A STOCHASTIC PARAMETER REGRESSION MODEL FOR LONG MEMORY TIME SERIES

by

Rose Marie Ocker

A thesis

submitted in partial fulfillment

of the requirements for the degree of

Master of Science in Mathematics

Boise State University

May 2014

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the thesis submitted by

Rose Marie Ocker

Thesis Title: A Stochastic Parameter Regression Model for Long Memory Time Series

Date of Final Oral Examination: 07 March 2014

The following individuals read and discussed the thesis submitted by student Rose Marie Ocker, and they evaluated her presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Jaechoeul Lee, Ph.D.	Chair, Supervisory Committee
Leming Qu, Ph.D.	Member, Supervisory Committee
Partha Mukherjee, Ph.D.	Member, Supervisory Committee

The final reading approval of the thesis was granted by Jaechoeul Lee, Ph.D., Chair of the Supervisory Committee. The thesis was approved for the Graduate College by John R. Pelton, Ph.D., Dean of the Graduate College.

dedicated to John and Mary Anne Ocker

ACKNOWLEDGMENTS

The author wishes to thank her family and friends for their unending support and encouragement and the many teachers who sparked her interest in math and science, especially Dr. Lynda Danielson and Dr. James Dull. She also wishes to thank her adviser, Dr. Jaechoul Lee, whose patience and guidance were of utmost importance in this project. Finally, she expresses gratitude to the National Science Foundation who funded her research in this topic.

ABSTRACT

In a complex and dynamic world, the assumption that relationships in a system remain constant is not necessarily a well-founded one. Allowing for time-varying parameters in a regression model has become a popular technique, but the best way to estimate the parameters of the time-varying model is still in discussion. These parameters can be autocorrelated with their past for a long time (long memory), but most of the existing models for parameters are of the short memory type, leaving the error process to account for any long memory behavior in the response variable. As an alternative, we propose a long memory stochastic parameter regression model, using a fractionally integrated (ARFIMA) noise model to take into account long memory autocorrelations in the parameter process. A fortunate consequence of this model is that it deals with heteroscedasticity without the use of transformation techniques. Estimation methods involve a Newton-Raphson procedure based on the Innovations Algorithm and the Kalman Filter, including truncated versions of each to decrease computation time without a noticeable loss of accuracy. Based on simulation results, our methods satisfactorily estimate the model parameters. Our model provides a new way to analyze regressions with a non-stationary long memory response and can be generalized for further application; the estimation methods developed should also prove useful in other situations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

SPRM – Stochastic parameter regression model

iid – Independent and identically distributed

ARMA – Autoregressive Moving-Average

ARIMA – Autoregressive Integrated Moving-Average

ARFIMA – Autoregressive Fractionally Integrated Moving-Average

ACF – Autocorrelation function

PACF – Partial autocorrelation function

OLS – Ordinary Least-Squares

FIA – Full Innovations Algorithm

TIA – Truncated Innovations Algorithm

FKF – Full Kalman Filter

TKF – Truncated Kalman Filter

LIST OF SYMBOLS

B backward shift operator; $BX_t = X_{t-1}$

∇ $(1 - B)$; $\nabla X_t = X_t - X_{t-1}$

Γ gamma function; $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ for positive α, unless otherwise noted

$\gamma(h)$ Autocovariance function at lag h (for stationary series)

$\rho(h)$ Autocorrelation function at lag h (for stationary series)

CHAPTER 1

INTRODUCTION

1.1 Motivation

In a complex and dynamic world, the assumption that relationships in a system remain constant is not necessarily a well-founded one. Classical time series regression analysis assumes that the relationships between the explanatory variables and the response variables are constant. In situations where this is not the case, applying a classical time series regression model can lead to non-negligible errors in our predictions of the response variables. Researchers in other fields have turned to modeling situations with non-constant coefficients to explain a more dynamic relationship between a predictor and its response. For example, changing house prices in the United Kingdom were modeled via time-varying coefficients in [3]. For simplicity, we concentrate on the univariate time series regression, classically represented as

$$y_t = z_t^T \beta + \epsilon_t \tag{1.1}$$

where y_t is the response variable, z_t is a known $r \times 1$ vector $(z_{1t}, z_{2t}, ...z_{rt})^T$ of explanatory variables, β is an unknown $r \times 1$ vector $(\beta_1, ...\beta_r)$ of regression coefficients, and ϵ_t is the error at time t. Typically, β_t and ϵ_t are assumed to be uncorrelated. This implies that $var(y_t) = var(\epsilon_t)$. That means any changes in the variance of the

response is allegedly due to a change in the variance of the errors. While this may be the case, a change in variance could also be due to a change in the relationship between the explanatory variables and the response variable. Theoretical fixes for non-stationary variance do exist and have been able to model some situations quite well [7]. However, a different approach, one where changes in the variance are due to the dynamic relationship between y_t and z_t, may prove more efficient and instructive.

A series, $\{y_t\}$, is said to be stationary if its mean and autocovariance function (ACVF), $\gamma_y(h)$, are independent of t. That is, a stationary series has the same mean for every time point in the series. Furthermore, a stationary series has $Cov(y_t, y_s) = Cov(y_{t+i}, y_{s+i})$ for all i such that $t + i,\ s + i \in 1, ..., n$. In that case, we can set $h = |t - s|$ and say $Cov(y_t, y_s) = Cov(y_i, y_{i+h}) = \gamma_y(h)$, as the covariance between the variables depends only on the lag $h = |t - s|$. In other words, the relationship between two observations only depends on how far apart they are, and not on *when* they appear in the series. If $\gamma_y(h)$ decreases very slowly, the data set is said to be modeled by a long memory time series. More precisely, a data set is long memory if

$$\sum_{h=-\infty}^{\infty} |\gamma_y(h)| = \infty. \tag{1.2}$$

The most widely used model for long memory time series is the ARFIMA(p, d, q), or autoregressive fractionally integrated moving-average model. The popularity of the ARFIMA(p, d, q) model may lie in its versatility; it can approximate well any long memory process [18] and can also account for some short memory behavior [10]. It is also common for an ARFIMA process to have local changes in mean even though it is stationary, which can account for otherwise inexplicable shifts in the mean of a time series.

There are many different ways to account for long memory behavior in the response variable of a time series regression. In the classical model, the long memory component of the response is taