulation I . . . . .	69
3.4.2	Simulation I: Results and Discussion . . . . .	69
3.4.3	Simulation II . . . . .	73
3.4.4	Simulation II: Results and Discussion . . . . .	74
3.5	Empirical Study . . . . .	77
3.5.1	Empirical Study: Results and Discussion . . . . .	79
3.6	Concluding Remarks . . . . .	81
3.7	Tables and Figures . . . . .	83
<b>4</b>	<b>Bootstrapping Multilevel Models: A State Space Approach</b>	<b>90</b>
4.1	Introduction . . . . .	90
4.2	The Bootstrap in Multilevel Models . . . . .	92
4.2.1	The Residual Bootstrap . . . . .	95
4.2.2	The Parametric Bootstrap . . . . .	97
4.3	State Space Models . . . . .	98



4.3.1	Kalman Filter Equations for Multilevel Models . . . . .	98
4.3.2	Parameter Estimation . . . . .	100
4.3.3	Casting Multilevel Models In State Space Form . . . . .	100
4.3.4	Example 1 . . . . .	102
4.4	Bootstrapping MLMs in State Space Form . . . . .	104
4.4.1	Residual Bootstrap . . . . .	104
4.4.2	Parametric Bootstrap . . . . .	105
4.5	Bootstrap Bias-Correction . . . . .	107
4.5.1	Simulation . . . . .	107
4.5.2	Simulation: Results and Discussion . . . . .	110
4.6	Empirical Study . . . . .	112
4.6.1	Empirical Study: Results and Discussion . . . . .	114
4.7	Concluding Remarks . . . . .	117
4.8	Tables and Figures . . . . .	119
4.8.1	Simulation: Tables . . . . .	119
4.8.2	Empirical Study: Tables . . . . .	121

**5 Conclusion** **125**

# List of Tables

2.1	Coverage Probabilities for Simulation I . . . . .	53
2.2	Length of Confidence Intervals . . . . .	55
2.3	Parameter Estimates and Running Times for Simulation II . .	56
2.4	Model Selections for Simulation III, $T = 20$ . . . . .	57
2.5	Model Selections for Simulation III, $T = 200$ . . . . .	57
3.1	Comparison of Asymptotic and Bootstrap Methods . . . . .	83
3.2	Average Absolute Differences Between $\hat{p}^*$ and $\tilde{p}^*$ . . . . .	84
3.3	Maximum Likelihood Estimates of Stochastic Parameter Regression Model (Equation 3.20), 1953/1 to 1965/2 . . . . .	84
3.4	Estimates of Stochastic Parameter Regression Model (Equation 3.20), 1953/1 to 1965/2 . . . . .	85
3.5	LR Statistics . . . . .	86
4.1	Estimates $J = 10, n = 100$ . . . . .	119
4.2	Estimates $J = 15, n = 100$ . . . . .	119
4.3	Estimates $J = 20, n = 100$ . . . . .	119
4.4	Estimates $J = 10, n = 100$ . . . . .	120
4.5	Estimates $J = 15, n = 100$ . . . . .	120

4.6	Estimates $J = 20, n = 100$ . . . . .	120
4.7	Estrone Assay Measurements (pg/mL) of a Single Blood Sample From Each of Five Postmenopausal Women . . . . .	121
4.8	Variance Estimates for One-Way Random Effects Model using VARCOMP . . . . .	122
4.9	Variance Estimates for One-Way Random Effects Model Using State Space Approach . . . . .	122
4.10	Bootstrap Estimates of the Variance of the Random Effects . .	122
4.11	Estimates of $\sigma_a^2$ . . . . .	123

# List of Figures

2.1	Coverage Probabilities for Simulation I . . . . .	54
3.1	LR Test Rejection Frequencies . . . . .	87
3.2	Wald Test Rejection Frequencies . . . . .	88
3.3	Score Test Rejection Frequencies . . . . .	89
4.1	Estimates of the Random Effects . . . . .	106
4.2	Empirical Cumulative Distribution Functions . . . . .	124

# Chapter 1

## Introduction

The bootstrap is a resampling technique that can be applied to a wide variety of problems including linear and nonlinear regressions, clustered data and time series models. However, its application to models in state space form has not been fully explored. A bootstrap for homoskedastic state space models was developed by [Stoffer and Wall \(1991\)](#) but this method is not valid in models with heteroskedastic errors. The goal of this research is to develop a bootstrap for state space models that would be viable under both homoskedasticity and heteroskedasticity. In this chapter we will first describe the general state space model and briefly discuss the current approach to bootstrapping models in the state space format. Next we will outline our objectives, explain the significance of the research and, finally, note its limitations.

Kalman filtering and state space models are highly effective tools for handling a wide range of data. Once a model has been cast in state space form, the Kalman filter can be applied and used for prediction, smoothing and likelihood evaluation. It effectively handles both stationary and non-

## Introduction

stationary series, including those with missing observations. Additionally, the state space framework accommodates fixed, systematically varying and stochastically varying parameters (Harvey, 1990). The linear, Gaussian state space model is given by:

$$\alpha_{t+1} = T_t \alpha_t + c_t + \eta_t, \quad t = 1, \dots, T \quad (1.1)$$

$$y_t = Z_t \alpha_t + d_t + \varepsilon_t, \quad t = 1, \dots, T \quad (1.2)$$

where

$$\begin{aligned} \varepsilon_t &\sim N(0, H_t), \\ \eta_t &\sim N(0, Q_t), \\ \alpha_0 &\sim N(a_0, P_0), \\ E(\varepsilon_t \varepsilon_j') &= H_t \delta_{t-j}, \\ E(\eta_t \eta_j') &= Q_t \delta_{t-j}, \\ E(\varepsilon_t \eta_j') &= 0, \\ E(\varepsilon_t \alpha_0') &= E(\eta_t \alpha_0') = 0. \end{aligned} \quad (1.3)$$

Equation (1.1) is called the state equation and equation (1.2) is known as the measurement equation.  $y_t$  is an observed, noisy time series,  $\alpha_t$  is the unobserved state vector,  $\alpha_0$  is the initial value of the state vector,  $Z_t$ ,  $d_t$ ,  $T_t$ ,  $c_t$ ,  $H_t$  and  $Q_t$  are time-varying system matrices and  $\eta_t$  and  $\varepsilon_t$  are the measurement and state equation disturbances, respectively.  $\delta_t$  is the Kronecker delta function, that is,  $\delta_{t-j} = 1$  if  $t = j$  and  $\delta_{t-j} = 0$  otherwise.

## Introduction

The model coefficients and covariance matrices are uniquely parameterised by a vector of hyperparameters,  $\psi$ , and the parameterisation is such that the model is completely identified ([Durbin and Koopman, 2012](#)):

$$Z_t = Z_t(\psi), T_t = T_t(\psi), H_t = H_t(\psi), Q_t = Q_t(\psi), c_t = c_t(\psi), d_t = d_t(\psi).$$

One of the primary functions of the Kalman filter is to estimate the hidden state variables,  $\alpha_t$ , using noisy measurements,  $y_t$ . Let  $a_{t|t-1}$  and  $P_{t|t-1}$  represent the estimate of the state and the covariance of the estimation error, respectively, based on information up to time  $t - 1$ , that is,

$$\begin{aligned} a_{t|t-1} &= E(\alpha_t | y_{t-1}, y_{t-2}, \dots, y_1) \\ P_{t|t-1} &= E \left[ (\alpha_t - a_{t|t-1})(\alpha_t - a_{t|t-1})' \right]. \end{aligned}$$

Similarly, let  $a_{t|t}$  and  $P_{t|t}$  be the estimate of the state and the covariance matrix of the estimation error given information up to time  $t$ , namely,

$$\begin{aligned} a_{t|t} &= E(\alpha_t | y_t, y_{t-1}, \dots, y_1) \\ P_{t|t} &= E \left[ (\alpha_t - a_{t|t})(\alpha_t - a_{t|t})' \right]. \end{aligned}$$

The discrete-time Kalman filter is given by:

$$\nu_t = y_t - Z_t a_{t|t-1} - d_t \tag{1.4}$$

$$F_t = Z_t P_{t|t-1} Z_t' + H_t \tag{1.5}$$

$$K_t = T_t P_{t|t-1} Z_t' F_t^{-1} \tag{1.6}$$

*Introduction*

$$\left. \begin{aligned} a_{t|t} &= a_{t|t-1} + P_{t|t-1} Z_t' F_t^{-1} (y_t - Z_t a_{t|t-1} - d_t) \\ P_{t|t} &= P_{t|t-1} - P_{t|t-1} Z_t' F_t^{-1} Z_t P_{t|t-1} \end{aligned} \right\} \text{updating equations} \quad (1.7)$$

$$(1.8)$$

$$\left. \begin{aligned} a_{t+1|t} &= T_t a_{t|t} + c_t \\ P_{t+1|t} &= T_t P_{t|t} T_t' + Q_t \end{aligned} \right\} \text{prediction equations.} \quad (1.9)$$

$$(1.10)$$

The one-step-ahead forecast errors,  $\nu_t$ , are sometimes referred to as prediction errors or innovations. They are normally distributed with mean zero and covariance,  $F_t$ . The matrix  $K_t$  is known as the Kalman gain. In the calculation of the updated state estimate,  $a_{t|t}$ , it is the weight given to the measurement,  $y_t$ , relative to the prior estimate of the state,  $a_{t|t-1}$ . Using the Kalman filter outputs, we can construct the prediction error decomposition form of the log-likelihood,  $L$ , for maximum likelihood estimation of the hyperparameters,  $\psi$ :

$$l_t = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log |F_t(\psi)| - \frac{1}{2} v_t'(\psi) F_t^{-1}(\psi) v_t(\psi) \quad (1.11)$$

$$L = \sum_{t=1}^T l_t$$

$$\mathcal{L} = \frac{1}{T} \sum_{t=1}^T l_t.$$

The likelihood in (1.11) is typically a highly nonlinear and complex function of the unknown parameters. It is usually maximised using quasi-Newton methods or the expectation maximisation algorithm. Under general conditions, the parameter estimates are consistent and asymptotically normal (Ljung and Caines, 1979). However, studies suggest that samples must be reasonably large



## *Introduction*

before the asymptotic results can be credibly applied (Dent and Min, 1978; Ansley and Newbold, 1980). Therefore, in the case of small to moderately sized data sets, appealing to asymptotics is no longer a practical option. In these cases, we find the bootstrap useful.

Bootstrapping is a technique for estimating the distribution of a test statistic or an estimator by resampling, either from the original sample data or from a model estimated from that data. The bootstrap can yield rejection probabilities of hypothesis tests and coverage probabilities of confidence intervals that are more accurate than the same measures calculated using asymptotic distribution theory. Additionally, bootstrapping allows us to make inferences in situations where finding the analytic approximation to a distribution is complicated or infeasible. A bootstrap for linear Gaussian state space models was developed by Stoffer and Wall (1991). In that study, the authors used the following mathematically equivalent version of the model described in (1.1) and (1.2), referred to as the innovations form representation:

$$\begin{aligned} a_{t+1|t} &= T_t a_{t|t-1} + c_t + K_t \nu_t & (1.12) \\ y_t &= Z_t a_{t|t-1} + d_t + \nu_t. \end{aligned}$$

In contrast to the standard model which has distinct errors,  $\eta_t$  and  $\varepsilon_t$ , in the state and measurement equations, respectively, the innovations form representation has the same error,  $\nu_t$ , in both equations. Reframing the standard model in this way is preferred since having only one error term simplifies the bootstrap methodology. The simulations performed by Stoffer and Wall (1991) demonstrated that, in small and moderately sized samples, the

## *Introduction*

bootstrap distributions of the parameter estimates more closely match their true distributions than do the distributions predicted by standard asymptotic theory. While ground-breaking, their technique is only valid for models in which the state and measurement equation errors are both homoskedastic. Since the seminal work by [Stoffer and Wall \(1991\)](#), to our knowledge, no further bootstraps have been developed for models in state space form.

The aim of this study is to develop a bootstrap that is applicable to heteroskedastic state space models. Currently, the most popular bootstrap for use in the presence of heteroskedasticity is the wild bootstrap. This was introduced into the literature by [Liu et al. \(1988\)](#) and is applicable in the case of heteroskedastic linear models with fixed regressors. [Kreiss \(1997\)](#) extended the work of [Liu et al. \(1988\)](#) and developed the fixed-design and recursive-design wild bootstraps for autoregressive models with heterogeneous errors. Another technique that can be used to address heteroskedasticity is the pairwise bootstrap of [Freedman \(1981\)](#). This approach stands out from the other methods because it resamples from the data, rather than from the residuals, and it does not condition on the regressors.

[Gonçalves and Kilian \(2004\)](#) generalised the work of [Kreiss \(1997\)](#) for use in autoregressive models with martingale difference errors and conditional heteroskedasticity of unknown form. They also showed that the fixed-design wild bootstrap, the recursive-design wild bootstrap and the pairwise bootstrap all perform better in small samples than the traditional large-sample approximations based on robust standard errors.

More recently, [Kline and Santos \(2012\)](#) introduced a bootstrap which is based on perturbing the scores of M-estimators. This score bootstrap

## *Introduction*

is applicable in the case of heteroskedastic, nonlinear models and is computationally efficient as it avoids recomputing the estimator in each bootstrap iteration. Our research builds on the methodology of [Kline and Santos \(2012\)](#), making it applicable to state space models. In contrast to the bootstrap of [Stoffer and Wall \(1991\)](#), our technique accommodates independent but not necessarily identically distributed error terms. Notably, the method will also work in homoskedastic models, providing an alternative to the methodology of [Stoffer and Wall \(1991\)](#). We examine the accuracy of our state space bootstrap by testing its performance in various models and by employing a variety of statistical tests.

While many models exhibit heteroskedasticity, it is not uncommon for researchers to assume homoskedasticity for convenience. However, homogeneous errors do not necessarily reflect reality. Many types of data exhibit heteroskedasticity and when samples are small to moderately-sized, bootstrap methods can be useful for their analysis. This kind of data is common in some areas of finance and economics. Examples include illiquid bonds with limited trading volume, microcap stocks with restricted liquidity and data availability, and data for newly issued securities such as initial public offerings. Moreover, economic research frequently grapples with a lack of historical indicators, especially when examining less developed countries or earlier time periods. This research will fill a gap in the literature by enabling practitioners to perform bootstrap inference in heteroskedastic state space models. Also, while a method for bootstrapping homoskedastic state space models already exists, the bootstrap developed herein will provide another, potentially faster, technique.

## *Introduction*

This study covers stationary state space models without missing observations, though we hope to extend it in the future.

The dissertation is divided into four distinct sections. Chapter 2 develops a bootstrap for heteroskedastic state space models. The performance of the new methodology is tested by calculating coverage probabilities for a selection of bootstrap confidence intervals. We show that they outperform the associated asymptotic confidence intervals. Using our bootstrap technique, an existing bootstrap model selection criterion is adjusted to make it applicable to state space models with heterogeneous error terms. We find that, in small samples, this bootstrap measure achieves better results than the standard Akaike Information Criterion. Additionally, a bootstrap procedure for use in homoskedastic state space models is developed, and proves to be faster than the approach of [Stoffer and Wall \(1991\)](#).

In Chapter 3, we review the approximate bootstrap of [Davidson and MacKinnon \(1999\)](#) and investigate its effectiveness in state space models for the purpose of hypothesis testing. The  $k$ -step method was developed to achieve the accuracy of a full bootstrap while reducing computational costs. Our findings indicate that their method is well-suited for state space models and that working in the state space framework has some advantages; Kalman filter output facilitates the construction of the log-likelihood function for estimation and enables the creation of certain chi-squared test statistics.

In Chapter 4, we emphasise the advantages of using the state space representation for estimating multilevel models. We demonstrate how the Kalman filter output can aid in performing residual and parametric bootstraps in clustered data. Additionally, we illustrate how bootstrapping can be

## *Introduction*

employed to mitigate the bias in the maximum likelihood estimates of the components of the level-2 covariance matrices of the vector of random effects.

Chapter 5 summarises and concludes.

# Chapter 2

## The State Space Model Bootstrap

### 2.1 Introduction

[Kline and Santos \(2012\)](#) proposed an alternative to the wild bootstrap of [Liu et al. \(1988\)](#). Their score bootstrap was proven to be less costly to implement than the conventional wild bootstrap because the procedure avoids re-computation of the estimator in each bootstrap replication. Our state space model bootstrap exploits the fact that the regression model considered in their work can be cast in state space form for estimation, at which point a similar gradient-based bootstrap can be applied, using the output of the Kalman filter. One benefit of adopting the state space formulation is that it allows for the inclusion of time-varying parameters. These stochastic parameter regression models are widely used in the literature and it is therefore worthwhile to extend the technique of [Kline and Santos \(2012\)](#) to include them. This research attempts to do so.

The coefficients in linear regression models can be estimated in one step

### *The State Space Model Bootstrap*

by means of ordinary least squares (OLS). However, in most cases, model estimation entails a series of steps. Normally, when a nonlinear model is bootstrapped, a multi-step optimisation procedure has to be undertaken in every bootstrap iteration, which can be time-consuming. In the case of  $B$  bootstrap replications,  $B + 1$  nonlinear optimisations must be performed. [Davidson and MacKinnon \(1999\)](#) developed a  $k$ -step process that reduces the computational cost of bootstrap testing in nonlinear models. The authors noted that while nonlinear estimation involves several iterations of a numerical algorithm, it is often sufficient to replace them with a small number of Newton or quasi-Newton steps. With their methodology, when performing  $B$  bootstrap replications, it is only required that we carry out one nonlinear optimisation, rather than  $B + 1$ ; the additional  $B$  estimations can be replaced with  $kB$  Newton steps where  $k$  is a small integer. In general, implementing the  $k$ -step procedure is less costly than performing  $B + 1$  nonlinear estimations. The score bootstrap of [Kline and Santos \(2012\)](#) can be interpreted as a special case of this  $k$ -step method when  $k = 1$ .

Inference in state space models is usually made tractable by appealing to asymptotic theory. Under certain conditions, the parameter estimates obtained by maximum likelihood (ML) are consistent and asymptotically normal ([Durbin and Koopman, 2012](#)). However, as asymptotic approximation to the distributions of ML estimators and related test statistics tend to perform poorly in finite samples, researchers often opt for bootstrap techniques. [Stoffer and Wall \(1991\)](#) applied the Monte Carlo bootstrap of [Efron \(1979\)](#) to homoskedastic state space models. They showed that, in small and medium-sized samples of 50 and 100 data points, respectively, their methodology

### *The State Space Model Bootstrap*

produces results which are superior to those achieved by conventional approaches. Specifically, [Stoffer and Wall \(1991\)](#) used Monte Carlo methods to simulate the true small-sample distributions of the ML estimators of the state space model parameters, focusing primarily on their means and standard deviations. They determined that the corresponding bootstrap estimates for these measures were closer to the true values than were the same quantities acquired via standard large-sample theory.

Widespread application of the bootstrap of [Stoffer and Wall \(1991\)](#) has partially been hindered by the fact that optimisation of the likelihood function must be carried out in every bootstrap iteration, rendering this practice computationally expensive, even with advancements in processor speeds. While our research in this chapter is primarily concerned with heteroskedastic models, we also derive a version of the state space model (SSM) bootstrap that is valid in homoskedastic models and which is less computationally taxing than the method introduced by [Stoffer and Wall \(1991\)](#).

The Akaike information criterion (AIC) is one of the most frequently used tools for model selection. However, when sample sizes are small, the AIC will overfit and favour models with a greater number of parameters. To account for this, the AICc was developed by [Sugiura \(1978\)](#). This measure imposes an extra penalty term for the number of parameters. Since the AICc was not formulated for use in the state space environment, [Cavanaugh and Shumway \(1997\)](#) devised a bootstrap-corrected version of the AIC, termed the AICb, specifically for small-sample state space model selection. They considered homoskedastic models and used the bootstrap procedure of [Stoffer and Wall \(1991\)](#) to calculate their criterion. Notably, the authors generated realisations



from an auto-regressive (AR) model and, for each realisation, AR models of varying lag lengths were fitted to the data. The AICb was then used to select the best candidate model. This experiment was repeated using the AIC and other information criteria. In all of the scenarios considered, the AICb outperformed the other measures. To examine the performance of the SSM bootstrap in the context of model selection, our investigation proceeds along the lines of [Cavanaugh and Shumway \(1997\)](#). We look at the effectiveness of a gradient-based, bootstrap-corrected variant of the AIC, deemed the AICg. Using this version of the bootstrap should allow us to extend the method of [Cavanaugh and Shumway \(1997\)](#) to include heteroskedastic state space models.

The study proceeds as follows: Section 2 reviews the score bootstrap of [Kline and Santos \(2012\)](#); Section 3 describes the bootstrap technique of [Stoffer and Wall \(1991\)](#); Section 4 introduces the methodology for the SSM bootstrap while Section 5 addresses model selection; simulation studies are provided in Section 6 and Section 7 concludes.

## 2.2 Review of the Score Bootstrap

[Kline and Santos \(2012\)](#) innovated a score based approach to the traditional wild bootstrap. The authors examined the linear model,

$$y_t = X_t' \beta + \varepsilon_t, \quad (2.1)$$

where  $\{y_t, X_t\}_{t=1}^T$  is an independent and identically distributed (*iid*) sequence

## The State Space Model Bootstrap

of random variables with  $y_t \in \mathbb{R}$  and  $X_t \in \mathbb{R}^m$ . They noted, from standard OLS algebra, that

$$\sqrt{T}(\hat{\beta} - \beta) = H_T^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T X_t \varepsilon_t. \quad (2.2)$$

Above,  $H_T = \frac{1}{T} \sum_{t=1}^T X_t X_t'$ , and  $X_t \varepsilon_t$  is the  $t^{\text{th}}$  score contribution. The bootstrap equivalent of (2.2) is

$$\sqrt{T}(\beta_j^* - \hat{\beta}) = H_T^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T X_t \varepsilon_t^*, \quad (2.3)$$

where  $\{\varepsilon_t^*\}_{t=1}^T$  are the bootstrap errors and  $\beta_j^*$  is the estimate from the  $j^{\text{th}}$  bootstrap sample. The authors examined (2.2) and determined that since the error terms only affect the limiting distribution of the OLS estimator through the score contributions,  $X_t \varepsilon_t$ , a numerically equivalent way of executing the wild bootstrap and estimating the distribution of  $\hat{\beta}$  would be to perturb the score contributions directly and avoid having to continually re-estimate  $\beta_j^*$ . For clarity, the algorithms for the wild and score bootstraps are outlined below.

### 2.2.1 The Wild Bootstrap

The steps for the wild bootstrap of [Liu et al. \(1988\)](#) are as follows:

1. Estimate  $y_t = X_t' \beta + \varepsilon_t$  by OLS to obtain  $\hat{\beta}$ .
2. Find the residuals  $e_t = y_t - X_t' \hat{\beta}$ ,  $t = 1, \dots, T$ .
3. Generate bootstrap errors of the form  $\varepsilon_t^* = W_t e_t$ , where  $\{W_t\}_{t=1}^T$  is

### *The State Space Model Bootstrap*

a randomly generated sequence that is independent of  $\{y_t, X_t\}_{t=1}^T$  and satisfies the conditions  $E(W_t) = 0$  and  $E(W_t^2) = 1$ .

4. Generate bootstrap replicates of the dependent variable  $y_t^* = X_t' \hat{\beta} + \varepsilon_t^*$ ,  $t = 1, \dots, T$ .
5. Perform OLS on the above to find  $\beta^*$ .
6. Monte Carlo approximation to the conditional distribution of  $\beta^*$  is obtained by repeating steps 3 to 5  $B$  times.
7. Use the distribution of  $\sqrt{T}(\beta^* - \hat{\beta})$  conditional on  $\{y_t, X_t\}_{t=1}^T$  to estimate the distribution of  $\sqrt{T}(\hat{\beta} - \beta)$ .

A number of distributions satisfy the requirements for generating the wild weights,  $W_t$ . Among these, two of the most widely used are,

- (a) The Rademacher distribution:

$$W_t = \begin{cases} -1, & \text{with probability } \frac{1}{2}. \\ +1, & \text{with probability } \frac{1}{2}. \end{cases}$$

Here,  $E[W_t] = 0$ ,  $E[W_t^2] = 1$ ,  $E[W_t^3] = 0$  and  $E[W_t^4] = 1$ . Clearly, when using Rademacher weights, if the distribution of the residuals is symmetric, there will be agreement between the first four moments of the distribution of the bootstrap errors and the distribution of the residuals.

(b) Mammen's distribution:

$$W_t = \begin{cases} \frac{\sqrt{5}+1}{2}, & \text{with probability } 1 - \frac{\sqrt{5}+1}{2\sqrt{5}}. \\ \frac{1-\sqrt{5}}{2}, & \text{with probability } \frac{\sqrt{5}+1}{2\sqrt{5}}. \end{cases}$$

With these weights,  $E[W_t] = 0$ ,  $E[W_t^2] = 1$ ,  $E[W_t^3] = 1$  and  $E[W_t^4] = 2$ .

In this case, since  $E[W_t^3] = 1$ , any skewness in the distribution of the residuals is captured in the distribution of the bootstrap errors.

In our research, we opt primarily for Rademacher weights as previous studies have demonstrated that they outperform Mammen's skew-correcting weights, even when the distribution of the residuals is skewed ([Davidson and Flachaire \(2008\)](#), [Davidson et al. \(2007\)](#)).

Apart from the two-point distributions mentioned here, the standard normal distribution also provides a viable alternative from which to draw wild weights. Nonetheless, due to computational limitations, we did not investigate additional distributions beyond the Rademacher distribution.

### **2.2.2 The Score Bootstrap**

The methodology for the score bootstrap of [Kline and Santos \(2012\)](#) is summarised below.

1. Estimate  $y_t = X_t'\beta + \varepsilon_t$  by OLS to acquire  $\hat{\beta}$  and the residuals  $e_t = (y_t - X_t'\hat{\beta})$ ,  $t = 1 \dots T$ .
2. Generate the fitted score contributions,  $X_t e_t$ ,  $t = 1 \dots T$ .

*The State Space Model Bootstrap*

3. Use random weights  $W_t$ , with previously described properties, to generate the perturbed score contributions  $\left\{ X_t \left( y_t - X_t' \hat{\beta} \right) W_t \right\}_{t=1}^T$ .
4. Multiply the perturbed score by  $H_T^{-1}$  to obtain

$$\sqrt{T} \left( \beta^* - \hat{\beta} \right) = H_T^{-1} \frac{1}{\sqrt{T}} \sum_t \left\{ X_t \left( y_t - X_t' \hat{\beta} \right) W_t \right\}. \quad (2.4)$$

5. Repeat steps 3 to 4  $B$  times.
6. Use the distribution of  $\sqrt{T} \left( \beta^* - \hat{\beta} \right)$  conditional on  $\{y_t, X_t\}_{t=1}^T$  to estimate the distribution of  $\sqrt{T} \left( \hat{\beta} - \beta \right)$ .

In equation (2.2), when  $\hat{\beta}$  is regarded as the maximum likelihood estimator,  $H_T$  is the Hessian of the log-likelihood<sup>1</sup> and  $\frac{1}{T} \sum_{t=1}^T X_t \varepsilon_t$  is the gradient calculated at the true parameter  $\beta$ .

Kline and Santos (2012) applied this technique to non-linear models. In contrast to the linear regression model where the Hessian matrix is always  $\frac{1}{T} \sum_t X'X$  in every bootstrap replication, in non-linear estimation, the Hessian will be different in each bootstrap round. To address this, the authors fix the Hessian at  $H_T$ , the matrix of second derivatives evaluated at the ML estimate,  $\hat{\beta}$ , and proceed with the algorithm as described. For this reason, their technique bears some similarity to the fixed-design wild bootstrap of Kreiss (1997) which holds all regressors constant in every bootstrap iteration.

The score bootstrap was developed for heteroskedastic models. However, in the presence of *iid* errors, a residual-type bootstrap can also be used with the score-based approach. While this was not addressed by Kline and Santos

---

<sup>1</sup> $H_T = \frac{1}{T} \sum_{t=1}^T \frac{\delta^2 l_t}{\delta \beta_k \delta \beta_j}$ , where  $l_t = \log p(y_t | y_{t-1}, y_{t-2}, \dots, y_1)$

### *The State Space Model Bootstrap*

(2012), a similar gradient-based technique can be performed in homoskedastic models. First, consider the standard linear regression model,

$$y_t = X_t' \beta + \varepsilon_t, \quad t = 1, \dots, T. \quad (2.5)$$

The algorithm for the standard residual bootstrap is,

1. Estimate (2.5) by OLS and find  $\hat{\beta}$ .
2. Calculate  $e_t = y_t - X_t' \hat{\beta}$ ,  $t = 1, \dots, T$ .
3. Find  $\{\varepsilon_t^*\}_{t=1}^T$  by resampling  $T$  times from  $\{e_t\}_{t=1}^T$ , with replacement.
4. Generate  $y_t^* = X_t' \hat{\beta} + \varepsilon_t^*$ ,  $t = 1, \dots, T$ .
5. Regress  $y_t^*$  on  $X_t$  and estimate  $\beta^*$ .
6. Repeat steps 3 to 5  $B$  times to obtain  $\{\beta_j^*\}_{j=1}^B$ .

In the same way that [Kline and Santos \(2012\)](#) calculated the perturbed score using wild bootstrap weights, if the errors are *iid*, we can also work directly with the score contributions using resampled residuals. That algorithm would proceed in the following manner:

1. Estimate (2.5) by OLS to find  $\hat{\beta}$  and the residuals,  $\{e_t\}_{t=1}^T$ .
2. Resample  $\{e_t\}_{t=1}^T$   $T$  times with replacement to generate the bootstrap errors,  $\{\varepsilon_t^*\}_{t=1}^T$ .
3. Generate the new score contributions,  $\{X_t \varepsilon_t^*\}_{t=1}^T$ .

*The State Space Model Bootstrap*

4. Multiply the perturbed score by  $H_T^{-1}$  to obtain

$$\sqrt{T}(\beta^* - \hat{\beta}) = H_T^{-1} \frac{1}{\sqrt{T}} \sum_t \{X_t \varepsilon_t^*\}. \quad (2.6)$$

5. Repeat steps 2 to 4  $B$  times to generate  $\{\beta_j^*; 1 \leq j \leq B\}$ .
6. Use the distribution of  $\sqrt{T}(\beta^* - \hat{\beta})$  conditional on  $\{y_t, X_t\}_{t=1}^T$  to estimate the distribution of  $\sqrt{T}(\hat{\beta} - \beta)$ .

Some care must be taken when using the gradient-based approach with the residual-type bootstrap. In the heteroskedastic case considered by [Kline and Santos \(2012\)](#), wild weights were applied directly to the score contributions,  $\{X_t e_t\}_{t=1}^T$ . However, in the homoskedastic case, we do not resample the score contributions. We first resample the residuals,  $\{e_t\}_{t=1}^T$ , with replacement, to acquire  $\{\varepsilon_t^*\}_{t=1}^T$ . Then, we construct  $\{X_t \varepsilon_t^*\}_{t=1}^T$ . If we resampled the  $X_t e_t$ 's, not only would the  $X_t$ 's be shuffled along with the  $e_t$ 's, but it is unlikely that all of the  $X_t$ 's would appear in the perturbed score since some of them might not be chosen in the resampling process.

Many non-linear models have extremely complex score functions. Another, and many times simpler, way of forming the perturbed score is via the bootstrap data generating process. Assume we have a model,

$$y_t = f(X_t, \beta, \varepsilon_t). \quad (2.7)$$

We can estimate this model to obtain  $\hat{\beta}$  and the residuals  $\{e_t\}_{t=1}^T$ . Now,

### *The State Space Model Bootstrap*

calculate  $\{\varepsilon_t^*\}_{t=1}^T$  and generate the bootstrap model

$$y_t^* = f(X_t, \hat{\beta}, \varepsilon_t^*). \quad (2.8)$$

The perturbed score can be found numerically from the log-likelihood of  $y_t^*$  by using finite differences.

## **2.3 The Bootstrap of Stoffer and Wall**

Consider the general state space model introduced in (1.1) and (1.2) but with homoskedastic errors, that is,

$$\begin{aligned} \alpha_{t+1} &= T_t \alpha_t + c_t + \eta_t, & t = 1, \dots, T \\ y_t &= Z_t \alpha_t + d_t + \varepsilon_t, & t = 1, \dots, T, \end{aligned}$$

where  $\eta_t \sim N(0, Q)$  and  $\varepsilon_t \sim N(0, H)$ . Let  $\psi$  be the parameter vector to be estimated. The algorithm for the bootstrap of [Stoffer and Wall \(1991\)](#) can be summarised as follows:

1. Maximise the log-likelihood of the state space model to obtain  $\hat{\psi}$ . Estimation can be accomplished using a variety of methods such as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) or Davidon-Fletcher-Powell algorithms (DFP).
2. Run the Kalman filter with  $\psi = \hat{\psi}$ . The output of the filter includes the estimate of the state given information up to time  $t - 1$ ,  $a_{t|t-1}$ , the Kalman gain,  $K_t$ , the prediction errors,  $\nu_t$ , and the prediction error



covariance,  $F_t$ . Recall the innovations form representation, introduced in Chapter 1,<sup>2</sup>

$$\begin{aligned} a_{t+1|t} &= T_t a_{t|t-1} + c_t + K_t \nu_t \\ y_t &= Z_t a_{t|t-1} + d_t + \nu_t. \end{aligned} \tag{2.9}$$

3. Generate bootstrap innovations,  $\{\nu^*\}_{t=1}^T$ , by drawing with replacement from the innovations,  $\{\nu_t\}_{t=1}^T$ .
4. Using the innovations form of the model, construct the bootstrap data set,  $\{y_t^*\}_{t=1}^T$ . Note that the parameter vector remains fixed at  $\hat{\psi}$ . In the bootstrap data generating process, the estimate of the state is calculated recursively and is initialised by setting  $a_{1|0}^* = a_{1|0}$ .

$$\begin{aligned} a_{t+1|t}^* &= T_t(\hat{\psi}) a_{t|t-1}^* + c_t(\hat{\psi}) + K_t v_t^* \\ y_t^* &= Z_t(\hat{\psi}) a_{t|t-1}^* + d_t(\hat{\psi}) + v_t^* \end{aligned} \tag{2.10}$$

5. Estimate (2.10) to find  $\psi^*$ .
6. Repeat steps 3 through 5  $B$  times, generating  $\{\psi_j^*; 1 \leq j \leq B\}$ . The Kalman filter should be initialised with the same values in each bootstrap iteration.
7. Use the conditional distribution of  $\sqrt{T}(\psi^* - \hat{\psi})$  to estimate the distribution of  $\sqrt{T}(\hat{\psi} - \psi_0)$ .

---

<sup>2</sup>See Anderson and Moore (2012) for the derivation of the innovations form representation from Kalman filter output.

This process requires that (2.10) be estimated  $B$  times.

## 2.4 Deriving the State Space Model Bootstrap

We consider a model similar to the one used by [Stoffer and Wall \(1991\)](#) but we allow for heteroskedasticity in the state and measurement equation errors. In particular, we look at

$$\alpha_{t+1} = T_t \alpha_t + c_t + \eta_t, \quad t = 1, \dots, T \quad (2.11)$$

$$y_t = Z_t \alpha_t + d_t + \varepsilon_t, \quad t = 1, \dots, T, \quad (2.12)$$

where  $\eta_t \sim N(0, Q_t)$  and  $\varepsilon_t \sim N(0, H_t)$ . Let  $\psi$  be the vector of parameters to be estimated.

Recall the prediction error decomposition form of the log-likelihood that is constructed using the output from the Kalman filter:

$$l_t = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log |F_t(\psi)| - \frac{1}{2} v_t'(\psi) F_t^{-1}(\psi) v_t(\psi). \quad (2.13)$$

We can define the average log-likelihood as,

$$\mathcal{L}(y_t; \psi) = \frac{1}{T} \sum_{t=1}^T l_t(y_t; \psi). \quad (2.14)$$

If (2.14) is differentiable and  $\hat{\psi}$  is an interior point, then the first order condition for  $\mathcal{L}(\psi)$  will be satisfied:

$$\frac{1}{T} \sum_{t=1}^T \nabla_{\psi} l_t(y_t; \hat{\psi}) = 0. \quad (2.15)$$

*The State Space Model Bootstrap*

Above,  $\nabla_{\psi} l_t(y_t; \hat{\psi})$  is the derivative of  $l_t$  with respect to  $\psi$ , evaluated at  $\hat{\psi}$ . Assuming that (2.14) is twice continuously differentiable, we can perform the following analysis:

Applying a Taylor expansion to (2.15) around the true parameter value  $\psi_0$  gives<sup>3</sup>,

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \nabla_{\psi} l_t(y_t; \hat{\psi}) &= \frac{1}{T} \sum_{t=1}^T \nabla_{\psi} l_t(y_t; \psi_0) \\ &\quad + \frac{1}{T} \left[ \sum_{t=1}^T \nabla_{\psi\psi} l_t(y_t; \psi_0) \right] (\hat{\psi} - \psi_0) + R_1 = 0, \end{aligned} \quad (2.16)$$

where  $\frac{1}{T} \sum_{t=1}^T \nabla_{\psi\psi} l_t(y_t; \psi)$  is the Hessian matrix of second derivatives of  $\mathcal{L}(y_t; \psi)$  with respect to  $\psi$  and  $R_1$  is the Taylor remainder term. In linear regression models such as (2.1), the third and higher derivatives of the log-likelihood with respect to  $\psi$  are zero. So, from (2.16),

$$\sqrt{T}(\hat{\psi} - \psi_0) = - \left[ \frac{1}{T} \sum_{t=1}^T \nabla_{\psi\psi} l_t(y_t; \psi_0) \right]^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T \nabla_{\psi} l_t(y_t; \psi_0). \quad (2.17)$$

This product of the inverse of the Hessian times the score is comparable to (2.2), the equation for the distribution of the OLS estimator. Along the lines of [Kline and Santos \(2012\)](#), we can construct an equation analogous to (2.17) in which we use perturbed score contributions,  $\{\nabla_{\psi} l_t(y_t^*; \hat{\psi})\}_{t=1}^T$ , to estimate the

---

<sup>3</sup>When  $\hat{\psi}$  is a vector, the Taylor expansion applies to each individual element in the partial derivatives.

conditional distribution of  $\psi^*$  without carrying out any further optimisation.

$$\sqrt{T}(\psi^* - \hat{\psi}) = - \left[ \underbrace{\frac{1}{T} \sum_{t=1}^T \nabla_{\psi\psi} l_t(y_t; \hat{\psi})}_{\text{Hessian}} \right]^{-1} \underbrace{\frac{1}{\sqrt{T}} \sum_{t=1}^T \nabla_{\psi} l_t(y_t^*; \hat{\psi})}_{\text{Score}} \quad (2.18)$$

Full details of the algorithm appear in the next section.

Observe from (2.18) that the bootstrap estimate,  $\psi^*$ , can be recovered from the maximum likelihood estimate,  $\hat{\psi}$ , as shown below:

$$\psi^* = \hat{\psi} - \left[ \sum_{t=1}^T \nabla_{\psi\psi} l_t(y_t; \hat{\psi}) \right]^{-1} \sum_{t=1}^T \nabla_{\psi} l_t^*(y_t; \hat{\psi}). \quad (2.19)$$

Equation (2.19) is one iteration of the familiar Newton-Raphson method. Since the first order conditions are linear in the parameters for models like (2.1), the algorithm will converge in just one Newton step. In non-linear estimation however, the remainder term in (2.16) only vanishes asymptotically so the relationship in (2.19) just holds approximately:

$$\psi^* \approx \hat{\psi} - \left[ \sum_{t=1}^T \nabla_{\psi\psi} l_t(y_t; \hat{\psi}) \right]^{-1} \sum_{t=1}^T \nabla_{\psi} l_t(y_t^*; \hat{\psi}) \quad (2.20)$$

### 2.4.1 The Wild State Space Model Bootstrap

When the error terms of the state space model exhibit heteroskedasticity, the bootstrap can be carried out in this manner:

1. Maximise the log-likelihood in (2.14) to find  $\hat{\psi} = \arg \max_{\psi} \mathcal{L}(y_t; \psi)$  using the preferred Newton procedure. The optimisation routine will

*The State Space Model Bootstrap*

also provide an estimate of the Hessian,  $H_T$ .

2. Run the Kalman filter with  $\psi = \hat{\psi}$  to obtain the innovations,  $\nu_t$ , and their covariances,  $F_t$ .
3. Using weights,  $W_t$ , with previously described properties, perturb the innovations,

$$\nu_t^* = \nu_t * W_t, \quad \forall t. \quad (2.21)$$

4. Create the bootstrap data set,  $\{y_t^*\}_{t=1}^T$ , according to step 4 of the [Stoffer and Wall \(1991\)](#) algorithm given in section 2.3.
5. Construct the bootstrap log-likelihood,  $l_t(y_t^*; \hat{\psi})$ ,

$$l_t(y_t^*; \hat{\psi}) = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log |F_t| - \frac{1}{2} \nu_t^* F_t^{-1} \nu_t^*. \quad (2.22)$$

6. Calculate the bootstrap score contributions,  $\nabla_{\psi} l_t(y_t^*; \hat{\psi})$ .
7. Find  $\psi^*$  using

$$\sqrt{T}(\psi^* - \hat{\psi}) = -H_T^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T \nabla_{\psi} l_t(y_t^*; \hat{\psi}). \quad (2.23)$$

8. Repeat steps 3 to 7  $B$  times to generate  $\{\psi_j^*; 1 \leq j \leq B\}$ .
9. Use the empirical distribution of  $\sqrt{T}(\psi_j^* - \hat{\psi})$  to estimate the distribution of  $\sqrt{T}(\hat{\psi} - \psi_0)$ .

### 2.4.2 The Residual State Space Model Bootstrap

If the error terms of the state space model are homoskedastic, the bootstrap is executed as shown below:

1. Maximise the log-likelihood in (2.14) to find  $\hat{\psi} = \arg \max_{\psi} \mathcal{L}(y_t; \psi)$  and the Hessian,  $H_T$ .
2. Run the Kalman filter with  $\psi = \hat{\psi}$  to obtain  $\nu_t$  and  $F_t$ ,  $t = 1, \dots, T$ .
3. Standardise the innovations,  $\nu_t$ , so that the first two moments of all innovations are the same:

$$\nu_t^s = F_t^{-\frac{1}{2}} \nu_t. \quad (2.24)$$

4. Resample with replacement from the standardised innovations,  $\nu_t^s$ , to generate  $\nu_t^r$ .
5. Reinflate  $\nu_t^r$  to create the bootstrap innovations,  $\nu_t^*$ ,

$$\nu_t^* = F_t^{\frac{1}{2}} * \nu_t^r. \quad (2.25)$$

6. Generate the bootstrap data set,  $\{y_t^*\}_{t=1}^T$ , as outlined in step 4 of the [Stoffer and Wall \(1991\)](#) algorithm presented in section 2.3.
7. Construct the bootstrap log-likelihood,  $l_t(y_t^*; \hat{\psi})$ ,

$$l_t(y_t^*; \hat{\psi}) = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log |F_t| - \frac{1}{2} \nu_t^* F_t^{-1} \nu_t^*. \quad (2.26)$$

*The State Space Model Bootstrap*

8. Compute the bootstrap score contributions,  $\nabla_{\psi} l_t(y_t^*; \hat{\psi})$ .
9. Calculate  $\psi^*$  from

$$\sqrt{T}(\psi^* - \hat{\psi}) = -H_T^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T \nabla_{\psi} l_t(y_t^*; \hat{\psi}). \quad (2.27)$$

10. Repeat steps 4 to 9  $B$  times to generate  $\{\psi_j^*; 1 \leq j \leq B\}$ .
11. Use the empirical distribution of  $\sqrt{T}(\psi_j^* - \hat{\psi})$  to approximate the distribution of  $\sqrt{T}(\hat{\psi} - \psi_0)$ .

In both algorithms, we can calculate the bootstrap score from the log-likelihood either numerically or analytically. The  $i^{th}$  element of the score vector for the prediction error decomposition form of the average log-likelihood in (2.14) is shown below.<sup>4</sup>

$$\begin{aligned} \frac{\partial \log \mathcal{L}}{\partial \psi_i} = & -\frac{1}{2T} \sum_t \left\{ \text{tr} \left[ \left[ F_t^{-1} \frac{\partial F_t}{\partial \psi_i} \right] (I - F_t^{-1} \nu_t \nu_t') \right] \right. \\ & \left. + 2 \frac{\partial \nu_t'}{\partial \psi_i} F_t^{-1} \nu_t \right\}, \quad i = 1, \dots, k. \end{aligned} \quad (2.28)$$

Like the score bootstrap of [Kline and Santos \(2012\)](#), the wild state space model (WSSM) and residual state space model (RSSM) bootstrap approaches are different from more conventional bootstraps as they require that  $\hat{\psi}$  be evaluated only once. This significantly increases the computational efficiency of the bootstrapping process.

The methodologies for both SSM bootstraps were developed using Gaussian errors. However, the techniques are robust to other types of errors.

---

<sup>4</sup>The recursions to calculate the exact score are given in [Harvey \(1990\)](#).

## *The State Space Model Bootstrap*

If the distribution of the error terms is known, then one can proceed with maximum likelihood estimation and the bootstrap procedures remain valid. The challenge arises when the model may be misspecified and it is unknown, *a priori*, whether or not the quasi-maximum likelihood estimates of the parameters of interest will be consistent. In these instances, we can appeal to the 2-step methodology of [White \(1982\)](#). In particular, we first employ the Information Matrix test to assess whether the model is misspecified. If it is not, we can confidently use maximum likelihood techniques and bootstrap inference as usual. On the other hand, if the model is determined to be misspecified, we conduct either a Hausman Test or a Gradient Test to evaluate whether this misspecification will lead to inconsistent parameter estimates. If the parameter estimates are deemed consistent, we can proceed as planned. However, if they are found to be inconsistent, the bootstrap methodologies will be invalid.<sup>5</sup>

## 2.5 Model Selection

The Akaike information criterion (AIC) is a well-known statistical measure for evaluating which model fits best for a given data set. The AIC function is,

$$\text{AIC} = -2\log\mathbb{L}(\hat{\psi}|Y_T) + 2k, \quad (2.29)$$

where  $k$  is the number of parameters to be estimated,  $\hat{\psi}$  is the ML estimate of the true parameter  $\psi$ ,  $\mathbb{L}$  is the likelihood function and  $Y_T$  represents the

---

<sup>5</sup>Details of the Information Matrix, Hausman and Gradient tests are given in [White \(1982\)](#).



### *The State Space Model Bootstrap*

observed data up to time  $T$ , that is,  $Y_T = (y_1, \dots, y_T)$ . One drawback of the AIC is its tendency to favour high-dimensional models when the sample size is small. To address this shortcoming, [Cavanaugh and Shumway \(1997\)](#) developed the AICb as a model selection criterion for use in small samples:

$$\text{AICb} = -2\log\mathbb{L}(\hat{\psi}|Y_T) + 2 \left\{ \frac{1}{B} \sum_{j=1}^B -2\log \frac{\mathbb{L}(\psi^*(j)|Y_T)}{\mathbb{L}(\hat{\psi}|Y_T)} \right\}. \quad (2.30)$$

In (2.30),  $\psi^*$  is the bootstrap estimate of  $\psi$  and  $B$  is the number of bootstrap replications. They considered univariate, stationary, autoregressive (AR) models with and without observation noise. The models with noise were labelled as ARN models. Specifically, the authors examined the following processes of order  $p$ , i)

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + \eta_t, \quad \eta_t \sim iid(0, \sigma_Q^2), \quad (2.31)$$

and ii)

$$\begin{aligned} y_t &= z_t + \varepsilon_t, & \varepsilon_t &\sim iid(0, \sigma_R^2) \\ z_t &= \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + \epsilon_t, & \epsilon_t &\sim iid(0, \sigma_Q^2). \end{aligned} \quad (2.32)$$

They cast the ARN model in (2.32) in state space form as shown below:

$$y_t = (1, 0, \dots, 0) \begin{pmatrix} z_t \\ z_{t-1} \\ \vdots \\ z_{t-p+1} \end{pmatrix} + \varepsilon_t \quad (2.33)$$

$$\begin{pmatrix} z_t \\ z_{t-1} \\ \vdots \\ z_{t-p+1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} z_{t-1} \\ z_{t-2} \\ \vdots \\ z_{t-p} \end{pmatrix} + \begin{pmatrix} \eta_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (2.34)$$

The state space representation for the AR( $p$ ) model is similar; the state equation (2.34) is the same as in the ARN( $p$ ) case but there is no error term,  $\varepsilon_t$ , in the observation equation (2.33).

The parameter vectors for the AR( $p$ ) and ARN( $p$ ) are given by  $\psi = (\phi_1, \dots, \phi_p, \sigma_Q^2)$  and  $\psi = (\phi_1, \dots, \phi_p, \sigma_R^2, \sigma_Q^2)$ , respectively. In order to evaluate  $\psi^*$  the authors made use of the bootstrap of [Stoffer and Wall \(1991\)](#), which is applicable in the case of homoskedastic models. Of course, if either of the errors in a state space model is heteroskedastic, this bootstrap cannot be applied. Under this scenario, we propose the use of the WSSM bootstrap to calculate  $\psi^*$  instead. The formula for the new information criterion based on the WSSM bootstrap is the same as for the AICb but now  $\psi^*$  is calculated using our gradient-based approach. This new measure is titled the AICg.

## 2.6 Simulations

### 2.6.1 Simulation I

The workhorse linear regression model (2.1) lends itself easily to the application of the score bootstrap. However, models with time-varying parameters have also been extensively studied in the literature, for example,

$$y_t = X'_{0,t}\beta_0 + X'_{1,t}\beta_{1,t} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2). \quad (2.35)$$

Such models can be estimated by rolling regressions, that is, by computing OLS estimates for sliding windows of a specified width, but this would greatly complicate the application of the score bootstrap. Fortunately, these models can be conveniently estimated and analysed using state space methods. For instance, (2.35) may be expressed as,

$$\begin{aligned} y_t &= X'_{0,t}\beta_0 + X'_{1,t}\beta_{1,t} + \varepsilon_t, & \varepsilon_t &\sim N(0, \sigma_\varepsilon^2) \\ \beta_{1,t+1} &= \gamma\beta_{1,t} + \eta_t, & \eta_t &\sim N(0, \sigma_\eta^2), \end{aligned} \quad (2.36)$$

where the time-varying parameter evolves according to a Markov process. By way of illustration, in the unobserved components Phillips curve<sup>6</sup>, a random walk is used to capture inflation persistence.

$$\begin{aligned} \pi_t &= \mu_t + \beta x_t + \varepsilon_t, & \varepsilon_t &\sim N(0, \sigma_\varepsilon^2) \\ \mu_t &= \mu_{t-1} + \eta_t, & \eta_t &\sim N(0, \sigma_\eta^2) \end{aligned} \quad (2.37)$$

---

<sup>6</sup>See [Harvey \(2014\)](#)

### *The State Space Model Bootstrap*

In (2.37),  $\pi_t$  and  $x_t$  represent inflation and the output gap, respectively.

Appealing to the state space formulation allows us to handle both the estimation of the stochastic parameters and the application of the bootstrap to the fixed parameters under investigation. In other words, the presence of time-varying parameters in no way hinders the execution of the bootstrap.

It is acknowledged that the errors in many regression models exhibit heteroskedasticity. However, for expedience, it is common to assume that they are *iid*. We estimate the linear, time-varying heteroskedastic model in (2.36) by Gaussian maximum likelihood, under the incorrect assumption of homogeneous errors. We then employ the WSSM bootstrap and compare the coverage probabilities of various confidence intervals constructed for  $\beta_0$ .

Equation (2.36) is simulated with  $X_0 \sim U(0,1)$ <sup>7</sup>,  $\beta_0 = 3$ ,  $X_1 \sim U(0,1)$ ,  $\varepsilon_t \sim N(0, \sigma_t^2)$  with  $\sigma_t = .3X_t$ ,  $\eta_t \sim N(0,1)$  and  $\beta_{1,t+1} = 0.9\beta_{1,t} + \eta_t$  where  $\beta_{1,1} = 0$ . Rademacher weights are applied in the WSSM bootstrap procedure.<sup>8</sup>

Coverage is calculated for the following regularly used confidence intervals. A nominal coverage of 95% provides the benchmark.

Standard normal confidence interval:

$$\hat{\beta}_0 \pm 1.96s_\infty$$

$\hat{\beta}_0$  is the maximum likelihood estimate of  $\beta_0$  and  $s_\infty$  is the corresponding asymptotic standard error.

---

<sup>7</sup>U represents the uniform distribution.

<sup>8</sup>Mammen weights produced similar results.

## *The State Space Model Bootstrap*

Normal bootstrap confidence interval:

$$\hat{\beta}_0 \pm 1.96s^*$$

$s^*$  is the bootstrap estimate of the coefficient standard error.

Simple percentile confidence interval:

The bootstrap estimates,  $\{\beta_{0,b}^*\}_{b=1}^B$ , are sorted in ascending order. The confidence limits are the .025 and the .975 quantiles of the ranked statistics.

Bootstrap- $t$  confidence interval:

For each of the  $B$  bootstrap samples,  $\beta_{0,b}^*$ , the bootstrap estimate of  $\beta_0$ , and its corresponding standard error,  $s_b^*$ , are calculated. We compute  $s_b^*$  by performing a second round of bootstrapping, employing the fast double bootstrap proposed by [Ouyse \(2013\)](#). We then construct,

$$z_b^* = \frac{\beta_{0,b}^* - \hat{\beta}_0}{s_b^*}, \quad b = 1, \dots, B.$$

In addition,  $s^*$ , the standard deviation of  $\{\beta_{0,b}^*\}_{b=1}^B$ , is evaluated. The  $\alpha^{th}$  percentile of  $\{z_b^*\}_{b=1}^B$  is estimated by  $\hat{t}_\alpha$  where

$$\frac{1}{B} \sum_{b=1}^B I(z_b^* \leq \hat{t}_\alpha) = \alpha.$$

A 95% confidence interval for  $\beta_0$  is given by

$$(\hat{\beta}_0 - \hat{t}_{.975}s^*, \hat{\beta}_0 - \hat{t}_{.025}s^*).$$

One thousand Monte Carlo simulations are performed, each using 999

bootstrap replications. The coverage probability is determined as the proportion of the time the true parameter value,  $\beta_0$ , falls into the relevant confidence interval. The lengths of the confidence intervals are also examined.

Note that, according to [MacKinnon \(2006\)](#) a minimum of  $B=999$  bootstrap replications will usually suffice in practice. However, when the bootstrap iterations are nested within a Monte Carlo simulation, a minimum of  $B=399$  is sufficient since the randomness due to  $B$  being small tends to average out across Monte Carlo rounds. These numbers guide our choice of  $B$  throughout the dissertation. The only exceptions occur when we are reproducing or making a direct comparison to another author's methodology; naturally, in those instances, we adhere to that author's choice for  $B$ .

## **2.6.2 Simulation I: Results and Discussion**

The purpose of simulation I was to compare the performance of Gaussian ML, under the false but frequently invoked assumption of homoskedasticity, with that of the WSSM bootstrap in a stochastic parameter regression model.

The data in [Table 2.1](#) indicates that, for all sample sizes, the confidence intervals generated using the WSSM bootstrap outperformed the normal confidence intervals which are based on asymptotic standard errors. The latter consistently under-covered and this outturn did not improve with increasing sample size. The poor coverage is consistent with the shorter standard normal confidence intervals that we see in [Table 2.2](#). These results are in keeping with the literature; for linear models like [\(2.36\)](#), in the presence of heteroskedasticity, the maximum likelihood estimator remains unbiased

but the conventional covariance matrix estimator is inconsistent. In this case, the estimate of the standard error is biased downward, leading to tighter confidence intervals.

Figure 1 highlights the relative functionality of the confidence intervals under study. In general, the normal bootstrap confidence intervals, which rely on the use of bootstrap standard errors, performed marginally better than the simple percentile intervals. Overall, the bootstrap- $t$  intervals were the most accurate. Unlike the other confidence intervals, they over-rejected for all sample sizes under consideration. However, the coverage of the bootstrap- $t$  intervals was closer to the nominal level of .95 than was the coverage of the other confidence intervals under investigation. This outcome was to be expected given the results of [Hall \(2013\)](#) which states that, while bootstrap- $t$  intervals can be computationally intensive, they tend to be more accurate than most other confidence intervals, particularly in small samples.

### **2.6.3 Simulation II**

Newton's method used for the maximisation of linear regression models converges in 1 step. However, in nonlinear estimation, multiple iterations are generally required to meet the convergence criterion of the optimisation routine. Using a gradient-based approach to the bootstrap, which requires numerical optimisation to be carried out only once, can result in considerable time savings. To investigate these efficiency gains, we consider the model,

$$\begin{aligned}\alpha_{t+1} &= \begin{bmatrix} \psi_1 & 1 \\ 0 & 0 \end{bmatrix} \alpha_t + \begin{bmatrix} 1 \\ \psi_2 \end{bmatrix} \eta_t, & \eta_t \sim N(0, 1) \\ y_t &= \begin{bmatrix} 1 & 0 \end{bmatrix} \alpha_t,\end{aligned}\quad (2.38)$$

with  $\psi_1 = 0.7$  and  $\psi_2 = 0.8$ . The parameter estimates and their standard errors are evaluated using the following methods:

1. Gaussian maximum likelihood.  $\hat{\psi}_1$  and  $\hat{\psi}_2$  are the ML estimates of  $\psi_1$  and  $\psi_2$ , respectively, while  $\text{std}(\hat{\psi}_1)$  and  $\text{std}(\hat{\psi}_2)$  are their associated standard errors.
2. The bootstrap of [Stoffer and Wall \(1991\)](#) introduced in Section 2.3. We generated  $B$  bootstrap samples and computed the maximum likelihood estimates,  $\{\psi_1^{*b}\}$  and  $\{\psi_2^{*b}\}$ , for  $b = 1, \dots, B$ , of each sample.  $\psi_{1,\text{stoffer\_wall}}$  and  $\psi_{2,\text{stoffer\_wall}}$  are the arithmetic means of  $\{\psi_1^{*b}\}$  and  $\{\psi_2^{*b}\}$ , for  $b = 1, \dots, B$ , respectively, while  $\text{std}(\psi_{1,\text{stoffer\_wall}})$  and  $\text{std}(\psi_{2,\text{stoffer\_wall}})$  are the corresponding standard deviations. These estimates are approximations of the means and standard deviations of the true sampling distributions of the maximum likelihood estimators,  $\hat{\psi}_1$  and  $\hat{\psi}_2$ .
3. The RSSM bootstrap developed in subsection 2.4.2. We generated  $B$  bootstrap samples and calculated  $\{\psi_{1,R}^{*b}\}$  and  $\{\psi_{2,R}^{*b}\}$ , for  $b = 1, \dots, B$ , for each sample. The values  $\psi_{1,RSSM}$  and  $\psi_{2,RSSM}$  are the averages of  $\{\psi_{1,R}^{*b}\}$  and  $\{\psi_{2,R}^{*b}\}$ , for  $b = 1, \dots, B$ , whereas  $\text{std}(\psi_{1,RSSM})$  and  $\text{std}(\psi_{2,RSSM})$  denote the respective standard deviations.



### *The State Space Model Bootstrap*

$B = 999$  bootstrap iterations are used in the investigation and all estimates are averaged over 1000 Monte Carlo replications. The results are compared with the means and standard deviations of the “true” finite-sample distributions of  $\hat{\psi}_1$  and  $\hat{\psi}_2$ . To generate the true distributions, we employ the same approach as [Stoffer and Wall \(1991\)](#), that is, equation (2.38) is simulated 1000 times with  $\psi_1 = 0.7$  and  $\psi_2 = 0.8$ , and each of these Monte Carlo rounds is estimated by maximum likelihood; the results are used to approximate the distributions of  $\hat{\psi}_1$  and  $\hat{\psi}_2$ . The means of these true sampling distributions of  $\hat{\psi}_1$  and  $\hat{\psi}_2$  are  $\psi_{1,true}$  and  $\psi_{2,true}$ , respectively, and their corresponding, respective standard deviations are  $\text{std}(\psi_{1,true})$  and  $\text{std}(\psi_{2,true})$ . The execution time for the two bootstrap methodologies is recorded and contrasted. All simulations are carried out on the same 16-core computer cluster.

#### **2.6.4 Simulation II: Results and Discussion**

For the model in (2.38), we simulated the true small sample distributions of the ML estimators of  $\psi_1$  and  $\psi_2$ . We approximated these distributions using the bootstrap of [Stoffer and Wall \(1991\)](#), the RSSM bootstrap and standard asymptotic theory. Overall, both the bootstrap of [Stoffer and Wall \(1991\)](#) and the RSSM bootstrap improved upon ML estimation in the sense that the standard deviations of the bootstrap distributions were closer to the true values determined by Monte Carlo simulation. The disparity in the dispersion estimates produced by the different methodologies was more pronounced in smaller samples, particularly for  $T = 50$ . (See Table 2.3). However, the differences between the ML and bootstrap results narrowed as the sample

### *The State Space Model Bootstrap*

size increased. By  $T = 200$ , the ML estimates were almost on par with those from the other two methods. This outcome suggests that utilising bootstrap standard errors, from either of the bootstrap methodologies, would be more reliable than employing asymptotic theory when dealing with limited sample sizes.

The bootstrap of [Stoffer and Wall \(1991\)](#) performed marginally better than the RSSM bootstrap when  $T = 50$ . However, as seen in [Table 2.3](#), the time saved by adopting the latter is considerable. At small sample sizes, there may be a slight trade-off between speed and accuracy when it comes to these two bootstrap algorithms.

At  $T = 50$ , the absolute difference between  $\text{std}(\psi_{1,true})$  and  $\text{std}(\psi_{1,stoffer\_wall})$  is 0.0137. By comparison, the absolute difference between  $\text{std}(\psi_{1,true})$  and the RSSM bootstrap estimate,  $\text{std}(\psi_{1,RSSM})$ , is 0.0171. So, we see that  $\text{std}(\psi_{1,stoffer\_wall})$  is closer to  $\text{std}(\psi_{1,true})$  than is  $\text{std}(\psi_{1,RSSM})$ . Similar results hold for  $\psi_2$ , namely,  $|\text{std}(\psi_{2,true}) - \text{std}(\psi_{2,stoffer\_wall})| = 0.0141$ , while  $|\text{std}(\psi_{2,true}) - \text{std}(\psi_{2,RSSM})| = 0.0181$ . However, when we consider  $T > 50$ , the estimates for both bootstrap methods are comparable. For the periods  $T = 100$ ,  $T = 150$  and  $T = 200$ , when considering  $\psi_1$ , the average absolute differences for the 2 bootstrap techniques are  $\sum_T \left| \frac{\text{std}(\psi_{1,true}) - \text{std}(\psi_{1,stoffer\_wall})}{3} \right| = 0.0036$  and  $\sum_T \left| \frac{\text{std}(\psi_{1,true}) - \text{std}(\psi_{1,RSSM})}{3} \right| = 0.00127$ . Additionally, with regard to  $\psi_2$ , the average absolute differences are  $\sum_T \left| \frac{\text{std}(\psi_{2,true}) - \text{std}(\psi_{2,stoffer\_wall})}{3} \right| = 0.0039$  and  $\sum_T \left| \frac{\text{std}(\psi_{2,true}) - \text{std}(\psi_{2,RSSM})}{3} \right| = 0.0052$ . These differences are an order of magnitude smaller than the differences at  $T = 50$  and can be considered minor.

When sample sizes are small, it is possible that a procedure which utilises

$k$  Newton steps, where  $k$  is a small integer, may perform better than our 1-step RSSM bootstrap. This idea will be explored further in the next chapter.

### 2.6.5 Simulation III

The aim of this investigation is to test the efficacy of the SSM bootstrap in the context of model selection. We compare the performance of the standard AIC with that of the AICg in state space models with heterogeneous errors. Recall,

$$\text{AICg} = -2\log\mathbb{L}(\hat{\psi}|Y_T) + 2 \left\{ \frac{1}{B} \sum_{j=1}^B -2\log \frac{\mathbb{L}(\psi^*(j)|Y_T)}{\mathbb{L}(\hat{\psi}|Y_T)} \right\}, \quad (2.39)$$

where the terms are as defined in section (2.5). Here,  $\psi^*$ , the bootstrap estimate of the true parameter vector,  $\psi$ , is found by implementing the WSSM bootstrap. We consider the following distributed lag model:

$$\begin{aligned} y_t &= \mu_t + 0.9X_t + 0.7X_{t-1} + 1.3X_{t-2} + \epsilon_t, & \epsilon_t &\sim N(0, 0.04X_t) \\ \mu_{t+1} &= 0.6\mu_t + \eta_t, & \eta_t &\sim N(0, 0.01), \end{aligned} \quad (2.40)$$

where,  $X_t \sim U(0, 1)$ . One hundred realisations of the true model (2.40) are generated and candidate models, with lag lengths on the exogenous variable ranging from 1 to 12, are fit to the data. Namely, we fit,

$$\begin{aligned} y_t &= \mu_t + \beta_0X_t + \beta_1X_{t-1} + \dots + \beta_pX_{t-p} + \epsilon_t, & \epsilon_t &\sim N(0, \sigma_{\epsilon_t}^2) \\ \mu_{t+1} &= \gamma\mu_t + \eta_t, & \eta_t &\sim N(0, \sigma_{\eta}^2), \end{aligned} \quad (2.41)$$

### *The State Space Model Bootstrap*

where  $1 \leq p \leq 12$ . The sample size considered in this scenario is  $T = 20$ . For every realisation, the AIC and AICg are calculated for each candidate model. In implementing the bootstrap, 250 replications are performed and the weights,  $W_t$ , are drawn from the Rademacher distribution. The selection results are recorded for the 100 realisations and presented in table 2.4. The choice of sample size and the number of bootstrap iterations was guided by those used in [Cavanaugh and Shumway \(1997\)](#).

To observe the performance of both selection criteria in large samples, the analysis was repeated for  $T = 200$ .

### **2.6.6 Simulation III: Results and Discussion**

We compared the performance of the AIC and the AICg under 2 different scenarios. When  $T = 20$ , it can be seen from Table 2.4 that the AIC, which is known to choose high-dimensional models in small samples, tended to overfit and select models with more parameters. By comparison, the AICg obtained a much greater number of correct order selections.

When  $T = 200$ , it is apparent from Table 2.5 that the AICg and the traditional AIC perform similarly. This is in accordance with established research regarding the favourable large-sample properties of the AIC ([Cavanaugh and Neath, 2019](#)). Naturally, we would not use the AICg when the computationally more efficient AIC would suffice. However, when sample sizes are small, the AICg is more successful at choosing the correct model dimension.

Overall, these results are similar to the outcomes observed in [Cavanaugh](#)

and Shumway (1997) where homoskedastic state space models were considered and the bootstrap of Stoffer and Wall (1991) was employed. The present results are significant in one respect. Since the approach of Stoffer and Wall (1991) cannot be used in heteroskedastic models, the WSSM bootstrap, via the AICg, could be a useful tool for model selection when working with smaller sample sizes in state space models with heterogeneous error terms.

## **2.7 Concluding Remarks**

The primary aim of this investigation was to formulate a bootstrap that was applicable to heteroskedastic state space models. To this end, we advanced the work of Kline and Santos (2012) and developed the wild state space model bootstrap. We discovered that, when we improperly assumed homoskedasticity, the bootstrap confidence intervals constructed using the WSSM bootstrap estimates of the standard error outperformed intervals constructed from standard errors that were based on asymptotic theory. Additionally, in model selection, the AICg measure, which uses WSSM bootstrap estimates, had a greater hit ratio than the AIC in small samples. We also devised the residual state space model bootstrap for use in homoskedastic state space models. While a bootstrap methodology for homoskedastic state space models already exists, the RSSM bootstrap was shown to be more computationally efficient.

## 2.8 Tables and Graphs

Table 2.1: Coverage Probabilities for Simulation I

$T$	Standard Normal	Normal Bootstrap	Simple Percentile	Bootstrap- $t$
50	0.748	0.841	0.835	0.979
75	0.788	0.882	0.873	0.987
100	0.836	0.901	0.896	0.980
125	0.811	0.897	0.891	0.983
150	0.768	0.898	0.898	0.980
175	0.758	0.911	0.910	0.981
200	0.824	0.909	0.905	0.982
225	0.796	0.913	0.912	0.977
250	0.804	0.911	0.905	0.973
275	0.813	0.907	0.908	0.986
300	0.811	0.913	0.910	0.982

Figure 2.1: Coverage Probabilities for Simulation I

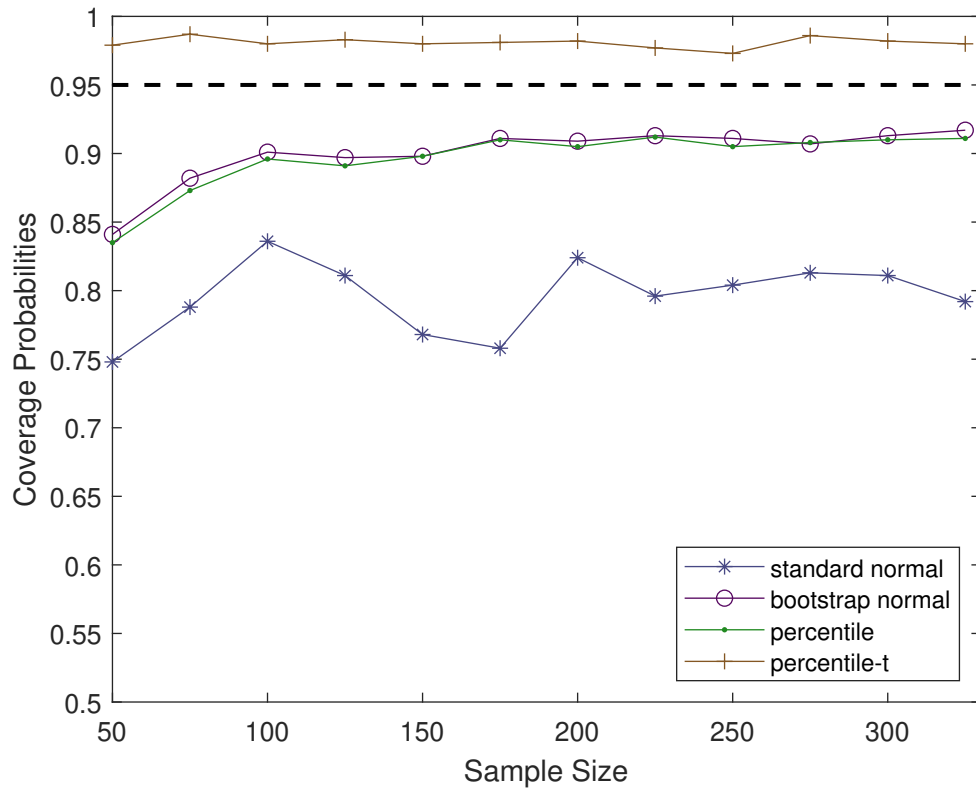


Table 2.2: Length of Confidence Intervals

$T$	Standard Normal	Normal Bootstrap	Simple Percentile	Bootstrap $_t$
50	.2009	.2961	.2942	1.5605
75	.2319	.2510	.2568	1.2172
100	.1521	.2168	.2109	0.98248
125	.1760	.2603	.2589	0.86873
150	.1367	.1743	.1727	0.81437
175	.1475	.1786	.1725	0.76662
200	.1250	.1390	.1385	0.68634
225	.1128	.1659	.1652	0.66164
250	.1396	.1326	.1336	0.59125
275	.1003	.1359	.1361	0.60973
300	.0962	.1246	.1259	0.54849



Table 2.3: Parameter Estimates and Running Times for Simulation II

$T$	True Distribution		ML		Stoffer-Wall		RSSM	
	$\psi_{1,true}$	$\psi_{2,true}$	$\hat{\psi}_1$	$\hat{\psi}_2$	$\psi_{1,stofer\_wall}$	$\psi_{2,stofer\_wall}$	$\psi_{1,RSSM}$	$\psi_{2,RSSM}$
50	0.6698 (0.1168)	0.8184 (0.1097)	0.7726 (0.0546)	0.8901 (0.0476)	0.7442 (0.1031)	0.9022 (0.0956)	0.7512 (0.0997)	0.8951 (0.0916)
100	0.6856 (0.0757)	0.8126 (0.0706)	0.7261 (0.0517)	0.8116 (0.0526)	0.7076 (0.0781)	0.8220 (0.0702)	0.7113 (0.0753)	0.8170 (0.0734)
150	0.6928 (0.0620)	0.8051 (0.0543)	0.6316 (0.0743)	0.7733 (0.0587)	0.6207 (0.0664)	0.7791 (0.0584)	0.6231 (0.0624)	0.7761 (0.0600)
200	0.6944 (0.0519)	0.8040 (0.0459)	0.6888 (0.0666)	0.7348 (0.0627)	0.6828 (0.0559)	0.7373 (0.0531)	0.6837 (0.0549)	0.7360 (0.0530)

---

<sup>9</sup>The standard errors appear in parentheses.

*The State Space Model Bootstrap*

Table 2.4: Model Selections for Simulation III,  $T = 20$ .

Lags	AIC	AICg
1	0	8
2	6	78
3	7	2
4	9	6
5	20	2
6	9	0
7	18	0
8	9	0
9 to 12	22	4

Table 2.5: Model Selections for Simulation III,  $T = 200$ .

Lags	AIC	AICg
1	0	0
2	80	80
3	8	14
4	2	2
5	0	0
6	4	2
7	2	2
8	0	0
9 to 12	4	0

# Chapter 3

## The $k$ -Step Bootstrap In State Space Models

### 3.1 Introduction

In the previous chapter we developed a gradient-based bootstrap for use in state space models. However, the performance of the state space model bootstrap in small samples ( $n \leq 50$ ) may be improved if we take more than one Newton step during the optimisation procedure. The main objective of this chapter is to assess the effectiveness of the  $k$ -step bootstrap introduced by [Davidson and MacKinnon \(1999\)](#), DM hereafter, when it is applied to models in state space form. Bootstrapping nonlinear models tends to be computationally expensive. DM proposed an approximate bootstrap methodology which attains the accuracy of standard bootstrap inference without requiring repeated nonlinear optimisation. The authors noted that bootstrap tests are, in general, not exact and that they make an error which

### *The $k$ -Step Bootstrap In State Space Models*

is proportional to the sample size raised to some negative power. Since the bootstrap test is already in error by a certain amount, it suffices to calculate the bootstrap test statistic to an order that is equal to the error made by the bootstrap test. This is achieved by taking  $k$  Newton or quasi-Newton steps toward the optimum, where  $k$  is a small integer. Therefore, rather than executing  $B + 1$  nonlinear estimations as in the case of the traditional bootstrap, where  $B$  is the number of bootstrap iterations, it is enough to perform one nonlinear optimisation along with  $kB$  Newton or quasi-Newton steps. This can significantly reduce computational costs.

This study also highlights the benefits of casting models in state space form for estimation and testing. The output of the Kalman filter can be used for likelihood evaluation, the construction of the score vector, the creation of the outer-product-of-the-gradient (OPG) form of the covariance matrix of parameter estimates and consequently for the development of several test statistics. The fact that these quantities can all be formulated by using output from the filter makes working in the state space framework extremely convenient.

The chapter is organised in the following way: Section 2 reviews the methodology of DM; Section 3 formulates the Score, Likelihood Ratio and Wald statistics using Kalman filter output; Section 4 investigates the accuracy of the  $k$ -step bootstrap applied to state space models and Section 5 concludes.

## 3.2 Summary of the $k$ -Step Bootstrap

Let  $f_n(\theta)$  be a random criterion function, such as a log-likelihood function, which is based on a sample of size  $n$ . Let  $\hat{\theta}$  be the maximum likelihood estimate of  $\theta$ . DM derived equations for the rate of convergence of both Newton and quasi-Newton optimisation algorithms<sup>1</sup>. Let the initial estimate of  $\theta$  be  $\theta_{(0)}$  and assume that  $(\theta_{(0)} - \hat{\theta}) = O_p(n^{-\frac{1}{2}})$ . Recall that one Newton step is given by

$$\theta_{(1)} = \theta_{(0)} - [H_n(\theta_{(0)})]^{-1} g_n(\theta_{(0)}), \quad (3.1)$$

where  $H_n(\theta)$  and  $g_n(\theta)$  are the Hessian matrix and score vector of  $f_n(\theta)$ , respectively. Let  $\theta_{(i)}$  denote the estimate of  $\theta$  after  $i$  iterations of Newton's formula. The authors demonstrated that, under certain regularity conditions,<sup>2</sup>

$$\theta_{(i)} - \hat{\theta} = O_p(n^{-2^{i-1}}). \quad (3.2)$$

Replacing  $H_n(\theta)$  in (3.1) with an approximation to the Hessian represents one quasi-Newton step. After  $i$  quasi-Newton iterations, DM found that,

$$\theta_{(i)} - \hat{\theta} = O_p\left(n^{-\frac{i+1}{2}}\right). \quad (3.3)$$

### 3.2.1 The Conventional Bootstrap

What follows is the typical procedure for the bootstrap method in the context of hypothesis testing. We partition  $\theta$ , the  $m$ -vector of parameters, into  $\theta =$

---

<sup>1</sup>These derivations were based on earlier work done by [Robinson \(1988\)](#).

<sup>2</sup>The regularity conditions are outlined in [Robinson \(1991\)](#) and [Davidson and MacKinnon \(1999\)](#).

### *The k-Step Bootstrap In State Space Models*

$[\theta_1 : \theta_2]$  where  $\theta_1$  is an  $m_1$ -vector,  $\theta_2$  is an  $m_2$ -vector and  $m_1 + m_2 = m$ . Without any loss of generality, we can test the null hypothesis that  $\theta_2 = 0$  and reject the null hypothesis for large values of a test statistic,  $\hat{\tau}$ , where  $\hat{\tau}$  is, for example, a Likelihood ratio, Score or Wald statistic. The steps to be undertaken are:

1. Compute the test statistic,  $\hat{\tau}$ , for the original data set and calculate a vector of parameter estimates,  $\tilde{\theta} = [\tilde{\theta}_1 : 0]$ , which satisfy the null hypothesis.
2. Use  $\tilde{\theta}$  in the bootstrap data generating process (DGP) to produce  $B$  bootstrap samples.
3. For the  $j^{\text{th}}$  bootstrap sample, calculate the bootstrap test statistic,  $\hat{\tau}_j^*$ , in the same way that  $\hat{\tau}$  was computed.
4. Estimate the bootstrap  $p$ -value as the proportion of bootstrap samples for which  $\hat{\tau}_j^*$  exceeds  $\hat{\tau}$ .

### **3.2.2 The Approximate Bootstrap**

The above technique requires that optimisation be carried out  $B+1$  times, once when we calculate  $\hat{\tau}$  and  $B$  times when we calculate  $\hat{\tau}_j^*$ . The high cost of nonlinear estimation can prove to be an obstacle when applying the bootstrap. DM proposed replacing the nonlinear optimisation in each bootstrap sample with a small number,  $k$ , of Newton or quasi-Newton steps. This is referred to as the  $k$ -step or approximate bootstrap. Performing  $kB$  steps is typically faster than conducting  $B$  nonlinear optimisations.

## *The k-Step Bootstrap In State Space Models*

Maximum likelihood estimation of state space models is customarily carried out using quasi-Newton methods such as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) and the Davidon-Fletcher-Powell (DFP) optimisation routines. These approaches are preferred to Newton's technique because they do not require the computation of the Hessian. In state space models, the analytic expression for the Hessian is complex. Moreover, Newton's procedure requires that the Hessian be inverted at each step. In contrast, quasi-Newton methods use an approximation to the Hessian which does not involve second derivatives and whose inverse is updated directly, rather than by solving a system of equations as is the case with Newton's approach. For these reasons, only quasi-Newton algorithms are considered in this paper.

### **The Score Test**

Let the criterion function under consideration be the log-likelihood,  $L(\theta)$ . The Score test statistic,  $\mathbb{T}_s$ , is given by

$$\mathbb{T}_s(\tilde{\theta}) = g'(\tilde{\theta})\mathcal{I}^{-1}(\tilde{\theta})g(\tilde{\theta}). \quad (3.4)$$

In equation (3.4),  $g(\tilde{\theta})$  is the gradient of  $L(\theta)$  and  $\mathcal{I}(\tilde{\theta})$  is a consistent estimate of the information matrix of  $L(\theta)$ , evaluated at the restricted parameter vector,  $\tilde{\theta}$ .

DM remarked that bootstrap tests are ordinarily not exact, and that it is normally possible to find an integer,  $l$ , so that the rejection probability for the bootstrap test at a nominal level  $\alpha$  differs from  $\alpha$  by an amount that is

*The k-Step Bootstrap In State Space Models*

$O(n^{-\frac{l}{2}})$ . Notably, the same order of accuracy can be realised if there is an error that is  $O_p(n^{-\frac{l}{2}})$  in the evaluation of the bootstrap  $p$ -values.

The bootstrap data is generated using the restricted estimates  $\tilde{\theta} = [\tilde{\theta}_1 : 0]$ . By construction,  $(\tilde{\theta}_1^* - \tilde{\theta}_1) = O_p(n^{-\frac{1}{2}})$ , where  $\tilde{\theta}_1^*$  is the restricted estimate from a bootstrap sample. Consequently,  $\tilde{\theta}_1$  is a suitable starting point for the quasi-Newton algorithm applied to the bootstrap model under the null restriction. DM found that

$$\tilde{\theta}_{1(i)}^* - \tilde{\theta}_1^* = O_p\left(n^{-\frac{(i+1)}{2}}\right). \quad (3.5)$$

In (3.5),  $\tilde{\theta}_{1(i)}^*$  is the parameter estimate from the restricted bootstrap model after  $i$  iterations of the quasi-Newton routine. Sequential approximations to the Score statistic are given by

$$\mathbb{T}_s(\tilde{\theta}_{(i)}^*) = g'(\tilde{\theta}_{(i)}^*)\mathcal{I}^{-1}(\tilde{\theta}_{(i)}^*)g(\tilde{\theta}_{(i)}^*). \quad (3.6)$$

The authors go on to show that

$$\mathbb{T}_s(\tilde{\theta}_{(i)}^*) = \mathbb{T}_s(\tilde{\theta}^*) + n^{\frac{1}{2}}O_p(\tilde{\theta}_{(i)}^* - \tilde{\theta}^*). \quad (3.7)$$

From (3.5) and (3.7), it is clear that the difference between  $\mathbb{T}_s(\tilde{\theta}_{(i)}^*)$  and  $\mathbb{T}_s(\tilde{\theta}^*)$  is on the order of  $n^{-\frac{l}{2}}$ . Assume the bootstrap  $p$ -values are in error at order  $n^{-\frac{l}{2}}$ . Then, the number of steps,  $k$ , needed to achieve at least the same order of accuracy as the bootstrap, should be chosen so that  $n^{-\frac{k}{2}} \leq n^{-\frac{l}{2}}$ , that is, so that  $k \geq l$ . Therefore, for  $l = 3$ , we would need to take at least three quasi-Newton steps. The approximate bootstrap procedure of DM for the



### *The k-Step Bootstrap In State Space Models*

Score test is therefore:

1. Calculate  $\mathbb{T}_s(\tilde{\theta})$  using the original data set.
2. Generate  $B$  bootstrap samples of size  $n$  using the bootstrap DGP characterised by  $\tilde{\theta}$ .
3. Choose the number of steps  $k \geq l$ .
4. Initialise the system with  $\tilde{\theta}_{1(0)}^* = \tilde{\theta}_1$ . For each bootstrap sample perform  $k$  steps of the quasi-Newton algorithm on the restricted model in order to obtain the iterated estimates  $\tilde{\theta}_{1(k)}^*$ .
5. Compute  $\mathbb{T}_s(\tilde{\theta}_{1(k)}^*)$  using equation (3.6).

### **The Likelihood Ratio Test**

The likelihood ratio (LR) statistic,  $\mathbb{T}_{LR}$ , is defined as,

$$\mathbb{T}_{LR}(\hat{\theta}, \tilde{\theta}) = 2 \left[ L(\hat{\theta}) - L(\tilde{\theta}) \right],$$

where  $L(\hat{\theta})$  and  $L(\tilde{\theta})$  are the unrestricted and restricted log-likelihoods, respectively. DM found that, if a quasi-Newton method is employed,

$$L(\hat{\theta}^*) - L(\hat{\theta}_{(i)}^*) = O_p(n^{-i}). \quad (3.8)$$

Similarly,

$$L(\tilde{\theta}^*) - L(\tilde{\theta}_{(i)}^*) = O_p(n^{-i}). \quad (3.9)$$

They demonstrate that  $k \geq \frac{l}{2}$  steps are required to guarantee that the error

### *The $k$ -Step Bootstrap In State Space Models*

in the LR statistic is at most on the order of  $n^{-\frac{l}{2}}$ . Therefore, for  $l = 4$ , we should choose  $k$  to be at least 2.

The algorithm for the approximate bootstrap LR test is:

1. Construct  $\mathbb{T}_{LR}$  using the initial data set.
2. Generate  $B$  bootstrap samples under the null hypothesis.
3. Determine the number of steps  $k \geq \frac{l}{2}$ .
4. For each bootstrap sample,
  - (a) Starting at  $\hat{\theta}_{(0)}^* = \hat{\theta}$ , perform  $k$  quasi-Newton steps on the unrestricted model to obtain  $\hat{\theta}_{(k)}^*$ .
  - (b) Beginning at  $\tilde{\theta}_{1(0)}^* = \tilde{\theta}_1$ , perform  $k$  quasi-Newton steps on the restricted model to obtain  $\tilde{\theta}_{1(k)}^*$  and construct  $\tilde{\theta}_{(k)}^* = [\tilde{\theta}_{1(k)}^* \vdots 0]$ .
5. Calculate  $\mathbb{T}_{LR}(\hat{\theta}_{(k)}^*, \tilde{\theta}_{(k)}^*) = 2 \left[ L(\hat{\theta}_{(k)}^*) - L(\tilde{\theta}_{(k)}^*) \right]$ .

### **The Wald Test**

The formula for the Wald statistic,  $\mathbb{T}_w$ , is

$$\mathbb{T}_w(\hat{\theta}) = (n^{\frac{1}{2}}\hat{\theta}_2)' \hat{V}^{-1} (n^{\frac{1}{2}}\hat{\theta}_2). \quad (3.10)$$

where  $\hat{V}$  is a consistent estimate of the asymptotic covariance matrix of  $n^{\frac{1}{2}}\hat{\theta}_2$ .

The approximation error was shown to be

$$\mathbb{T}_w(\hat{\theta}_{(i)}^*) - \mathbb{T}_w(\hat{\theta}^*) = n^{\frac{1}{2}}O_p(\hat{\theta}_{(i)}^* - \hat{\theta}^*).$$

### The $k$ -Step Bootstrap In State Space Models

As in the case of the score test,  $k \geq l$  steps are needed to ensure that the error in  $\mathbb{T}_w(\hat{\theta}_{(i)}^*)$  is at most  $n^{-\frac{l}{2}}$ .

The procedure for the approximate bootstrap Wald test is:

1. Compute  $\mathbb{T}_w(\hat{\theta})$  from the primary data set.
2. Generate  $B$  bootstrap samples under the null hypothesis.
3. Choose the number of steps  $k \geq l$ .
4. For each bootstrap sample, perform  $k$  quasi-Newton steps on the unrestricted model, starting from  $\hat{\theta}_{(0)}^* = \hat{\theta}$ .
5. Compute  $\mathbb{T}_w(\hat{\theta}_{(k)}^*) = \left( n^{\frac{1}{2}} \hat{\theta}_{2(k)}^* \right)' \left[ \hat{V}(\hat{\theta}_{2(k)}^*) \right]^{-1} \left( n^{\frac{1}{2}} \hat{\theta}_{2(k)}^* \right)$ .

If we know the order of the error that a bootstrap test will make *a priori*, then determining  $k$  is straightforward and proceeds as previously described. However, when we have no prior information, we take the most conservative approach to its calculation. The rejection probability of a bootstrap test should never, under normal circumstances, be in error by more than  $O(n^{-1})$  (Davidson and MacKinnon, 1999). So, when we have no information regarding the error in rejection probability, we use  $O(n^{-1})$  to derive  $k$ .

## 3.3 Constructing Test Statistics Using Kalman Filter Output

We now discuss how to construct the three classical test statistics using output from the Kalman filter. As in the previous section, we consider the null hypothesis that  $\theta_2 = 0$ .

*The k-Step Bootstrap In State Space Models*

Recall, from Chapter 1, the prediction error decomposition form of the log-likelihood (1.11),

$$l_t = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log|F_t| - \frac{1}{2}\nu_t'F_t^{-1}\nu_t, \quad (3.11)$$

$$L(\theta) = \sum_{t=1}^n l_t, \quad (3.12)$$

where  $l_t$  is the log-likelihood of observation  $t$ . The prediction error,  $\nu_t$ , and its covariance,  $F_t$ , are functions of the model parameters,  $\theta$ .

The likelihood ratio statistic,  $\mathbb{T}_{LR}$ , can be constructed by evaluating  $L(\theta)$  under the null and alternative hypotheses,

$$\mathbb{T}_{LR} = 2[L(\hat{\theta}) - L(\tilde{\theta})]. \quad (3.13)$$

The exact gradient can also be derived from the Kalman filter equations.<sup>3</sup> Consider the partial derivative of  $l_t$  with respect to  $\theta_i$ . The  $(m \times n)$  matrix  $G(y, \theta)$  has typical element

$$G_{ti}(\theta) = \frac{\partial l_t}{\partial \theta_i} = -\frac{1}{2} \left[ \text{tr} \left[ F_t^{-1} \frac{\partial F_t}{\partial \theta_i} \right] [I - F_t^{-1} \nu_t \nu_t^{-1}] \right] - \left( \frac{\partial \nu_t}{\partial \theta_i} \right)' F_t^{-1} \nu_t, \quad i = 1, \dots, m. \quad (3.14)$$

---

<sup>3</sup>Recursions for the gradient can be found in [Harvey \(1990\)](#).

*The k-Step Bootstrap In State Space Models*

The  $(m \times 1)$  gradient vector,  $g(\theta)$ , has typical element,

$$g_i(\theta) = \frac{\partial L}{\partial \theta_i} = -\frac{1}{2} \sum_t \left\{ \text{tr} \left[ \left[ F_t^{-1} \frac{\partial F_t}{\partial \theta_i} \right] (I - F_t^{-1} \nu_t \nu_t^{-1}) \right] + 2 \left( \frac{\partial \nu_t}{\partial \theta_i} \right)' F_t^{-1} \nu_t \right\}, \quad i = 1, \dots, m, \quad (3.15)$$

where each element of the gradient vector is the sum of the elements of one of the rows of the matrix  $G(y, \theta)$ .

The outer product of the gradient (OPG) form of the Fisher information is defined as

$$\mathcal{I}(\theta) = E_\theta \left[ G(\theta) G'(\theta) \right], \quad (3.16)$$

and we can estimate  $n^{-1} \mathcal{I}(\hat{\theta})$  consistently by  $n^{-1} G(\hat{\theta}) G'(\hat{\theta})$ . The OPG estimator of the covariance matrix of  $\sqrt{n} \hat{\theta}$  is,

$$\widehat{Var}_{OPG}(\sqrt{n} \hat{\theta}) = \left( n G(\hat{\theta}) G'(\hat{\theta}) \right)^{-1}. \quad (3.17)$$

Using (3.14) and (3.15), the Score statistic,  $\mathbb{T}_s(\tilde{\theta})$ , can therefore be entirely constructed from Kalman filter output:

$$\mathbb{T}_s(\tilde{\theta}) = g'(\tilde{\theta}) \left[ G(\tilde{\theta}) G'(\tilde{\theta}) \right]^{-1} g(\tilde{\theta}).$$

Also, using (3.14), the Wald statistic can be expressed as

$$\mathbb{T}_w(\hat{\theta}) = (\sqrt{n} \hat{\theta}_2)' \left[ n G(\hat{\theta}_2) G'(\hat{\theta}_2) \right]^{-1} (\sqrt{n} \hat{\theta}_2).$$

## 3.4 Simulations

### 3.4.1 Simulation I

In Simulation II of Chapter 2 we considered the model,

$$\begin{aligned}\alpha_{t+1} &= \begin{bmatrix} \psi_1 & 1 \\ 0 & 0 \end{bmatrix} \alpha_t + \begin{bmatrix} 1 \\ \psi_2 \end{bmatrix} \eta_t, & \eta_t &\sim N(0, 1) & (3.18) \\ y_t &= \begin{bmatrix} 1 & 0 \end{bmatrix} \alpha_t.\end{aligned}$$

To approximate the distributions of the maximum likelihood estimators,  $\hat{\psi}_1$  and  $\hat{\psi}_2$ , we employed the bootstrap of [Stoffer and Wall \(1991\)](#), normal asymptotic theory and the 1-step RSSM bootstrap. In addition, the “true” small sample distributions of the parameter estimates were simulated using Monte Carlo methods. In the current analysis, we compare the results of the previous chapter to the ones we get from implementing the  $k$ -step bootstrap of DM. We use  $k=2$  and  $k=3$  quasi-Newton steps in our approximate bootstraps. The estimates are averaged over 1000 Monte Carlo rounds and the time taken for the various bootstrap procedures is recorded. Optimisation is carried out using the BFGS algorithm.  $B = 999$  bootstrap iterations are used in the study.

### 3.4.2 Simulation I: Results and Discussion

Simulation I contrasted the performances of the 2-step and 3-step bootstraps with those of the RSSM bootstrap and the bootstrap of [Stoffer and Wall \(1991\)](#). Table [3.1](#) presents the results.

### The $k$ -Step Bootstrap In State Space Models

$\psi_{1,true}$  and  $\psi_{2,true}$  were calculated from Monte Carlo simulations. One thousand data sets were generated from model (3.18), using the true parameters,  $\psi_1 = 0.7$  and  $\psi_2 = 0.8$ . Each data set was then estimated by maximum likelihood to provide  $\{\hat{\psi}_{1,i}\}_{i=1}^{1000}$  and  $\{\hat{\psi}_{2,i}\}_{i=1}^{1000}$ . In Table 3.1, the values  $\psi_{1,true}$  and  $\psi_{2,true}$  are the arithmetic means of  $\{\hat{\psi}_{1,i}\}_{i=1}^{1000}$  and  $\{\hat{\psi}_{2,i}\}_{i=1}^{1000}$ , respectively. The estimates  $\text{std}(\psi_{1,true})$  and  $\text{std}(\psi_{2,true})$  are the standard deviations of  $\{\hat{\psi}_{1,i}\}_{i=1}^{1000}$  and  $\{\hat{\psi}_{2,i}\}_{i=1}^{1000}$ , respectively. The means and standard deviations of  $\{\hat{\psi}_{1,i}\}_{i=1}^{1000}$  and  $\{\hat{\psi}_{2,i}\}_{i=1}^{1000}$  represent the means and standard deviations of the true sampling distributions of the maximum likelihood estimators.

$\hat{\psi}_{1,ML}$  and  $\hat{\psi}_{2,ML}$  are the maximum likelihood estimates of the model in (3.18), while  $\text{std}(\hat{\psi}_{1,ML})$  and  $\text{std}(\hat{\psi}_{2,ML})$  are their respective asymptotic standard errors.

When we executed the bootstrap of [Stoffer and Wall \(1991\)](#), we produced  $B$  bootstrap samples. We obtained the maximum likelihood estimates,  $\{\psi_1^{*b}\}$  and  $\{\psi_2^{*b}\}$ , for  $b = 1, \dots, B$ , of each sample. In Table 3.1,  $\psi_{1,\text{stoffer\_wall}}$  and  $\psi_{2,\text{stoffer\_wall}}$  are the arithmetic means of  $\{\psi_1^{*b}\}$  and  $\{\psi_2^{*b}\}$ , for  $b = 1, \dots, B$ , respectively, while  $\text{std}(\psi_{1,\text{stoffer\_wall}})$  and  $\text{std}(\psi_{2,\text{stoffer\_wall}})$  are the associated standard deviations. These estimates approximate of the means and standard deviations of the true sampling distributions of the maximum likelihood estimators,  $\hat{\psi}_1$  and  $\hat{\psi}_2$ .

The estimates based on the RSSM bootstrap were calculated in Simulation II of the previous chapter and are reproduced in the Table 3.1. These 1-step estimates of the means of the true distributions of the maximum likelihood estimators are denoted by  $\psi_{1,RSSM}$  and  $\psi_{2,RSSM}$ . Their standard deviations

## The $k$ -Step Bootstrap In State Space Models

are given by  $\text{std}(\psi_{1,RSSM})$  and  $\text{std}(\psi_{2,RSSM})$ , respectively.

When we applied the approximate bootstrap of DM to the model in (3.18), we first generated  $B$  bootstrap samples. For each of the  $B$  samples, we found the estimates  $\{\psi_{1,k=2}^{*b}\}$  and  $\{\psi_{2,k=2}^{*b}\}$ , for  $b = 1, \dots, B$ , by implementing 2 steps of the BFGS quasi-Newton algorithm. In Table 3.1,  $\psi_{1,k=2}^*$  and  $\psi_{2,k=2}^*$  are the averages of  $\{\psi_{1,k=2}^{*b}\}$  and  $\{\psi_{2,k=2}^{*b}\}$ , for  $b = 1, \dots, B$ , respectively, while  $\text{std}(\psi_{1,k=2}^*)$  and  $\text{std}(\psi_{2,k=2}^*)$  are the respective standard deviations. These are the 2-step estimates of the means and standard deviations of the true small sample distributions of the maximum likelihood estimators,  $\hat{\psi}_1$  and  $\hat{\psi}_2$ . We also calculated these estimates for  $k = 3$  steps. They appear in Table 3.1 as  $\psi_{1,k=3}^*$ ,  $\psi_{2,k=3}^*$ ,  $\text{std}(\psi_{1,k=3}^*)$  and  $\text{std}(\psi_{2,k=3}^*)$ .

The entries are colour-coded for easy interpretation. Specifically, the estimates of  $\psi_1$  obtained via the various methodologies are shaded in a light blue colour, facilitating comparison across different techniques. The corresponding standard errors for these estimates are shaded in a darker blue. Similarly, the estimates of  $\psi_2$  are shaded in a light purple colour, while the corresponding standard errors are shaded in a darker purple.

In Simulation II of the previous chapter, we noted that when  $n = 50$ , the bootstrap of [Stoffer and Wall \(1991\)](#) performed slightly better than the RSSM bootstrap. In particular, we found that  $|\text{std}(\psi_{1,true}) - \text{std}(\psi_{1,stoffer\_wall})| = 0.0137$  while  $|\text{std}(\psi_{1,true}) - \text{std}(\psi_{1,RSSM})| = 0.0171$ . When we examine Table 3.1, we see that, for  $k=3$ ,  $|\text{std}(\psi_{1,true}) - \text{std}(\psi_{1,k=3})| = 0.0147$ , that is,  $\text{std}(\psi_{1,k=3})$  is closer to  $\text{std}(\psi_{1,true})$  than is  $\text{std}(\psi_{1,RSSM})$ . Conversely,  $|\text{std}(\psi_{2,true}) - \text{std}(\psi_{2,k=3})| = 0.0203$ , compared to a difference of 0.0181 for  $|\text{std}(\psi_{2,true}) - \text{std}(\psi_{2,RSSM})|$ . So, in the case of  $\psi_2$ , the 3-step estimate is further



### *The $k$ -Step Bootstrap In State Space Models*

away from the true value than is the RSSM estimate. Therefore, the findings when  $k = 3$  are ambiguous.<sup>4</sup> The values of  $k$  were chosen to be in line with the work of DM. However, we know that, as  $k$  increases, the  $k$ -step bootstrap estimates will eventually converge to those of the full bootstrap of [Stoffer and Wall \(1991\)](#). Therefore, it is reasonable to expect that if we continue to increase  $k$ , we will eventually reach a point where the approximate bootstrap definitively outperforms the RSSM bootstrap. In this model, that scenario occurs when  $k = 6$ . Specifically, for  $k = 6$ , the standard errors of  $\psi_1$  and  $\psi_2$  are 0.1025 and 0.919, respectively, values which are closer to the true standard errors than are the standard errors calculated from the RSSM bootstrap.

It is worth noting that the  $k$ -step bootstrap approximations to the distributions of the maximum likelihood estimators outmatched the corresponding asymptotic approximations based on large-sample results. The time required to complete the 3-step bootstrap differed significantly from that of the bootstrap proposed by [Stoffer and Wall \(1991\)](#). On average, the 3-step bootstrap was 25% faster than the full bootstrap. Convergence of the optimisation algorithm required an average of eight steps in this relatively simple model. However, when dealing with larger and more complex state space models, employing the approximate bootstrap method instead of [Stoffer and Wall's](#) approach could result in substantial time savings.

---

<sup>4</sup>The analysis for  $k = 2$  produces similar results.

### 3.4.3 Simulation II

The tests considered in this section concern the transition matrix,  $T$ , in the state equation of the following state space model:

$$\begin{aligned} y_t &= Z\alpha_t + d + \varepsilon_t, & \varepsilon_t &\sim N(0, H) \\ (\alpha_{t+1} - a) &= T(\alpha_t - a) + \eta_t, & \eta_t &\sim N(0, Q). \end{aligned} \quad (3.19)$$

Equation (3.19) is simulated with  $T = 0$ ,  $Z = 2$ ,  $d = -0.5$ ,  $a = 0.85$ ,  $H = 0.13$ , and  $Q = 0.15$ . For simplicity we assume that  $Z$ ,  $H$  and  $Q$  are known and we estimate  $d$ ,  $T$ , and  $a$ . The aim is to test the hypothesis,

$$H_0 : T = 0$$

$$H_A : T \neq 0.$$

$\theta = [d, a : T]$  is the vector of parameters to be estimated. We first employ maximum likelihood to calculate the vector of unrestricted estimates,  $\hat{\theta} = [\hat{d}, \hat{a} : \hat{T}]$ . Next, we impose  $H_0$  and find the restricted estimates,  $\tilde{\theta} = [\tilde{d}, \tilde{a} : 0]$ . Optimisation is undertaken using the BFGS algorithm. After estimation, the Kalman filter is applied to each of the fitted models. The LR, Wald and Score statistics are formulated from the output of the filter as outlined in Section 3.3.

Rejection frequencies at a nominal level of 0.05 are calculated for the asymptotic tests, as well as for their bootstrap and approximate bootstrap analogues. Namely, we employ the bootstrap of [Stoffer and Wall \(1991\)](#) and the  $k$ -step bootstrap of DM with  $k = 3, 4, 5$  and 8 steps and we compute

the bootstrap LR, Wald and Score statistics. The experiments are carried out using 399 bootstrap iterations and 10 000 Monte Carlo replications. Additionally, for select sample sizes, the absolute difference is calculated between the  $p$ -value associated with the [Stoffer and Wall \(1991\)](#) method ( $\hat{p}^*$ ) and the  $p$ -value calculated from the  $k$ -step bootstrap ( $\tilde{p}^*$ ).

### **3.4.4 Simulation II: Results and Discussion**

The primary purpose of simulation II was to calculate the rejection probabilities for the classical LR, Wald and Score tests, as well as for their bootstrap equivalents. We were particularly interested in the adequacy of the  $k$ -step bootstraps relative to that of the full bootstrap of [Stoffer and Wall \(1991\)](#). Figures [3.1](#), [3.2](#) and [3.3](#) depict the rejection frequencies for the asymptotic and bootstrap tests.

(i) The LR test:

As shown in [Figure 3.1](#), the classical LR test markedly over-rejects when compared to the bootstrap tests. This effect is much more pronounced in small samples ( $n < 50$ ). Overall, as the number of steps,  $k$ , increases, the performance of the  $k$ -step bootstrap becomes closer to that of the [Stoffer and Wall](#) bootstrap. Although it outperforms the asymptotic test, the 3-step bootstrap still functions relatively poorly when compared to the other approximate bootstraps. By  $n = 90$ , all of the approximate bootstraps, with the exception of  $k = 3$ , have almost achieved the nominal rejection rate of 5 percent. While the rejection probability of the standard LR test clearly

### *The $k$ -Step Bootstrap In State Space Models*

improves as the series gets longer, this progress occurs considerably more slowly than for most of the bootstrap tests. At  $n = 90$  the LR test is still rejecting at around 9 percent, about 3 percentage points higher than most of the bootstrap tests.

(ii) The Wald test:

From the graph in Figure 3.2, we see that the outcome for the Wald test is similar to that of the LR test in that it over-rejects considerably in small samples. The performance of the associated bootstrap tests surpasses that of the asymptotic test for all sample sizes under consideration. Not surprisingly, in the case of the approximate bootstraps, the results improve as  $k$  increases. Much like in the LR scenario, the rejection probability of the classical Wald test approaches the nominal rejection frequency more slowly than those of the corresponding bootstrap tests. By  $n = 90$ , all of the bootstrap tests are rejecting at rates close to 0.05, in contrast to a rejection frequency of almost 0.20 exhibited by the standard Wald test.

(iii) The Score test:

Looking at Figure 3.3, what stands out is that, in contrast to the LR and Wald tests, the Score test under-rejects in small samples. However, as in the previous two cases, the error in rejection probability improves as the sample size grows. The bootstrap tests outperformed the asymptotic test but the rejection frequencies were, in comparison with the LR and Wald simulations, surprisingly erratic. We know, however, that the Score test is very sensitive

to the way in which the Fisher information is estimated<sup>5</sup>. It is plausible that alternative formulations of the Score statistic which are not based on the OPG estimate of the Fisher information could yield improved results. Also, it is possible that additional bootstrap and/or Monte Carlo iterations could potentially achieve a smoother series. However, limitations imposed by technological constraints prevented the exploration of increased replications.

We calculated the absolute differences between the  $p$ -values for the full bootstrap of [Stoffer and Wall](#) ( $\hat{p}^*$ ) and  $p$ -values from the  $k$ -step bootstrap ( $\check{p}^*$ ). Our findings exhibited some similarities with those of DM. In their study, they utilised a Tobit model to test the null hypothesis that a specific slope coefficient was zero against the alternative hypothesis that it was nonzero. They applied their  $k$ -step bootstrap to the LR, Score, and Wald statistics, among others, and determined that  $k=2$  steps were sufficient to replace a full bootstrap in the case of the LR tests, while two, three, or four steps were needed for the Score and Wald tests.

Table 3.2 displays the average absolute differences between  $\hat{p}^*$  and  $\check{p}^*$ . According to DM, an average absolute difference of around 0.001 can be considered negligibly small. In our case, it can be seen from the table that the Score and LR tests necessitate a minimum of three or four steps of the approximate bootstrap to replace the full bootstrap. However, the approximate bootstrap did not perform as well for the Wald statistic and eight steps were needed in this instance.

It is worth noting that in the Tobit model examined by DM, when performing the full bootstrap, the maximum number of steps required for the

---

<sup>5</sup>See [Godfrey \(1988\)](#).

optimisation algorithm to converge was, on average, 4.06 for the unrestricted estimates and 3.87 for restricted estimates. Therefore, taking  $k = 2$ ,  $k = 3$ , or  $k = 4$  steps resulted in only modest computational savings. In contrast, for our state space model, the maximum number of steps needed for convergence in the case of the unrestricted estimates was 12.03, while for the restricted estimates, it was 10.37.

### **3.5 Empirical Study**

[Newbold and Bos \(1985\)](#) investigated the relationship between the 3-month treasury bill rate ( $R_t$ ) and quarterly inflation ( $I_t$ ) in the United States. They utilised maximum likelihood to estimate the stochastic parameter regression model,

$$\begin{aligned} I_t &= \alpha + \beta_t R_t + v_t \\ (\beta_t - \beta) &= \varphi(\beta_{t-1} - \beta) + w_t, \end{aligned} \tag{3.20}$$

where  $v_t \sim iid(0, \sigma_v^2)$  and  $w_t \sim iid(0, \sigma_w^2)$ . [Stoffer and Wall \(1991\)](#) reproduced the work of [Newbold and Bos \(1985\)](#) for the period 1953/1 to 1965/2.<sup>6</sup> However, in addition to estimating (3.20) by maximum likelihood, they also used their methodology to calculate bootstrap sample means and bootstrap sample standard deviations, which they used to approximate the means and standard deviations of the sampling distributions of the maximum likelihood estimators.

---

<sup>6</sup>19xx/y denotes year 19xx, quarter y.

### *The $k$ -Step Bootstrap In State Space Models*

Newbold and Bos (1985) went on to test the hypothesis of a random coefficient model against the hypothesis of a stationary stochastic parameter regression model by applying the LR test, that is, for model (3.20) they tested,

$$H_0 : \varphi = 0 \tag{3.21}$$

$$H_1 : \varphi \neq 0. \tag{3.22}$$

They considered 6 subperiods of varying lengths between 1953/1 and 1980/2 and they tested the hypothesis for each period. Significance levels were set at 5%.

The goal of our experiment is to continue the investigation of Newbold and Bos (1985) by bootstrapping the LR statistics, repeating the hypothesis tests and noting whether we arrive at the same conclusions as they did. First, we perform maximum likelihood estimation of (3.20). We then apply the bootstrap of Stoffer and Wall (1991) and determine the means and standard deviations of the bootstrap distributions; we use these results to approximate the means and standard errors of the distributions of the ML estimators. These two steps are undertaken to ensure that we can reproduce the results of those authors. Next, we evaluate the bootstrap LR statistics using the methodology of Stoffer and Wall (1991), and find the corresponding  $p$ -values. We then go on to calculate the approximate bootstrap LR statistics and their related  $p$ -values by employing the  $k$ -step bootstrap of DM; we utilise  $k = 3, k = 4, k = 6, k = 8$  and  $k = 10$  steps. Our analysis uses 999 bootstrap replications and a  $p$ -value less than 0.05 is considered significant.

### **3.5.1 Empirical Study: Results and Discussion**

We replicated the ML estimation of (3.20) that was first performed by [Newbold and Bos \(1985\)](#) and later by [Stoffer and Wall \(1991\)](#). The time series considered covered the period 1953/1 to 1965/2. All three sets of results appear in Table 3.3 and are notably similar. Differences in the numerical optimisation routines employed can account for the minor discrepancies among the outcomes.<sup>7</sup>

The bootstrap estimates for (3.20) are presented in Table 3.4. Our findings are consistent with those of [Stoffer and Wall \(1991\)](#). The asymptotic standard errors are biased downwards in three of the five parameters under consideration when compared to the corresponding bootstrap values. In particular, for the estimates of the transition matrix,  $\varphi$ , and the standard deviations of the noise processes,  $\sigma_w$  and  $\sigma_v$ , the asymptotic standard errors are 60% to 70% the size of the bootstrap standard errors. Additionally, the maximum likelihood estimate of  $\varphi$  is noticeably higher (0.8414) than our bootstrap estimate (0.6034) and the bootstrap estimate of [Stoffer and Wall \(1991\)](#) (0.5897). Regarding the estimates of the constants  $\alpha$  and  $\beta$ , the asymptotic and bootstrap standard errors are comparable. These results highlight the widely-known fact that while asymptotic approximations are frequently used in statistical inference due to their simplicity, they may not always accurately represent the behaviour of estimators in finite samples. In this instance, the maximum likelihood estimation was performed using 50 data points.

Table 3.5 presents the LR statistics from [Newbold and Bos \(1985\)](#), along with their related  $p$ -values. Our bootstrap  $p$ -values, in conjunction with the

---

<sup>7</sup>In Table 3.3,  $\sigma_w$  and  $\sigma_v$  are the square roots of the corresponding variances calculated by [Newbold and Bos \(1985\)](#)



### *The $k$ -Step Bootstrap In State Space Models*

$k$ -step bootstrap  $p$ -values, are also reported. The notation  $p_k^*$  represents the  $p$ -value associated with the  $k$ -step bootstrap. We see from the column headed pVal, which contains the asymptotic  $p$ -values, that the LR test does not reject the null hypothesis for the first period ( 1953/1 to 1965/2) but rejects it from there onward. In comparison, the results in the column labelled  $p^*$ , which contains the full bootstrap  $p$ -values, indicate that the bootstrap LR test does not reject the null for the first two periods under consideration but rejects it thereafter. So, for the period 1953/1 to 1968/2, inferences based on the asymptotic and bootstrap  $p$ -values diverge. This result seems to support the literature which indicates that the finite sample distribution of the LR and other test statistics may differ somewhat from the chi-squared distribution that is usually employed <sup>8</sup>. The bootstrap distribution of a statistic is usually closer to its true small sample distribution than is the statistic's asymptotic distribution (Stoffer and Wall, 2004). This makes bootstrap inference in small samples more reliable than inference based on asymptotic theory.

In Table 3.5, columns 5 through 9 indicate that the  $k$ -step and standard bootstrap LR tests give rise to the same conclusions. For the time periods 1953/1 to 1965/2 and 1953/1 to 1968/2, when  $k = 4$ ,  $p_4^*$  differs from  $p^*$  by a maximum of 0.024. This suggests that when  $n = 50$  and  $n = 62$ , taking only 4 steps may be inadequate. At least six steps seem to be required for satisfactory results; when  $k = 6$ , for the aforementioned time periods,  $p_6^*$  diverges from  $p^*$  by at most 0.004. After 10 steps, the approximate bootstrap  $p$ -values deviate from the conventional ones by no more than 0.002, an amount which is, acceptably small. From the period 1953/1 to 1974/2 onward, that

---

<sup>8</sup>See Mittelhammer et al. (2000) for details.

is, for longer series, it is sufficient to take 3 steps. From column (3) we see that the average number of steps required for the convergence of the BFGS algorithm varies from 20 to 25 as the sample length increases from  $n = 50$  to  $n = 110$ . So, even if the approximate bootstrap is implemented with  $k = 10$ , it is still possible to cut the number of steps taken during optimisation by at least fifty percent.

### **3.6 Concluding Remarks**

This study set out to evaluate the effectiveness of the approximate bootstrap when it is applied to models in state space form. In addition, the paper argued that it is more expedient to work within the state space framework since the Kalman filter enables likelihood function evaluation via prediction error decomposition, facilitates estimation of any unknown parameters and furnishes components which allow for statistical testing. The filter update and prediction equations can be used to derive recursions for the exact gradient and consequently provide an estimate for the OPG form of the covariance matrix of parameter estimates. With all of this information, we are able to derive test statistics based solely on the output of the Kalman filter.

The empirical findings of this study suggest that in some models, when sample sizes are small, the approximate bootstrap with  $k = 3$  steps performs similarly to the RSSM bootstrap. Furthermore, regardless of the relatively poor performance of some of the bootstrap tests, the asymptotic results were even less satisfactory. Despite the limited number of models examined,

*The k-Step Bootstrap In State Space Models*

this observation can have significant practical implications. In the context of bootstrapping in the state space setting, implementing a small number of quasi-Newton steps can substantially reduce the computational cost associated with repeated nonlinear optimisation. In addition, the results are likely to be superior to those obtained via large-sample asymptotic approximations, particularly when working with small samples.

### 3.7 Tables and Figures

Table 3.1: Comparison of Asymptotic and Bootstrap Methods

	n=50	n=100	n=150	n=200
$\psi_{1,true}$	0.6698	0.6856	0.6928	0.6944
$\text{std}(\psi_{1,true})$	0.1168	0.0757	0.0620	0.0519
$\psi_{2,true}$	0.8184	0.8126	0.8051	0.8040
$\text{std}(\psi_{2,true})$	0.1097	0.0706	0.0543	0.0459
$\hat{\psi}_{1,ML}$	0.7726	0.7260	0.6316	0.6888
$\text{std}(\hat{\psi}_{1,ML})$	0.0546	0.0517	0.0743	0.0666
$\hat{\psi}_{2,ML}$	0.8901	0.8116	0.7733	0.7348
$\text{std}(\hat{\psi}_{2,ML})$	0.0476	0.0526	0.0587	0.0627
$\psi_{1, \text{stoffer\_wall}}$	0.7442	0.7076	0.6207	0.6828
$\text{std}(\psi_{1, \text{stoffer\_wall}})$	0.1031	0.0781	0.0664	0.0559
$\psi_{2, \text{stoffer\_wall}}$	0.9022	0.8220	0.7791	0.7373
$\text{std}(\psi_{2, \text{stoffer\_wall}})$	0.0956	0.0702	0.0584	0.0531
$\text{time}_{\text{stoffer\_wall}}(\text{hrs})$	41.77	64.57	91.67	116.24
$\psi_{1,RSSM}$	0.7512	0.7113	0.6231	0.6837
$\text{std}(\psi_{1,RSSM})$	0.0997	0.0753	0.0624	0.0549
$\psi_{2,RSSM}$	0.8951	0.8170	0.7761	0.7360
$\text{std}(\psi_{2,RSSM})$	0.0916	0.0734	0.0600	0.0530
$\text{time}_{RSSM}(\text{hrs})$	22.38	27.47	39.64	53.12
$\psi_{1,k=2}^*$	0.7458	0.7050	0.6214	0.6831
$\text{std}(\psi_{1,k=2}^*)$	0.1047	0.0747	0.0656	0.0554
$\psi_{2,k=2}^*$	0.9014	0.7143	0.7791	0.7372
$\text{std}(\psi_{2,k=2}^*)$	0.0891	0.0737	0.0574	0.0524
$\text{time}_{k=2}(\text{hrs})$	26.64	34.50	50.74	65.97
$\psi_{1,k=3}^*$	0.7452	0.7085	0.6208	0.6829
$\text{std}(\psi_{1,k=3}^*)$	0.1021	0.0778	0.0662	0.0558
$\psi_{2,k=3}^*$	0.9018	0.8219	0.7791	0.7373
$\text{std}(\psi_{2,k=3}^*)$	0.0894	0.0704	0.0584	0.0531
$\text{time}_{k=3}(\text{hrs})$	32.26	47.84	65.45	88.19

Table 3.2: Average Absolute Differences Between  $\hat{p}^*$  and  $\check{p}^*$

n	k	LR	Wald	Score
30	3	0.0165	0.0679	0.0017
	4	0.0026	0.0418	0.0005
	5	0.0011	0.0145	0.0001
	8	0.0003	0.0022	0.0001
60	3	0.0315	0.0202	0.0003
	4	0.0004	0.0017	0.0002
	5	0.0011	0.0101	0.0001
	8	0.0004	0.0004	0.0001
90	3	0.0233	0.0135	0.0003
	4	0.0013	0.0009	0.0007
	5	0.0009	0.0078	0.0001
	8	0.0009	0.001	0.0000

Table 3.3: Maximum Likelihood Estimates of Stochastic Parameter Regression Model (Equation 3.20), 1953/1 to 1965/2

Parameter	Simulation III		Stoffer and Wall		Newbold and Bos	
	Estimate	s.e. <sup>a</sup>	Estimate	s.e.	Estimate	s.e.
$\varphi$	.8414	.2005	.8414	.1997	.8414	.2122
$\beta$	.8584	.2784	.8584	.2776	.8584	.2591
$\alpha$	-.7714	.6466	-.7714	.6449	-.7715	.6033
$\sigma_w$	.1269	.0923	.1269	.0924	.1268	NA
$\sigma_v$	1.1306	.1424	1.1306	.1419	1.1306	NA

<sup>a</sup>In the table, s.e. represents the standard error.

Table 3.4: Estimates of Stochastic Parameter Regression Model  
(Equation 3.20), 1953/1 to 1965/2

Parameter	ML		Bootstrap (SW1991 <sup>a</sup> method)		Results taken from SW1991	
	Estimate	s.e. <sup>b</sup>	Estimate	Standard Deviation	Estimate	Standard Deviation
$\varphi$	.8414	.2005	.6034	.2565	.5897	.2775
$\beta$	.8584	.2784	.8432	.2770	.8416	.2737
$\alpha$	-.7714	.6466	-.7826	.6500	-.7652	.6315
$\sigma_w$	.1269	.0923	.1743	.1533	.1562	.1272
$\sigma_v$	1.1306	.1424	1.0643	.2278	1.0121	.2421

<sup>a</sup>SW1991 represents [Stoffer and Wall \(1991\)](#)

<sup>b</sup>In the table, s.e. represents the standard error.

Table 3.5: LR Statistics

1953/1 to	LR	Avg # steps	pVal	$p^*$	$p_3^*$	$p_4^*$	$p_6^*$	$p_8^*$	$p_{10}^*$
1965/2	3.7245	20	0.0536	0.0601	0.1001	0.0811	0.0602	0.0622	0.0621
1968/2	3.9815	20.5	0.046	0.1051	0.1431	0.1291	0.1101	0.1091	0.1071
1974/2	22.0688	23.5	0	0.001	0	0	0	0	0
1977/2	24.9097	24.5	0	0	0	0	0	0	0
1980/2	27.3407	25	0	0	.005	0.002	0	0	0

Figure 3.1: LR Test Rejection Frequencies

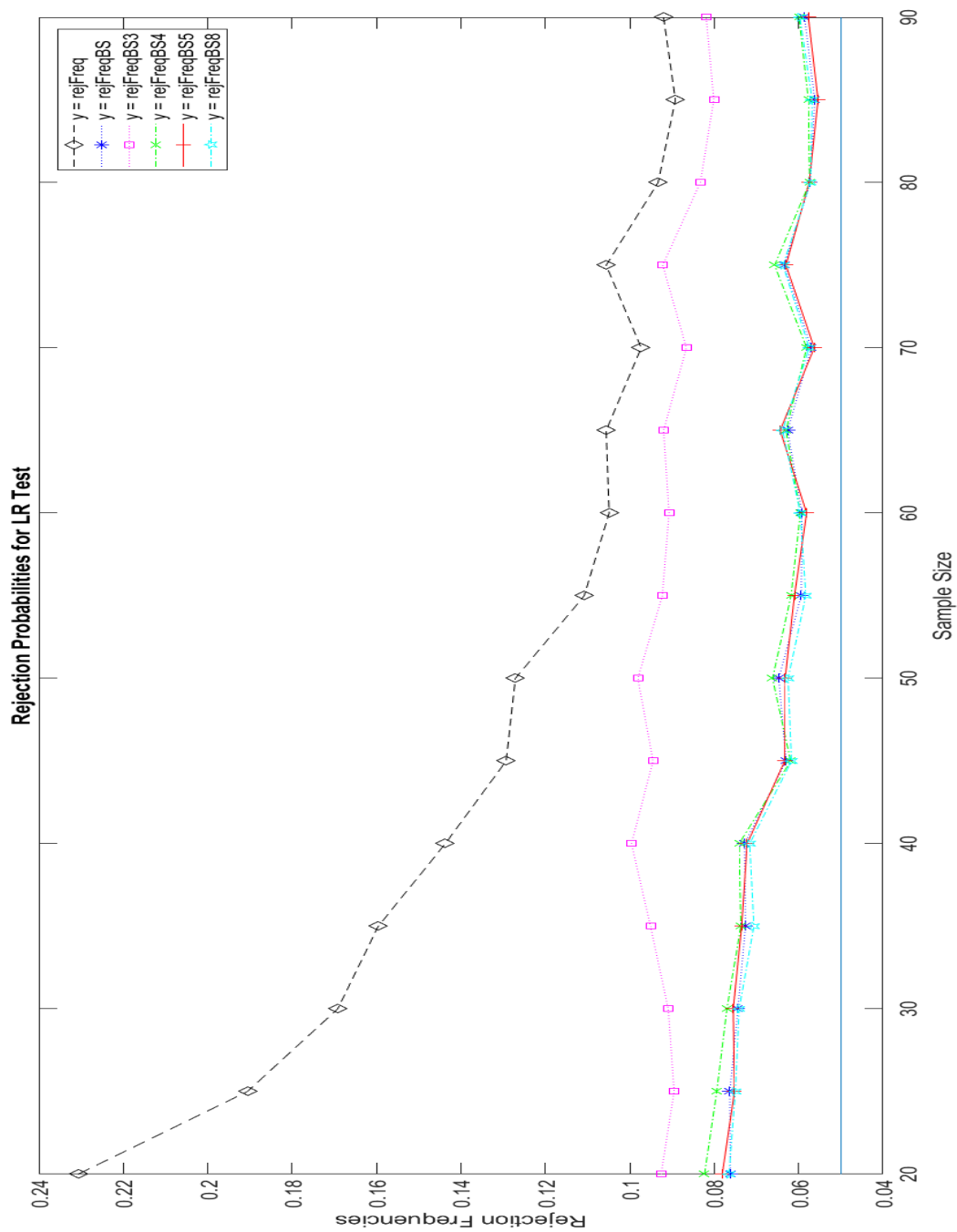




Figure 3.2: Wald Test Rejection Frequencies

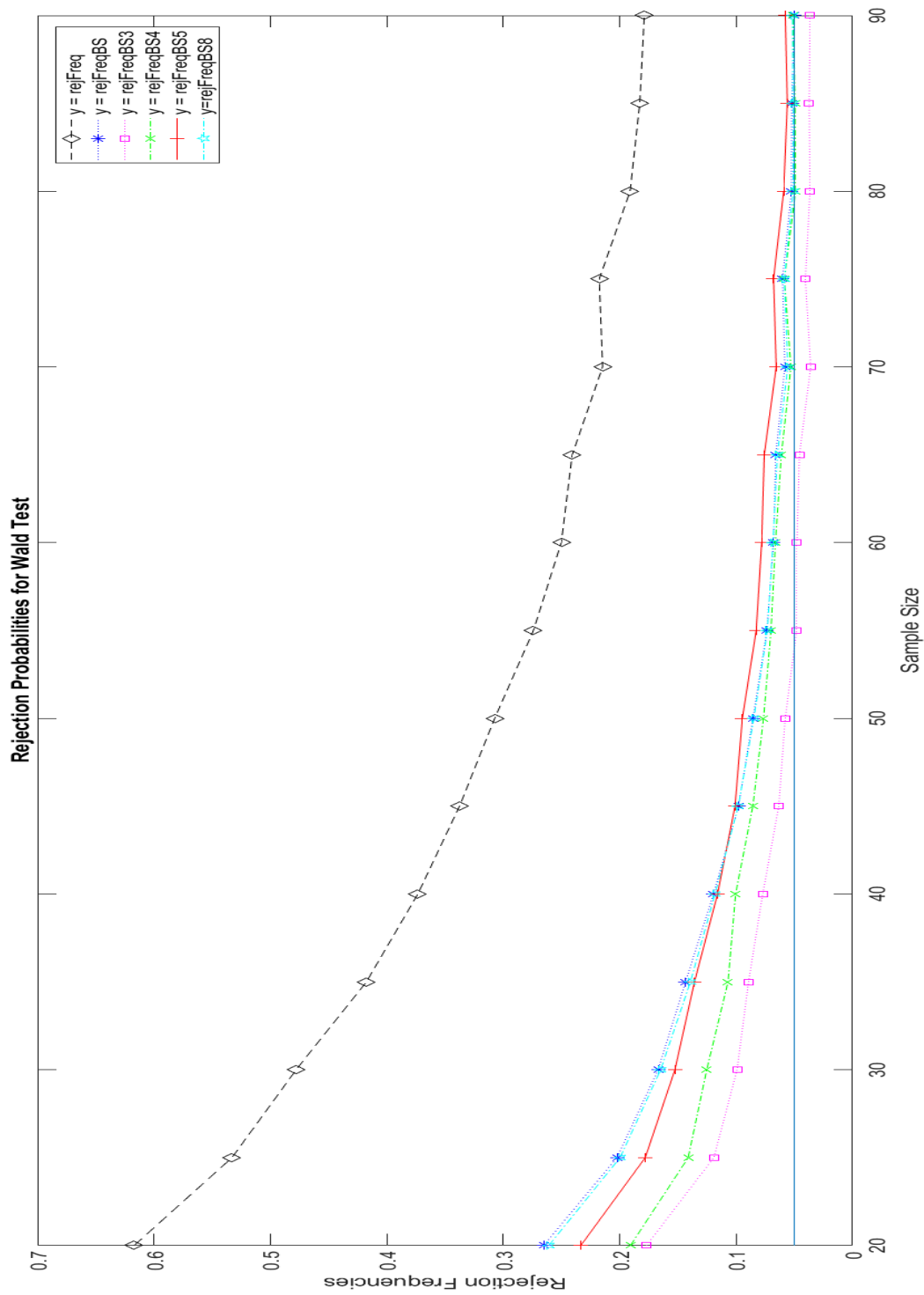
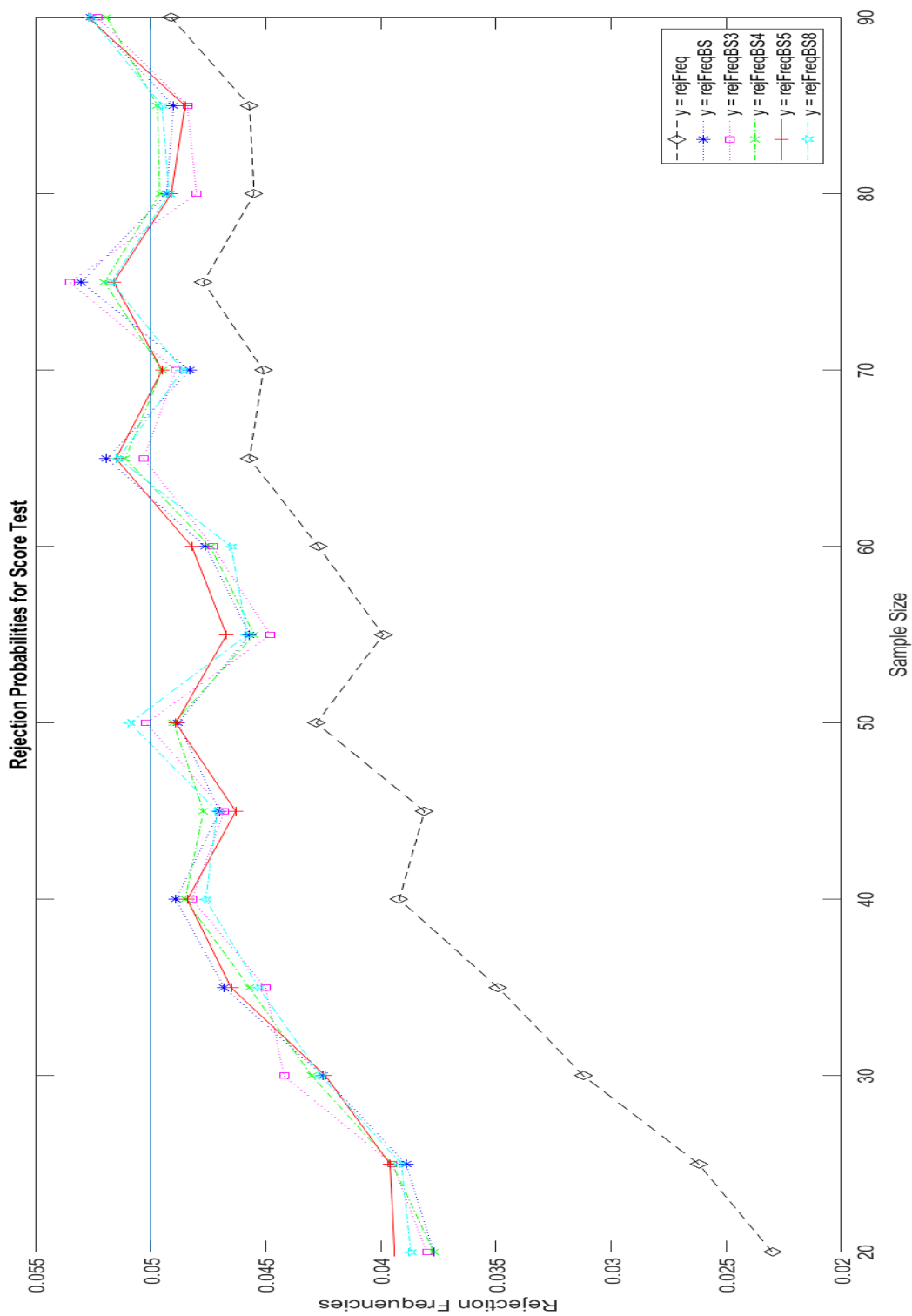


Figure 3.3: Score Test Rejection Frequencies



## Chapter 4

# Bootstrapping Multilevel Models: A State Space Approach

### 4.1 Introduction

In the time series literature, casting models in state space form and employing the Kalman filter is a well-known method of estimation. However, the effectiveness of the state space approach has received little attention within the context of clustered data. Multilevel models (MLMs) have applications across a wide range of disciplines. They possess a hierarchical structure and differ from standard regression models in that they allow for different error terms at each tier in the hierarchy. MLMs are generally estimated either by Full Information Maximum Likelihood (FIML) or by Restricted Maximum Likelihood (REML). Since FIML estimation does not take into account the loss in degrees of freedom which results from estimating the regression parameters, estimates of the variance components in the covariance matrix

of the vector of random effects tend to be biased downward. Some studies suggest that 2-tiered models with fewer than 50 clusters are particularly susceptible to these small-sample biases (McNeish, 2017). Conversely, in the REML approach, estimates of the variance components are calculated after the fixed effects have been removed and, therefore, suffer from less bias (Patterson and Thompson, 1971). Regardless, FIML estimation is still widely used since chi-squared difference tests can be employed to compare nested models with differing numbers of regression or variance parameters. With REML estimation, only differences in the variance entries can be compared. In this chapter, FIML estimation will be the maximum likelihood approach that is primarily employed; when REML estimation is used, it will be stated explicitly.

A few authors have explored the use of bootstrap methods in multilevel models. As the hierarchical structure of the data must be considered, resampling procedures for these models require modifications to the classic bootstrap techniques. Van der Leeden et al. (2008) and Goldstein (2011) discussed the extension of the parametric bootstrap and the residual bootstrap for use in multilevel analysis. Additionally, Modugno and Giannerini (2015) proposed a method for implementing the wild bootstrap in multilevel data sets. Our research showcases how employing the output of the Kalman filter can assist in facilitating the parametric and residual bootstrap procedures once the MLM has been converted into state space form.

State space models are powerful devices that allow for a multitude of operations within a unified framework. Gu et al. (2014) argued that estimating multilevel models in their state space configuration can offer computational

efficiency compared to alternative methods. Traditional FIML and REML approaches often require complex matrix operations at the cluster level, making the estimation process time-consuming. However, this is not the case when using the Kalman filter to perform optimisation. Despite these potential benefits, state space modelling for MLMs has not been widely embraced. This research aims to demonstrate the practicality of employing the state space formulation in hierarchical models.

The chapter proceeds as follows: Section 2 reviews the implementation of the bootstrap in MLMs; Section 3 illustrates how MLMs can be converted into state space format; Section 4 describes how Kalman filter output can be used in bootstrapping routines; a simulation is carried out in Section 5; an empirical study is undertaken in Section 6 and Section 7 concludes.

## **4.2 The Bootstrap in Multilevel Models**

When applying the bootstrap in MLMs, it is necessary to adapt traditional procedures to account for the hierarchical structure. Following, we provide a comprehensive review of the residual and parametric bootstraps within the multilevel setting.

Consider the 2-level model,

$$y_{ij} = \beta_{0j} + \beta_{1j}v_{ij} + e_{ij} \tag{4.1}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}w_{1j} + \eta_{0j} \tag{4.2}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}w_{1j} + \eta_{1j}, \tag{4.3}$$

for  $i = 1, \dots, N$ ,  $j = 1, \dots, J$ ,  $e_{ij} \sim N(0, \sigma_e^2)$  and  $\eta_j = \begin{bmatrix} \eta_{0j} \\ \eta_{1j} \end{bmatrix} \sim N(0, \Psi)$ .

The subscript  $j$  indicates a level-2 unit, that is, a cluster or group, and  $i$  represents a level-1 unit, namely one of the individual observations which are categorised into clusters. The model consists of  $J$  clusters, each containing  $n_j$  elements. There are  $N$  individual observations in total.  $y_{ij}$  is the dependent variable corresponding to the  $i^{th}$  element within the  $j^{th}$  group.  $v_{ij}$  and  $w_{1j}$  are level-1 and level-2 regressors, respectively. The random effects,  $\eta_j$ , and the level-1 errors,  $e_{ij}$ , are independent.

Variables at any given level are uncorrelated with error terms at the same level. Additionally, variables at a particular level are uncorrelated with the error terms on other levels. For example, the level-1 predictors,  $v_{ij}$ , are uncorrelated with the random effects,  $\eta_j$ , that is,  $E(v_{ij}\eta_j) = 0$ .

Substituting (4.2) and (4.3) into (4.1) gives,

$$y_{ij} = \gamma_{00} + \gamma_{01}w_{1j} + \eta_{0j} + \gamma_{10}v_{ij} + \gamma_{11}w_{1j}v_{ij} + \eta_{1j}v_{ij} + e_{ij} \quad (4.4)$$

$$= \begin{bmatrix} 1 & w_{1j} & v_{ij} & w_{1j}v_{ij} \end{bmatrix} \begin{bmatrix} \gamma_{00} \\ \gamma_{01} \\ \gamma_{10} \\ \gamma_{11} \end{bmatrix} + \begin{bmatrix} 1 & v_{ij} \end{bmatrix} \begin{bmatrix} \eta_{0j} \\ \eta_{1j} \end{bmatrix} + e_{ij}. \quad (4.5)$$

Expressing (4.5) in the standard mixed model format produces,

$$y_{ij} = x_{ij}\beta + z_{ij}\eta_j + e_{ij}, \quad (4.6)$$

where  $x_{ij} = \begin{bmatrix} 1 & w_{1j} & v_{ij} & w_{1j}v_{ij} \end{bmatrix}$ ,  $\beta = \begin{bmatrix} \gamma_{00} \\ \gamma_{01} \\ \gamma_{10} \\ \gamma_{11} \end{bmatrix}$  and  $z_{ij} = \begin{bmatrix} 1 & v_{ij} \end{bmatrix}$ .  $x_{ij}$  is

associated with the vector of fixed coefficients,  $\beta$ , and  $z_{ij}$  is connected to the random effects,  $\eta_j$ . We can also express the model in matrix form for each cluster,  $j$ :

$$Y_j = X_j\beta + Z_j\eta_j + e_j. \quad (4.7)$$

Here  $e_j \sim N(0, \sigma_e^2 I_{n_j})$ , where  $I_{n_j}$  is the  $n_j \times n_j$  identity matrix and  $\eta_j \sim N(0, \Psi)$ . Also,  $Y_j$ ,  $X_j$  and  $Z_j$  are matrices formed from vertically stacking the rows of  $y_{ij}$ ,  $x_{ij}$  and  $z_{ij}$ , for  $i = 1, \dots, n_j$ . Define,

$$V_j = \text{Var}(Y_j | X_j\beta) = Z_j\Psi Z_j' + \sigma_e^2 I_{n_j}. \quad (4.8)$$

Multilevel models can be estimated by full information maximum likelihood. If we assume that the residuals follow normal distributions, we can obtain the FIML estimates by maximising the log-likelihood function,

$$L = - \sum_j \left[ \frac{n_j}{2} \log(2\pi) + \frac{1}{2} \log |V_j| + \frac{1}{2} (Y_j - X_j\beta)' V_j^{-1} (Y_j - X_j\beta) \right], \quad (4.9)$$

with respect to all model parameters. Details of this method can be found in [Van der Leeden et al. \(1997\)](#). When the data is normally distributed and there are no missing observations, iterative generalised least squares (IGLS) and FIML estimation produce identical results.

MLMs are essentially extensions of random effects panel data models to

cases where there can be any number of levels in the data hierarchy (Rice et al., 2002). This enables the application of the methods discussed herein to random effects panels. It is worth noting, however, that these techniques are not suitable for fixed effects panels due to the correlation of the level-1 regressors with the random effects; this violates the assumption of uncorrelatedness discussed earlier in this chapter.

### 4.2.1 The Residual Bootstrap

The residual bootstrap for MLMs was introduced into the literature by Carpenter et al. (2003). There are no distributional assumptions on the error terms but the homogeneity of variances across groups is assumed.

In hierarchical models, both the level-1 and level-2 residuals can be biased. This means that empirical estimates of the variance, calculated from the raw residuals, do not line up with those estimated via full information maximum likelihood. To account for this, Carpenter et al. (2003) centres the residuals and multiplies them by the ratio of their estimated and empirical covariances. This adjustment results in estimates of the variance parameters that are numerically identical to the FIML estimates.

Consider the model described in equations (4.1) to (4.3). The methodology for transforming the level-2 residuals is as follows.

1. Use FIML estimation to find  $\hat{\Psi}$ ,  $\hat{V}_j$  and  $\hat{\beta}$  and calculate the level-2 residuals,  $\hat{\eta}_j, j = 1, \dots, J$ .<sup>1</sup>

$$\hat{\eta}_j = \hat{\Psi} Z_j' \hat{V}_j^{-1} (Y_j - X_j \hat{\beta}), \quad \forall j. \quad (4.10)$$

---

<sup>1</sup>See Jiang (1997) for the derivation of the formula for  $\hat{\eta}_j$ .



2. Centre the residuals:  $\tilde{\eta}_j = \hat{\eta}_j - \bar{\eta}_j$ , where  $\bar{\eta}_j$  is the arithmetic mean of  $\hat{\eta}_j$ .
3. Assume that each  $\eta_j$  is of dimension  $(q \times 1)$ . Create a  $J \times q$  matrix,  $\hat{Y}$ , from the  $J$  sets of centred residuals and calculate the empirical covariance matrix,  $S = \frac{\hat{Y}'\hat{Y}}{J}$ .
4. Let  $\hat{\Psi}$  be the FIML estimate of the covariance matrix of  $\eta_j, j = 1, \dots, J$ .
5. Find a transformation of the form  $\hat{Y}^\dagger = \hat{Y}A$  such that  $\frac{\hat{Y}^\dagger'\hat{Y}^\dagger}{J} = \frac{A'\hat{Y}'\hat{Y}A}{J} = A'SA = \hat{\Psi}$ . In essence, we want to rescale the residuals so that the transformed residuals have an empirical variance that is equal to the variance estimate obtained via FIML estimation. To construct  $A$ , we make use of the Cholesky decompositions of  $S$  and  $\hat{\Psi}$ :

$$S = L_S L_S'$$

$$\hat{\Psi} = L_{\hat{\Psi}} L_{\hat{\Psi}}',$$

where  $L_S$  and  $L_{\hat{\Psi}}$  are lower triangular matrices. We note that

$$\frac{L_{\hat{\Psi}} L_S^{-1} \hat{Y}' \hat{Y} (L_{\hat{\Psi}} L_S^{-1})'}{J} = L_{\hat{\Psi}} L_S^{-1} S L_S^{-1'} L_{\hat{\Psi}}' = L_{\hat{\Psi}} L_{\hat{\Psi}}' = \hat{\Psi}.$$

Hence we can choose  $A = (L_{\hat{\Psi}} L_S^{-1})'$ .

6. Construct  $\hat{Y}^\dagger = \hat{Y}A$ , the rows of which are the  $J$  sets of centred, rescaled level-2 residuals,  $\left\{ \hat{\eta}_j^\dagger \right\}_{j=1}^J$ .

An analogous procedure can be implemented for the level-1 residuals,  $\hat{e}_{ij}$ . Note that  $\hat{e}_{ij} = y_{ij} - x_{ij}\hat{\beta} - z_{ij}\hat{\eta}_j$ . The vector  $\{\hat{e}_{ij}\}_{i=1}^N$  can then be centred and

rescaled by the factor  $A = \frac{\hat{\sigma}_e}{sd(\hat{e}_{ij})}$  where  $\hat{\sigma}_e$  is the estimated standard error and  $sd(\hat{e}_{ij})$  is the empirical standard deviation. So,  $\{\hat{e}_{ij}^\dagger\}_{i=1}^N = \{\hat{e}_{ij}A\}_{i=1}^N$ .

The bootstrap proceeds as follows:

1. Estimate the model in equations (4.1) to (4.3) and find  $\{\hat{\beta}, \hat{\Psi}, \hat{\sigma}_e^2\}$ .
2. Resample  $N$  times, with replacement, from  $\{\hat{e}_{ij}^\dagger\}_{i=1}^N$  to form  $\{e_{ij}^*\}_{i=1}^N$ , the bootstrap sample of level-1 residuals.
3. Resample  $J$  times, with replacement, from the rows of  $\hat{\Upsilon}^\dagger$  to form the bootstrap sample of level-2 residuals,  $\{\eta_j^*\}_{j=1}^J$ .
4. Generate the bootstrap data set  $y_{ij}^* = x_{ij}\hat{\beta} + z_{ij}\eta_j^* + e_{ij}^*$ ,  $\forall i, j$  and use it to obtain bootstrap analogues,  $\{\beta^*, \Psi^*, \sigma_e^{2*}\}$ , of the FIML estimates in Step 1.
5. Repeat steps 2 to 4  $B$  times.
6. The estimates,  $\{\beta^*, \Psi^*, \sigma_e^{2*}\}_{b=1}^B$ , can be used to provide approximations for various measures such as bias, standard errors and confidence intervals.

### 4.2.2 The Parametric Bootstrap

Again, consider the model outlined in equations (4.1) to (4.3). The parametric bootstrap requires that both the model and the distributions of the error terms be correct. The bootstrap resamples are generated as follows:

1. Estimate the model to find  $\{\hat{\beta}, \hat{\Psi}, \hat{\sigma}_e^2\}$ .

2. Generate  $N$  level-1 bootstrap errors,  $e_{ij}^*$ , by drawing, with replacement, from the distribution,  $\hat{f}_e \sim N(0, \hat{\sigma}_e^2)$ .
3. Generate  $J$  level-2 bootstrap errors,  $\eta_j^*$ , by drawing, with replacement, from the distribution,  $\hat{f}_\eta \sim N(0, \hat{\Psi})$ .
4. Construct the bootstrap sample  $y_{ij}^* = x_{ij}\hat{\beta} + z_{ij}\eta_j^* + e_{ij}^*, \forall i, j$ , and estimate it to obtain  $\{\beta^*, \Psi^*, \sigma_e^{2*}\}$ , the bootstrap counterparts of the FIML estimates in Step 1.
5. Repeat steps 2 to 4  $B$  times.
6. The estimates,  $\{\beta^*, \Psi^*, \sigma_e^{2*}\}_{b=1}^B$ , can now be used for statistical inference.

## 4.3 State Space Models

### 4.3.1 Kalman Filter Equations for Multilevel Models

When dealing with MLMs in state space form, a second index,  $j$ , must be added to the state and measurement equations to account for the different clusters. The general state space format for handling hierarchical models is, <sup>2</sup>

$$y_{ij} = Z_{ij}\alpha_{ij} + B_{ij}u_{ij} + e_{ij} \quad \forall i, j \quad (4.11)$$

$$\alpha_{ij} = T_{ij}\alpha_{i-1,j} + C_{ij}u_{ij} + \xi_{ij}. \quad \forall i, j \quad (4.12)$$

---

<sup>2</sup>In previous chapters, we employed a state space model which is mathematically equivalent to the one presented here. The reason for adopting this particular format was to align with the work of [Gu et al. \(2014\)](#), which served as a key reference for this paper.

*Bootstrapping Multilevel Models: A State Space Approach*

where  $e_{ij} \sim N(0, H_{ij})$  and  $\xi_{ij} \sim N(0, Q_{ij})$ .<sup>3</sup> Also,

1. The initial state vector for each cluster  $j$ ,  $\alpha_{0j}$ , is distributed such that  $E(\alpha_{0j}) = a_{0|0,j}$  and  $\text{Var}(\alpha_{0j}) = P_{0|0,j}$ .
2. The error terms  $e_{ij}$  and  $\xi_{ij}$  are uncorrelated with each other, i.e.,  $E(e_{ij}\xi'_{sk}) = 0, \forall i, s$  and  $\forall j, k$ .
3. The error terms  $e_{ij}$  and  $\xi_{ij}$  are uncorrelated with the initial state, so  $E(e_{ij}\alpha'_{0j}) = 0$  and  $E(\xi_{ij}\alpha'_{0j}) = 0, \forall i, j$ .

Likewise, another index is included in the Kalman Filter equations:

$$v_{ij} = y_{ij} - Z_{ij}a_{i|i-1,j} - B_{ij}u_{ij} \quad (4.13)$$

$$F_{ij} = Z_{ij}P_{i|i-1,j}Z'_{ij} + H_{ij} \quad (4.14)$$

$$\left. \begin{aligned} a_{i|i-1,j} &= T_{ij}a_{i-1,j} + C_{ij}u_{ij} \\ P_{i|i-1,j} &= T_{ij}P_{i-1,j}T'_{ij} + Q_{ij} \end{aligned} \right\} = \text{prediction equations} \quad (4.15)$$

$$\left. \begin{aligned} a_{i|i,j} &= a_{i|i-1,j} + P_{i|i-1,j}Z'_{ij}F_{ij}^{-1}v_{ij} \\ P_{i|i,j} &= P_{i|i-1,j} - P_{i|i-1,j}Z'_{ij}F_{ij}^{-1}Z_{ij}P_{i|i-1,j} \end{aligned} \right\} = \text{updating equations.}^4 \quad (4.16)$$

<sup>3</sup>The exact relationship between the quantities that appear in (4.11) and (4.12) and the quantities that appear in the general MLM in (4.1) are given in subsection 4.3.3.

<sup>4</sup>A word on notation: Generally, with two subscripts, we use  $k_{ij}$  to refer to the  $i^{\text{th}}$  element in the  $j^{\text{th}}$  block. However, when the subscripts are more complex, as in the case of  $k_{i-1,j}$ , we use a comma between the two subscripts for clarity. Also, for improved readability, terms of the form  $X_{i-1|i-1,j}$  will be expressed as  $X_{i-1,j}$ .

### 4.3.2 Parameter Estimation

For each cluster,  $j$ , the Kalman filter is initialised with the mean and variance of the initial state vector,  $a_{0|0,j}$  and  $P_{0|0,j}$ , respectively<sup>5</sup>, and then executed. The output of the filter is used to construct the prediction error decomposition (p.e.d) form of the log-likelihood. In MLMs, this is obtained by summing the individual log-likelihoods over the  $J$  clusters,

$$L_{mlm}(\theta) = \sum_{j=1}^J \sum_{i=1}^{n_j} \left[ -\frac{n_j}{2} \log 2\pi - \frac{1}{2} \sum_{i=1}^{n_j} \left\{ \log |F_{ij}(\theta)| + v'_{ij}(\theta) F_{ij}^{-1}(\theta) v_{ij}(\theta) \right\} \right]. \quad (4.17)$$

This function can then be maximised using the preferred optimisation procedure. Recall that  $n_j$  represents the number of elements in each cluster  $j$  and that  $N = \sum_{j=1}^J n_j$  is the total number of elements in the model.

### 4.3.3 Casting Multilevel Models In State Space Form

Consider the following random intercept/random slope model,

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{ij} \quad (4.18)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}w_{1j} + \eta_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}w_{1j} + \eta_{1j}, \quad \forall i, j,$$

where  $e_{ij} \sim N(0, \sigma_e^2)$  and  $\eta_j = \begin{bmatrix} \eta_{0j} \\ \eta_{1j} \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{bmatrix} \right)$ .

---

<sup>5</sup>See [Durbin and Koopman \(2012\)](#) for details.

*Bootstrapping Multilevel Models: A State Space Approach*

Substitute  $\beta_{0j}$  and  $\beta_{1j}$  into  $y_{ij}$  to get,

$$y_{ij} = [\gamma_{00} + \gamma_{01}w_{1j} + \gamma_{10}x_{ij} + \gamma_{11}w_{1j}x_{ij}] + [\eta_{0j} + \eta_{1j}x_{ij}] + e_{ij} \quad (4.19)$$

Separate the right hand side of (4.19) into the fixed and random effects, along with their coefficient matrices.

$$y_{ij} = \begin{bmatrix} \gamma_{00} & \gamma_{01} & \gamma_{10} & \gamma_{11} \end{bmatrix} \begin{bmatrix} 1 \\ w_{1j} \\ x_{ij} \\ w_{1j}x_{ij} \end{bmatrix} + \begin{bmatrix} \eta_{0j} & \eta_{1j} \end{bmatrix} \begin{bmatrix} 1 \\ x_{ij} \end{bmatrix} \quad (4.20)$$

Now, rearrange the terms in (4.20) to match the format of the measurement equation, (4.11). That is,

$$y_{ij} = \begin{bmatrix} 1 & x_{ij} \end{bmatrix} \begin{bmatrix} \eta_{0j} \\ \eta_{1j} \end{bmatrix} + \begin{bmatrix} \gamma_{00} & \gamma_{01} & \gamma_{10} & \gamma_{11} \end{bmatrix} \begin{bmatrix} 1 \\ w_{1j} \\ x_{ij} \\ w_{1j}x_{ij} \end{bmatrix} + e_{ij}. \quad (4.21)$$

Since the random effects,  $\begin{bmatrix} \eta_{0j} \\ \eta_{1j} \end{bmatrix}$ , do not change within a given group  $j$ , the state equation (4.12) is given by,

$$\begin{bmatrix} \eta_{0j} \\ \eta_{1j} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_{0j} \\ \eta_{1j} \end{bmatrix} \quad (4.22)$$

Referring to equations (4.11) and (4.12), the matrices are:  $Z_{ij} = \begin{bmatrix} 1 & x_{ij} \end{bmatrix}$ ,

$$\alpha_{ij} = \begin{bmatrix} \eta_{0j} \\ \eta_{1j} \end{bmatrix}, B_{ij} = \begin{bmatrix} \gamma_{00} & \gamma_{01} & \gamma_{10} & \gamma_{11} \end{bmatrix}, u_{ij} = \begin{bmatrix} 1 \\ w_{1j} \\ x_{ij} \\ w_{1j}x_{ij} \end{bmatrix}, T_{ij} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, C_{ij} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, Q_{ij} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \text{ and } H_{ij} = \sigma_e^2.$$

Given that the Kalman filter is initialised using the mean and variance of the initial state vectors, we start the recursion with,  $a_{0|0,j} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$  and

$$P_{0|0,j} = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{bmatrix}, \forall j.$$

#### 4.3.4 Example 1

Consider the random intercept/random slope model,

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \epsilon_{ij} \tag{4.23}$$

$$\beta_{0j} = \gamma_0 + \eta_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}w_{1j} + \eta_{1j}.$$

Here,  $i = 1, \dots, N$ ,  $j = 1, \dots, J$ ,  $\epsilon_{ij} \sim N(0, \sigma_e^2)$  and  $\begin{bmatrix} \eta_{0j} \\ \eta_{1j} \end{bmatrix} \sim$

*Bootstrapping Multilevel Models: A State Space Approach*

$N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{bmatrix} \right)$ . Substituting  $\beta_{0j}$  and  $\beta_{1j}$  into  $y_{ij}$  gives,

$$y_{ij} = (\gamma_0 + \eta_{0j}) + (\gamma_{10} + \gamma_{11}w_{1j} + \eta_{1j})x_{ij} + \epsilon_{ij}. \quad (4.24)$$

Expressing (4.24) in state space form yields the following:

$$y_{ij} = \begin{bmatrix} 1 & x_{ij} \end{bmatrix} \begin{bmatrix} \eta_{0j} \\ \eta_{1j} \end{bmatrix} + \begin{bmatrix} \gamma_0 & \gamma_{10} & \gamma_{11} \end{bmatrix} \begin{bmatrix} 1 \\ x_{ij} \\ w_{1j}x_{ij} \end{bmatrix} + \epsilon_{ij} \quad (4.25)$$

$$\begin{bmatrix} \eta_{0j} \\ \eta_{1j} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_{0j} \\ \eta_{1j} \end{bmatrix} \quad (4.26)$$

So,  $Z_{ij} = \begin{bmatrix} 1 & x_{ij} \end{bmatrix}$ ,  $\alpha_{ij} = \begin{bmatrix} \eta_{0j} \\ \eta_{1j} \end{bmatrix}$ ,  $B_{ij} = \begin{bmatrix} \gamma_0 & \gamma_{10} & \gamma_{11} \end{bmatrix}$ ,  $u_{ij} = \begin{bmatrix} 1 \\ x_{ij} \\ w_{1j}x_{ij} \end{bmatrix}$ ,  $T_{ij} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ,  $C_{ij} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ ,  $Q_{ij} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ ,  $e_{ij} = \epsilon_{ij}$ ,  $H_{ij} = \sigma_\epsilon^2$ ,  $a_{0|0,j} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$  and  $P_{0|0,j} = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{bmatrix}$ .



## 4.4 Bootstrapping MLMs in State Space Form

### 4.4.1 Residual Bootstrap

Consider the MLM in equation (4.18). Let  $\theta = [\sigma_e^2, \gamma_0, \gamma_{10}, \gamma_{11}, \tau_{00}, \tau_{01}, \tau_{11}]$  be the vector of parameters to be estimated. The steps for performing the residual bootstrap in the state space framework are outlined below:

1. Cast the MLM in state space form and initialise the Kalman filter as illustrated in the previous section.
2. Use an optimisation algorithm to maximise the log-likelihood (4.17) and find  $\hat{\theta} = \arg \max_{\theta} L_{mlm}(\theta)$ . Run the Kalman filter with  $\theta = \hat{\theta}$ .
3. Obtain the level-2 residuals,  $\{\hat{\eta}_j\}_{j=1}^J$ . These can conveniently be procured from the Kalman filter output. In the previous section we saw that the random effects are modelled as the random state vector in the Kalman filter equations. Recall that there are  $n_j$  elements in each of the  $J$  blocks. The estimate of the random effect ( $\hat{\eta}_j$ ) for each block  $j$  is the filtered state estimate,  $a_{n_j|n_j,j}$ . Specifically, the entries in each block are numbered from 1 to  $n_j$ , for  $j = 1, \dots, J$ . The random effect for block  $j$  is estimated at the terminal element in that block,  $n_j$ , because that estimate is based on all of the data points in the block. For example, assume that the first cluster has 50 elements, i.e,  $j = 1$  and  $n_1 = 50$ . The estimate of the associated random effect,  $\hat{\eta}_1$ , is given by  $a_{n_1|n_1,1} = a_{50|50,1}$ . Figure 4.1 demonstrates how the filtered state estimates (shown in bold) for each block are collected to form the vector of level-2 residuals. Notice

that these values are automatically produced by the updating equations (4.16) so no additional calculations are required.

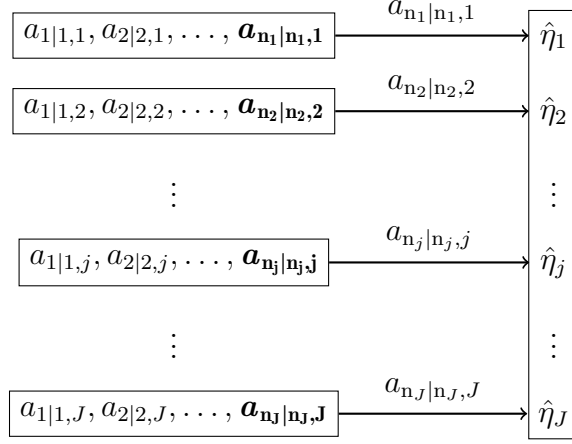
4. Rescale the level-2 residuals acquired in step 3 as described in subsection (4.2.1) to find  $\left\{\hat{\eta}_j^\dagger\right\}_{j=1}^J$ .
5. Recover the level-1 residuals,  $\hat{e}_{ij} = y_{ij} - Z_{ij}\hat{\eta}_j - \hat{B}_{ij}u_j, \forall i, j$ , and rescale them as demonstrated in subsection (4.2.1) to produce  $\left\{\hat{e}_{ij}^\dagger\right\}_{i=1}^N$ .
6. Resample  $J$  times from the rescaled level-2 residuals, with replacement, to generate the level-2 bootstrap errors,  $\left\{\eta_j^*\right\}_{j=1}^J$ .
7. Resample  $N$  times from the rescaled level-1 residuals, with replacement, to generate the level-1 bootstrap errors,  $\left\{e_{ij}^*\right\}_{i=1}^N$ .
8. Create the bootstrap sample  $y_{ij}^* = Z_{ij}\eta_j^* + \hat{B}_{ij}u_j + e_{ij}^*$ .
9. Compute  $\theta_b^*$  by applying the definition of  $\hat{\theta}$  to the  $b^{\text{th}}$  bootstrap dataset, in place of the original data.
10. Repeat steps 6 to 9  $B$  times.

These estimates,  $\left\{\theta_b^*\right\}_{b=1}^B$ , can be used to approximate standard errors, biases, confidence intervals, and so on.

#### 4.4.2 Parametric Bootstrap

Again, consider the model in equation (4.18) and let  $\theta = [\sigma_e^2, \gamma_0, \gamma_{10}, \gamma_{11}, \tau_{00}, \tau_{01}, \tau_{11}]$  be the vector of parameters to be estimated. To implement the parametric bootstrap we,

Figure 4.1: Estimates of the Random Effects



1. Express the model in its state space format then initialise and run the Kalman filter.
2. Construct the p.e.d form of the log-likelihood, (4.17), and obtain the estimate,  $\hat{\theta} = [\hat{\sigma}_e^2, \hat{\gamma}_0, \hat{\gamma}_{10}, \hat{\gamma}_{11}, \hat{\tau}_{00}, \hat{\tau}_{01}, \hat{\tau}_{11}]$ .
3. Draw a sample,  $\{\eta_j^*\}_{j=1}^J$ , from a multivariate normal distribution with mean zero and covariance matrix  $\hat{\Psi} = \begin{bmatrix} \hat{\tau}_{00} & \hat{\tau}_{01} \\ \hat{\tau}_{01} & \hat{\tau}_{11} \end{bmatrix}$ .
4. Draw a sample,  $\{e_{ij}^*\}_{i=1}^N$ , from a normal distribution with mean zero and variance  $\hat{H}_{ij} = \hat{\sigma}_e^2$ .
5. Generate the bootstrap responses,  $y_{ij}^* = Z_{ij}\eta_j^* + \hat{B}_{ij}u_j + e_{ij}^*$ ,  $\forall i, j$ .
6. Compute  $\theta_b^*$ , the FIML estimate of  $\theta$ , based on the  $b^{th}$  bootstrap sample.
7. Repeat steps 3 to 6  $B$  times to obtain  $\{\theta_b^*\}_{b=1}^B$ , which can be used to facilitate statistical inference.

Bootstrapping within the state space environment is very efficient. Both FIML and IGLS require the inversion of  $V_j$ , which is an  $n_j \times n_j$  matrix<sup>6</sup>. With the standard Kalman filter we only need to invert  $F_{ij}$  whose dimensions are  $\Pi \times \Pi$ , where  $\Pi$  is the number of dependent variables in the model<sup>7</sup>. Typically,  $n_j$  is substantially larger than  $\Pi$ . Notice that  $F_{ij}$  is independent of  $n_j$  while  $V_j$  is not. Therefore, increases in  $n_j$ , the number of entries per cluster, will rapidly increase the dimension of  $V_j$ , making the matrix inversion in FIML and IGLS more burdensome and time-consuming. More specifically, the standard matrix inversion algorithm has a time complexity of  $O(r^3)$ , where  $r$  is the dimension of the matrix being inverted. This means that the computation time increases rapidly with the size of the matrix, growing cubically with the number of components. For example, in a MLM with one dependent variable and ten elements per cluster, the time required for matrix inversion using state space estimation is  $O(1)$ , whereas the time for matrix inversion using IGLS is  $O(1000)$ . Given that bootstrapping generally involves hundreds of matrix inversions, opting for the state space approach can markedly reduce computational costs.

## 4.5 Bootstrap Bias-Correction

### 4.5.1 Simulation

The objective of this exercise is to apply the bootstrap methodologies developed in section 4.4 and bias-correct the FIML estimates of the variance

---

<sup>6</sup>See equation (4.8).

<sup>7</sup>See equation (4.14).

components of a multilevel model.

Consider the following model:

$$\begin{aligned} y_{ij} &= \beta_{0j} + \beta_{1j}x_{ij} + e_{ij} \\ \beta_{0j} &= \gamma_{00} + \gamma_{01}w_{1j} + \eta_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}w_{1j} + \eta_{1j}, \forall i, j, \end{aligned} \tag{4.27}$$

where  $e_{ij} \sim N(0, \sigma_e^2)$  and  $\eta_j = \begin{bmatrix} \eta_{0j} \\ \eta_{1j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{bmatrix}\right)$ .

The model is expressed in state space form as shown in (4.28).

$$y_{ij} = \begin{bmatrix} 1 & x_{ij} \end{bmatrix} \begin{bmatrix} \eta_{0j} \\ \eta_{1j} \end{bmatrix} + \begin{bmatrix} \gamma_{00} & \gamma_{01} & \gamma_{10} & \gamma_{11} \end{bmatrix} \begin{bmatrix} 1 \\ w_{1j} \\ x_{ij} \\ w_{1j}x_{ij} \end{bmatrix} + e_{ij} \tag{4.28}$$

$$\begin{bmatrix} \eta_{0j} \\ \eta_{1j} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_{0j} \\ \eta_{1j} \end{bmatrix}.$$

The filter is initialised with

$$a_{0|0,j} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \text{ for } j = 1, \dots, J \tag{4.29}$$

$$P_{0|0,j} = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{bmatrix}, \text{ for } j = 1, \dots, J. \quad (4.30)$$

Let  $\theta = [\gamma_{00}, \gamma_{01}, \gamma_{10}, \gamma_{11}, \sigma_e^2, \tau_{00}, \tau_{01}, \tau_{11}]$  be the vector of parameters to be estimated. We use the BFGS algorithm to maximise the log-likelihood in equation (4.17) and find  $\hat{\theta}$ . As was mentioned in the introduction of this chapter, FIML estimates of the variance components tend to be biased downward in small samples. After estimating  $\tau_{00}, \tau_{01}$  and  $\tau_{11}$ , we use the residual and parametric bootstraps to help correct for this bias. The bias-corrected FIML estimates are then compared with the true parameter values. Additionally, REML<sup>8</sup> estimates are evaluated for comparison since they are less susceptible to severe bias.

The model is simulated with  $\gamma_{00} = 3$ ,  $\gamma_{01} = 4.8$ ,  $\gamma_{10} = 5.0$ ,  $\gamma_{11} = 5.2$ ,  $\sigma_e^2 = 2$ ,  $\tau_{00} = 2$ ,  $\tau_{01} = 0.5$  and  $\tau_{11} = 2$ . The elements  $x_{ij}$  and  $w_{ij}$  are the  $ij^{th}$  entries in  $N \times 1$  and  $J \times 1$  vectors of standard normal variables, respectively. Simulations are carried out for varying numbers of groups,  $J$ , all with group size  $n_j = 100$ . Specifically, the number of groups we analyse are:

$J$	$n_j$	$N=J \times n_j$
10	100	1000
15	100	1500
20	100	2000

In hierarchical models, it is the level-2 sample size ( $J$ ) that is generally responsible for determining whether the model will suffer from small-sample

---

<sup>8</sup>All REML estimates in this document are evaluated using restricted iterative generalised least squares (RIGLS). See Goldstein (1989).

problems and it is recommended that  $J$  be larger than 20 in order to have relatively unbiased estimates (Snijders and Bosker, 1993). Additionally, Snijders and Bosker (2011) advises against using MLMs if  $J$  is smaller than 10. These two sources informed our choices for  $J$ . In this simulation, despite the fact that there are  $n_j = 100$  elements in each block, there are relatively few blocks, with  $J$  ranging from only  $J = 10$  to  $J = 20$ . These data sets would therefore be considered small.  $B = 999$  bootstraps are used in this analysis. One thousand Monte Carlo rounds are conducted and the averages of the parameter estimates are calculated. The results are reported in Tables 4.1 through 4.6.

### 4.5.2 Simulation: Results and Discussion

In tables 4.1 through 4.6, from left to right, the columns display the Monte Carlo averages of the REML estimates ( $\hat{\theta}_{REML}$ ), the FIML estimates ( $\hat{\theta}_{FIML}$ ), the bootstrap estimates ( $\theta^*$ ), the bias-corrected FIML estimates ( $\hat{\theta}_{FIMLBC}$ ), the proportional bias of the FIML estimates and the proportional bias of the bias-corrected FIML estimates.

The bias of the FIML estimator is given by,

$$\text{Bias}_{true} = E \left[ \hat{\theta}_{FIML} \right] - \theta_0,$$

where  $\theta_0$  is the underlying true parameter value. The corresponding bootstrap bias is given by

$$\text{Bias}_{boot} = \bar{\theta}^* - \hat{\theta}_{FIML},$$

where  $\bar{\theta}^* = \frac{1}{B} \sum_{b=1}^B \theta_b^*$  and  $\theta_b^*$  is the estimate of  $\theta$  from the  $b^{th}$  bootstrap sample. The bias-corrected FIML estimator is

$$\begin{aligned}\hat{\theta}_{FIML-BC} &= \hat{\theta}_{FIML} - \text{Bias}_{boot} \\ &= 2\hat{\theta}_{FIML} - \bar{\theta}^*.\end{aligned}\tag{4.31}$$

The results of the study highlight that the bias in the FIML estimates shrinks as the number of clusters increases. Consider  $\tau_{00}$ , the first element on the leading diagonal in the initial state covariance matrix in (4.30). When  $J=10$ , the FIML estimate of  $\tau_{00}$  is 1.546, 22.7% below its true value, as shown in Table 4.1. However, at  $J = 20$ , the FIML estimate for  $\tau_{00}$  is 1.818, which is only 9.1% below the actual value, as seen in Table 4.3.

By comparison, for the parametric bootstrap, at  $J=10$ , the bias-corrected FIML estimate of  $\tau_{00}$  is 1.860, 7.0% below the true value (Table 4.1) and when the number of clusters is increased to 20, the bias-corrected figure differs from the real value by 0.1% (Table 4.3).

When we analyse the effectiveness of the residual bootstrap, we notice that it closely mirrors the performance of the parametric bootstrap. Again, consider  $\tau_{00}$ . For  $J=10$  clusters, the residual bootstrap-bias-corrected FIML estimate is 1.811, 9.5% lower than the true value of  $\tau_{00} = 2$ , as depicted in Table 4.4. By  $J=20$ , the bias-corrected estimate differs from the true value by only 0.1%, as illustrated in Table 4.6.

For both bootstrap schemes, undertaking the same analysis for the remaining parameters,  $\tau_{01}$  and  $\tau_{11}$ , yields comparable findings.

Clearly, when level-2 sample sizes are small, bias-correction of the FIML



outcomes can be helpful. When  $J = 20$ , the bias-corrected estimates are virtually on par with the REML figures. However, in very small samples, as is the case for  $J = 10$ , performing a single round of bias-correction still produces estimates with greater biases than those observed with REML estimation. In these instances, it is likely that the bias-correction could be improved by implementing an iterative bootstrap procedure. For the interested reader, many authors, including [Martin \(1992\)](#), [Davidson and MacKinnon \(2007\)](#) and [Ouyssse \(2013\)](#) describe, in detail, procedures for implementing the double bootstrap. For the motivated researcher, [Davidson and Trokić \(2020\)](#) provides a triple bootstrap technique.

## 4.6 Empirical Study

The present research builds upon the findings of [Fears et al. \(1996\)](#). Those authors examined the estrone levels of 5 postmenopausal women. Sixteen samples were taken from each participant and the estrone levels were recorded for each sample. Table 4.7 contains that data. They estimated the model,

$$\begin{aligned} y_{ij} &= \beta_{0j} + \varepsilon_{ij} \\ \beta_{0j} &= \mu + \mathbf{a}_j, \end{aligned} \tag{4.32}$$

for  $i = 1, \dots, 16$  and  $j = 1, \dots, 5$ . In (4.32),  $y_{ij}$  represents the logarithm (base 10) of the estrone measurement for the  $i^{th}$  sample from the  $j^{th}$  woman and  $\mu$  is the overall mean. The error terms,  $\mathbf{a}_j$  and  $\varepsilon_{ij}$ , are assumed to be independent, normally distributed random variables with zero means and with variances of  $\sigma_{\mathbf{a}}^2$  and  $\sigma_{\varepsilon}^2$ , respectively. [Fears et al. \(1996\)](#) estimated  $\sigma_{\mathbf{a}}^2$  and  $\sigma_{\varepsilon}^2$  by employing

the maximum likelihood option in the SAS procedure, VARCOMP, which computes FIML estimates. They constructed a Wald statistic to test the hypothesis:

$$H_0 : \sigma_{\mathbf{a}}^2 = 0 \quad (4.33)$$

$$H_A : \sigma_{\mathbf{a}}^2 > 0. \quad (4.34)$$

The null hypothesis was not rejected, indicating homogeneity in estrone levels among the study participants. Following this, the authors generated the exact bootstrap distribution of  $\hat{\sigma}_{\mathbf{a}}^2$ , the FIML estimator of the variance of the random effects  $\mathbf{a}_j$ , and noted that the distribution deviated significantly from normality.

In order to estimate it, we first transform the model of [Fears et al. \(1996\)](#) into the state space format. We rewrite (4.32) as

$$\begin{aligned} y_{ij} &= \begin{bmatrix} 1 \end{bmatrix} \mathbf{a}_j + \mu \begin{bmatrix} 1 \end{bmatrix} + \varepsilon_{ij} \\ \mathbf{a}_j &= \begin{bmatrix} 1 \end{bmatrix} \mathbf{a}_j, \end{aligned} \quad (4.35)$$

for  $i = 1, \dots, 16$ ,  $j = 1, \dots, 5$ ,  $\varepsilon_{ij} \sim N(0, \sigma_{\varepsilon}^2)$  and  $\mathbf{a}_j \sim N(0, \sigma_{\mathbf{a}}^2)$ . In terms of the matrices we defined in (4.11) and (4.12),  $Z_{ij} = 1$ ,  $\alpha_{ij} = \mathbf{a}_j$ ,  $B_{ij} = \mu$ ,  $u_{ij} = 1$ ,  $T_{ij} = 1$ ,  $c_{ij} = 0$ ,  $\xi_{ij} = 0$ ,  $Q_{ij} = 0$ ,  $H_{ij} = \sigma_{\mathbf{a}}^2$  and  $e_{ij} = \varepsilon_{ij}$ . The Kalman filter is initialised with  $a_{0|0,j} = 0$  and  $P_{0|0,j} = \sigma_{\mathbf{a}}^2$  and we maximise the prediction error decomposition form of the log-likelihood (4.17) to estimate  $\mu$ ,  $\sigma_{\varepsilon}^2$  and  $\sigma_{\mathbf{a}}^2$ . For completeness, the REML estimates of the variance parameters are also evaluated.

We compare the empirical distribution functions of the bootstrap replicates of  $\hat{\sigma}_{\mathbf{a}}^2$  obtained using the parametric bootstrap and the exact bootstrap of [Fears et al. \(1996\)](#). Our bootstrap resampling process employs 599 bootstrap replications.

We use equation (4.31) to bias-correct the FIML estimate,  $\hat{\sigma}_{\mathbf{a}}^2$ . Additionally, we bootstrap the Wald statistic using the method outlined in section 3.2.1 of the previous chapter and we retest the null hypothesis (4.33).

#### 4.6.1 Empirical Study: Results and Discussion

Our objective was to demonstrate that Kalman filter output could be used to estimate and bootstrap the model of [Fears et al. \(1996\)](#).

Table 4.8 presents their FIML estimates for  $\sigma_{\varepsilon}^2$  and  $\sigma_{\mathbf{a}}^2$ . Additionally, Table 4.9 presents our FIML estimates for these error variances, acquired via BFGS maximisation of the log-likelihood, along with the REML estimates.

When we compare the figures in Tables 4.8 and 4.9, we see that the estimates found by [Fears et al. \(1996\)](#) are identical to the ones obtained using the Kalman filter approach. Specifically,  $\hat{\sigma}_{\varepsilon}^2 = 0.00325$  and  $\hat{\sigma}_{\mathbf{a}}^2 = 0.01395$  in both cases.

As we mentioned in Section 4.1, relative to the REML estimates, FIML estimates of the level-2 variance terms tend to exhibit a downward bias in small samples; this effect is evident from the data presented in Table 4.9. In particular, the FIML estimate of  $\sigma_{\mathbf{a}}^2$  is 0.01395, compared to the REML estimate of 0.01750.

An approximation to the finite sample CDF of  $\hat{\sigma}_{\mathbf{a}}^2$ , based on the exact

bootstrap, is shown in Figure 4.2. The exact bootstrap differs from other resampling methods in that it uses the entire space of resamples. This means that given a data set of  $n$  observations, we draw  $B = n^n$  samples of size  $n$ , with replacement. The probability of selecting a particular sample is therefore  $n^{-n}$ . With other resampling techniques, where we work with a subset of the resample space,  $B$  is generally much smaller than  $n^n$ . For the purpose of comparison, the approximate CDF of  $\hat{\sigma}_a^2$ , based on the parametric bootstrap procedure outlined in subsection 4.4.2, is overlaid on the same axes. Figure 4.2 reveals that the two distribution functions are remarkably similar.

The exact bootstrap distribution of  $\sigma_a^2$  has a mean of 0.0111 and a standard deviation of 0.0061, as reported in Fears et al. (1996). The parametric bootstrap distribution provides an almost identical mean and standard deviation of 0.0110 and 0.0061, respectively (Table 4.10). From these outcomes, we see that, in this instance, the performance of the parametric bootstrap is comparable to that of the exact bootstrap.

There are usually two sources of error in the bootstrapping process: the error introduced by using a particular sample to represent the background population and the error resulting from the failure to enumerate all possible bootstrap resamples. Since the exact bootstrap involves the generation of all possible resamples, this facilitates the calculation of every realisation of the quantity of interest. So, the advantage of the exact bootstrap lies in the fact that it avoids the latter cause of inaccuracy. Nevertheless, it rapidly becomes intractable. Fears et al. (1996) employed 5 level-2 units ( $J = 5$ ) and computed

a total of 3125 resamples, out of which 126 were unique.<sup>9</sup> It is important to point out that 5 clusters represents an extremely small sample size. If  $J$  is increased to 15, the number of unique resamples jumps to approximately  $7.76 \times 10^7$ . Clearly, even with advancements in computing speeds, the exact bootstrap can quickly become cumbersome. Having a viable alternative, such as the parametric bootstrap, can be helpful.

Since we know that the FIML estimates of the level-2 variances tend to be biased downward, we employed the parametric bootstrap to bias-correct  $\hat{\sigma}_a^2$ . The results are given in Table 4.11. To further refine our estimation, we performed a second round of bootstrapping using a fast double bootstrap (Ouyssse, 2013). The results in Table 4.11 indicate that the double bootstrap bias-corrected estimate of  $\hat{\sigma}_a^2$  is comparable to the REML estimate.

The Wald statistic in Fears et al. (1996), had a  $p$ -value of 0.1191. The authors remarked that this outcome indicated there was no variation in estrone levels among the subjects, which seemed improbable. They attributed this to the small sample size and suggested that a larger sample could yield more reliable outcomes. We bootstrapped the Wald statistic using the method outlined in Section 3.2.1 of the preceding chapter. Namely, we generated bootstrap samples under the null hypothesis, calculated the Wald statistic for each sample, and determined the proportion of bootstrap samples for which the Wald statistic exceeded the original Wald test statistic. The resulting bootstrap  $p$ -value was 0.001. As a result, we rejected the null hypothesis of

---

<sup>9</sup>The total number of resamples is  $n^n = 5^5 = 3125$ , in which some draws contain identical elements. The number of distinct samples is found from the formula for a combination with replacement, i.e.,  ${}^{n+r-1}C_r = \frac{(n+r-1)!}{r!(n-1)!}$  where  $n$  is the total number of objects in a set and  $r$  is the number of items being selected. In the Fears et al. (1996) example with  $n = r = 5$ , we find  ${}^9C_5 = 126$ .

homogeneity in estrone levels among the participants. This situation is similar to that of [Newbold and Bos \(1985\)](#) in the previous chapter, where the use of the bootstrap led to a different outcome of the hypothesis test.

## **4.7 Concluding Remarks**

Multilevel models are typically not analysed using the state space format. However, this chapter argued that casting multilevel models in state space form may be advantageous for estimation. Output from the Kalman filter recursions is valuable, not only for parameter estimation by seamlessly facilitating the prediction error decomposition form of the log-likelihood, but also for executing the parametric and residual bootstraps. On occasion, researchers are faced with limited sample sizes. In these instances, bootstrap bias-correction may be useful to provide improvements on the FIML estimates of the variance elements in the covariance matrix of the vector of random effects. Additionally, in extremely small samples like the ones considered in our investigation, double bootstrap bias-correction may be required to bring the FIML estimates in line with their REML counterparts.

In our empirical example, we observed, once again, that our hypothesis test resulted in a different decision when we bootstrapped the test statistic. This emphasises the fact that using bootstrap methods can be advantageous over asymptotic approaches when dealing with small sample sizes. However, it is important to note that a cluster size of 5, as utilised in [Fears et al. \(1996\)](#), is atypical. Most researchers aim to work with at least 10 clusters, if at all possible, to avoid the difficulties associated with small samples.

*Bootstrapping Multilevel Models: A State Space Approach*

The estimation and bootstrapping of the multilevel models in this paper were successfully implemented using the state space representation. As highlighted by [Gu et al. \(2014\)](#), estimating the state space formulation has the additional advantage of being faster than traditional FIML estimation in hierarchical models. This underscores the idea that working in the state space framework offers a practical alternative to more conventional methods when dealing with clustered data.

## 4.8 Tables and Figures

### 4.8.1 Simulation: Tables

#### Parametric Bootstrap

Table 4.1: Estimates  $J = 10, n = 100$

	$\hat{\theta}_{REML}$	$\hat{\theta}_{FIML}$	$\theta^*$	$\hat{\theta}_{FIML-BC}$	$\frac{\hat{\theta}_{FIML}-\theta_0}{\theta_0}$	$\frac{\hat{\theta}_{FIML-BC}-\theta_0}{\theta_0}$
$\tau_{00} = 2.0$	1.938	1.546	1.232	1.860	-0.227	-0.070
$\tau_{01} = 0.5$	0.468	0.374	0.300	0.448	-0.252	-0.103
$\tau_{11} = 2.0$	1.921	1.533	1.223	1.843	-0.234	-0.078

Table 4.2: Estimates  $J = 15, n = 100$

	$\hat{\theta}_{REML}$	$\hat{\theta}_{FIML}$	$\theta^*$	$\hat{\theta}_{FIML-BC}$	$\frac{\hat{\theta}_{FIML}-\theta_0}{\theta_0}$	$\frac{\hat{\theta}_{FIML-BC}-\theta_0}{\theta_0}$
$\tau_{00} = 2.0$	2.005	1.735	1.501	1.969	-0.132	-0.015
$\tau_{01} = 0.5$	0.515	0.446	0.387	0.506	-0.107	0.012
$\tau_{11} = 2.0$	2.086	1.805	1.562	2.048	-0.097	0.024

Table 4.3: Estimates  $J = 20, n = 100$

	$\hat{\theta}_{REML}$	$\hat{\theta}_{FIML}$	$\theta^*$	$\hat{\theta}_{FIML-BC}$	$\frac{\hat{\theta}_{FIML}-\theta_0}{\theta_0}$	$\frac{\hat{\theta}_{FIML-BC}-\theta_0}{\theta_0}$
$\tau_{00} = 2.0$	2.022	1.818	1.634	2.002	-0.091	0.001
$\tau_{01} = 0.5$	0.490	0.441	0.397	0.485	-0.119	-0.031
$\tau_{11} = 2.0$	2.025	1.820	1.636	2.005	-0.090	0.003



**Residual Bootstrap**

Table 4.4: Estimates  $J = 10, n = 100$

	$\hat{\theta}_{REML}$	$\hat{\theta}_{FIML}$	$\theta^*$	$\hat{\theta}_{FIML-BC}$	$\frac{\hat{\theta}_{FIML}-\theta_0}{\theta_0}$	$\frac{\hat{\theta}_{FIML-BC}-\theta_0}{\theta_0}$
$\tau_{00} = 2.0$	1.938	1.546	1.281	1.811	-0.227	-0.095
$\tau_{01} = 0.5$	0.468	0.374	0.309	0.439	-0.252	-0.122
$\tau_{11} = 2.0$	1.921	1.533	1.264	1.802	-0.234	-0.099

Table 4.5: Estimates  $J = 15, n = 100$

	$\hat{\theta}_{REML}$	$\hat{\theta}_{FIML}$	$\theta^*$	$\hat{\theta}_{FIML-BC}$	$\frac{\hat{\theta}_{FIML}-\theta_0}{\theta_0}$	$\frac{\hat{\theta}_{FIML-BC}-\theta_0}{\theta_0}$
$\tau_{00} = 2.0$	2.005	1.735	1.579	1.891	-0.132	-0.055
$\tau_{01} = 0.5$	0.515	0.446	0.371	0.521	-0.107	0.042
$\tau_{11} = 2.0$	2.086	1.805	1.464	2.146	-0.097	0.073

Table 4.6: Estimates  $J = 20, n = 100$

	$\hat{\theta}_{REML}$	$\hat{\theta}_{FIML}$	$\theta^*$	$\hat{\theta}_{FIML-BC}$	$\frac{\hat{\theta}_{FIML}-\theta_0}{\theta_0}$	$\frac{\hat{\theta}_{FIML-BC}-\theta_0}{\theta_0}$
$\tau_{00} = 2.0$	2.022	1.818	1.635	2.001	-0.091	0.001
$\tau_{01} = 0.5$	0.490	0.441	0.394	0.488	-0.119	-0.024
$\tau_{11} = 2.0$	2.025	1.820	1.583	2.057	-0.090	0.029

### 4.8.2 Empirical Study: Tables

Table 4.7: Estrone Assay Measurements (pg/mL) of a Single Blood Sample From Each of Five Postmenopausal Women

	Individual				
Sample	P1	P2	P3	P4	P5
1	23	25	38	14	46
2	23	33	38	16	36
3	22	27	41	15	30
4	20	27	38	19	29
5	25	30	38	20	36
6	22	28	32	22	31
7	27	24	38	16	30
8	25	22	42	19	32
9	22	26	35	17	32
10	22	30	40	18	31
11	23	30	41	20	30
12	23	29	37	18	32
13	27	29	28	12	25
14	19	37	36	17	29
15	23	24	30	15	31
16	18	28	37	13	32

*a*

Table 4.8: Variance Estimates for One-Way Random Effects Model using VARCOMP

Parameter	SAS Estimates
$\sigma_{\epsilon}^2$	0.00325 (0.00053)
$\sigma_a^2$	0.01395 (0.00895)

Table 4.9: Variance Estimates for One-Way Random Effects Model Using State Space Approach

Parameter	Kalman Filter FIML Estimates	REML Estimates
$\sigma_{\epsilon}^2$	0.00325 (0.00053)	0.00325. (0.00053)
$\sigma_a^2$	0.01395 (0.00895)	0.01750 (0.0090)

Table 4.10: Bootstrap Estimates of the Variance of the Random Effects

	Estimates	
	Exact Bootstrap	Parametric Bootstrap
$\sigma_a^2$	0.0111 (0.0061)	0.0110 (0.0061)

<sup>a</sup>In the tables, the standard errors appear in parentheses.

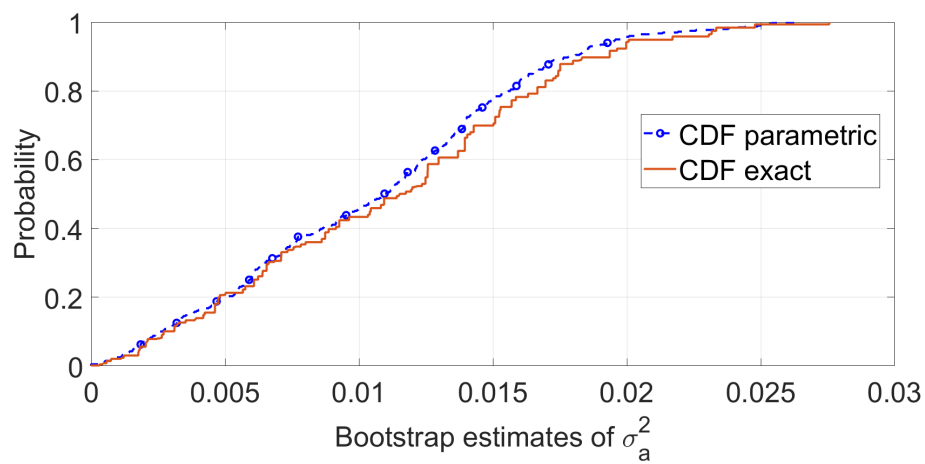
Table 4.11: Estimates of  $\sigma_a^2$

	Estimate
$\hat{\sigma}_{a, FIML}^2$	0.0140
<sup>a</sup> $\hat{\sigma}_{a, bc}^2$	0.0169
$\hat{\sigma}_{a, dbc}^2$	0.0176
$\hat{\sigma}_{a, REML}^2$	0.0175

---

<sup>a</sup> $\hat{\sigma}_{a, bc}^2$  and  $\hat{\sigma}_{a, dbc}^2$  are the bootstrap and double bootstrap bias-corrected estimates of  $\hat{\sigma}_{a, FIML}^2$ , respectively.

Figure 4.2: Empirical Cumulative Distribution Functions



# Chapter 5

## Conclusion

The primary objective of this study was to extend the research of [Stoffer and Wall \(1991\)](#) by introducing a bootstrap technique that is applicable in heteroskedastic state space models. To this end, we developed the state space model (SSM) bootstrap which utilises a gradient-based approach to the bootstrapping process. Our findings suggest that our bootstrap performs well in a variety of situations. Furthermore, we aimed to illustrate that Kalman filter output can be employed for both estimation and bootstrapping purposes if the models are represented in their state space format.

Up to this point, as far as we know, the bootstrap of [Stoffer and Wall \(1991\)](#) was the only one developed for use in state space models. However, it has the limitation of being created solely for homoskedastic models, along with being computationally expensive. In Chapter 2, we crafted a bootstrap technique, termed the wild state space model bootstrap (WSSM), specifically for state space models with heterogeneous errors. We tested its efficacy by examining the coverage probabilities of a variety of bootstrap confidence intervals and

## *Conclusion*

observed that they outperformed their asymptotic counterparts. Additionally, we employed the WSSM bootstrap to develop the AICg, a bootstrap version of the Akaike Information Criterion (AIC). When it came to model selection in small samples, this bootstrap variant achieved greater success than the conventional AIC. We also demonstrated that our bootstrap methodology could be adjusted for use in homoskedastic models, providing an alternative to the approach of [Stoffer and Wall \(1991\)](#).

Our bootstrap procedure bears some similarity to the  $k$ -step bootstrap of [Davidson and MacKinnon](#) and, in Chapter 3, we examined the feasibility of adopting their methodology for use in state space models. In smaller samples, their approximate bootstrap yielded superior results when compared to asymptotic methods. Additionally, working in the state space format for bootstrap hypothesis testing conferred some advantages since the Kalman filter output could be used to formulate the Score, Likelihood Ratio, and Wald test statistics.

Although the state space format is not commonly used for analysing multilevel models, we contended that employing this approach could offer benefits for estimation and bootstrapping. In Chapter 4, we explored the use of the residual and parametric bootstraps in multilevel models and investigated how to implement them when those models were expressed in the state space format.

It is well known that in clustered models, full information maximum likelihood (FIML) estimates of the variance components in the covariance matrix of the random effects vector tend to exhibit a downward bias. In contrast, restricted maximum likelihood estimates (REML) are subject to

## *Conclusion*

less bias. Despite this, FIML estimation remains prevalent due to its ability to use chi-squared difference tests in comparing nested models with varying numbers of regression or variance parameters. On the other hand, REML estimation only permits comparisons of variance entries. Our study explored the potential for correcting the downward bias in FIML estimates through the application of bootstrap methods. Our results revealed that employing either single or double bootstrap bias-correction techniques effectively aligned the FIML and REML estimates.

The development of the SSM bootstrap is significant as it addresses a critical gap in the literature. Previously, the sole bootstrap technique for state space models could only be applied when the error terms in both the state and measurement equations were homoskedastic. Although it is not uncommon for researchers to assume homoskedasticity for simplicity, this assumption can sometimes be unrealistic. Therefore, our methodology offers the potential to bootstrap a much wider range of state space models. In addition, while a bootstrap version of the AIC is available, as it uses the bootstrap of [Stoffer and Wall \(1991\)](#), it, too, can only be used for homoskedastic models. Our AICg measure provides a means for model selection in heteroskedastic state space models. Furthermore, although a method for bootstrapping homogeneous state space models already exists, its execution can be time-consuming. Our RSSM bootstrap offers a valuable alternative because it has been shown to be notably faster than the method of [Stoffer and Wall \(1991\)](#), resulting in significant time savings.

While our research offers some insights, there are also some limitations. Firstly, our investigations only focused on stationary state space models. As



## *Conclusion*

a result, our technique is not applicable to non-stationary models such as structural time series models; these are frequently used in the state space literature but are typically non-stationary. Secondly, our SSM bootstrap relies on one quasi-Newton step during implementation, which may not be sufficient in very small samples. Thus, our methodology may be better suited to samples of size 50 or more. Finally, due to limited access to computer cores, the number of Monte Carlo simulations that could have been performed in each investigation was restricted.

Looking ahead, there is ample opportunity to build on our work and extend the application of our bootstrap methodology. In Chapter 2, when we developed the SSM bootstrap, we worked with the perturbed score contributions. One potential area for extending our research is to perturb the Hessian as well, rather than working with a fixed Hessian. We suspect that doing so could improve the performance of our bootstrap procedure.

Our work could also be extended to explore the application of the SSM bootstrap to models with missing observations, an issue which presents a common challenge in research. In this scenario, both the Kalman filter and the Kalman smoother would be required to provide estimates for the missing observations. Examining the effectiveness of the SSM bootstrap in this context could be a valuable direction for future investigation.

Our study suggested that using our bootstrap technique could lead to significant time savings. However, further analysis is needed to explore the potential time savings that could result from variations in the form of the Kalman filter itself. For instance, in the case of multilevel models with a large number of clusters, using the sequential Kalman filter could be

### *Conclusion*

advantageous when bootstrapping. This variant of the Kalman filter, which is mathematically equivalent to the forms used in our research, avoids any matrix inversion during execution and could be particularly useful in situations where computational time is a crucial factor. A comparative study between the sequential and standard Kalman filters would provide useful insights into potential time savings.

# Bibliography

- Anderson, B. D. and Moore, J. B. (2012). *Optimal filtering*. Courier Corporation.
- Ansley, C. F. and Newbold, P. (1980). Finite sample properties of estimators for autoregressive moving average models. *Journal of Econometrics*, 13(2):159–183.
- Carpenter, J. R., Goldstein, H., and Rasbash, J. (2003). A novel bootstrap procedure for assessing the relationship between class size and achievement. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(4):431–443.
- Cavanaugh, J. E. and Neath, A. A. (2019). The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *Wiley Interdisciplinary Reviews: Computational Statistics*, 11(3):e1460.
- Cavanaugh, J. E. and Shumway, R. H. (1997). A bootstrap variant of AIC for state-space model selection. *Statistica Sinica*, pages 473–496.
- Davidson, J., Monticini, A., and Peel, D. (2007). Implementing the wild

- bootstrap using a two-point distribution. *Economics Letters*, 96(3):309–315.
- Davidson, R. and Flachaire, E. (2008). The wild bootstrap, tamed at last. *Journal of Econometrics*, 146(1):162–169.
- Davidson, R. and MacKinnon, J. G. (1999). Bootstrap testing in nonlinear models. *International Economic Review*, 40(2):487–508.
- Davidson, R. and MacKinnon, J. G. (2007). Improving the reliability of bootstrap tests with the fast double bootstrap. *Computational Statistics & Data Analysis*, 51(7):3259–3281.
- Davidson, R. and Trokić, M. (2020). The fast iterated bootstrap. *Journal of Econometrics*, 218(2):451–475.
- Dent, W. and Min, A.-S. (1978). A Monte Carlo study of autoregressive integrated moving average processes. *Journal of Econometrics*, 7(1):23–55.
- Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*, volume 38 of *Oxford Statistical Science Series*. OUP Oxford.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7:1–26.
- Fears, T. R., Benichou, J., and Gail, M. H. (1996). A reminder of the fallibility of the Wald statistic. *The American Statistician*, 50(3):226–227.
- Freedman, D. A. (1981). Bootstrapping regression models. *The Annals of Statistics*, 9(6):1218–1228.

- Godfrey, L. G. (1988). *Misspecification tests in econometrics: The Lagrange multiplier principle and other approaches*, volume 16 of *Cambridge studies in probability, induction, and decision theory*. Cambridge University Press.
- Goldstein, H. (1989). Restricted unbiased iterative generalized least-squares estimation. *Biometrika*, 76(3):622–623.
- Goldstein, H. (2011). *Multilevel Statistical Models*, volume 922 of *Wiley Series in Probability and Statistics*. John Wiley & Sons, 4th edition.
- Gonçalves, S. and Kilian, L. (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics*, 123(1):89–120.
- Gu, F., Preacher, K. J., Wu, W., and Yung, Y.-F. (2014). A computationally efficient state space approach to estimating multilevel regression models and multilevel confirmatory factor models. *Multivariate Behavioral Research*, 49(2):119–129.
- Hall, P. (2013). *The Bootstrap and Edgeworth Expansion*. Springer Science & Business Media.
- Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman filter*. Cambridge University Press.
- Harvey, A. C. (2014). Modelling the Phillips curve with unobserved components. In *Perspectives on Econometrics and Applied Economics*. Routledge.

- Jiang, J. (1997). A derivation of BLUP - Best linear unbiased predictor. *Statistics Probability Letters*, 32(3):321–324.
- Kline, P. and Santos, A. (2012). A score based approach to wild bootstrap inference. *Journal of Econometric Methods*, 1(1):23–41.
- Kreiss, J.-P. (1997). *Asymptotical properties of residual bootstrap for autoregressions*. Citeseer.
- Liu, R. Y. et al. (1988). Bootstrap procedures under some non-*iid* models. *The Annals of Statistics*, 16(4):1696–1708.
- Ljung, L. and Caines, P. (1979). Asymptotic normality of prediction error estimators for approximate system models. *IEEE Transactions on Automatic Control*, 24(5):742–746.
- MacKinnon, J. G. (2006). Bootstrap methods in econometrics. *Economic Record*, 82:S2–S18.
- Martin, M. A. (1992). On the double bootstrap. In *Computing Science and Statistics*, pages 73–78. Springer New York.
- McNeish, D. (2017). Small sample methods for multilevel modeling: A colloquial elucidation of REML and the Kenward-Roger correction. *Multivariate Behavioral Research*, 52(5):661–670.
- Mittelhammer, R. C., Judge, G. G., and Miller, D. J. (2000). *Econometric Foundations*. Cambridge University Press.

- Modugno, L. and Giannerini, S. (2015). The wild bootstrap for multilevel models. *Communications in Statistics-Theory and Methods*, 44(22):4812–4825.
- Newbold, P. and Bos, T. (1985). *Stochastic Parameter Regression Models*, volume 51 of *Quantitative Applications in the Social Sciences*. Sage Publications.
- Ouysse, R. (2013). A fast iterated bootstrap procedure for approximating the small-sample bias. *Communications in Statistics - Simulation and Computation*, 42(7):1472–1494.
- Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58(3):545–554.
- Rice, N., Jones, A. M., and Goldstein, H. (2002). Multilevel models where the random effects are correlated with the fixed predictors. *Centre for Health Economics, University of York, York*.
- Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects. *Statistical Science*, 6(1):15–32.
- Robinson, P. M. (1988). The stochastic difference between econometric statistics. *Econometrica: Journal of the Econometric Society*, pages 531–548.
- Snijders, T. A. and Bosker, R. J. (2011). *Multilevel Analysis: An introduction to basic and advanced multilevel modeling*. Sage Publications.

- Snijders, T. A. B. and Bosker, R. J. (1993). Standard errors and sample sizes for two-level research. *Journal of Educational Statistics*, 18(3):237.
- Stoffer, D. S. and Wall, K. D. (1991). Bootstrapping state-space models: Gaussian maximum likelihood estimation and the Kalman filter. *Journal of the American Statistical Association*, 86(416):1024–1033.
- Stoffer, D. S. and Wall, K. D. (2004). Resampling in state space models. In *State space and Unobserved Component Models: Theory and Applications*, pages 227–258. Cambridge University Press.
- Sugiura, N. (1978). Further analysis of the data by Akaike’s information criterion and the finite corrections. *Communications in Statistics-Theory and Methods*, 7(1):13–26.
- Van der Leeden, R., Busing, F., and Meijer, E. (1997). Bootstrap methods for two-level models. *Journal of Educational and Behavioral Statistics*, 22(3):310–335.
- Van der Leeden, R., Meijer, E., and Busing, F. M. (2008). Resampling multilevel models. In *Handbook of Multilevel Analysis*, pages 401–433. Springer.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the econometric society*, pages 1–25.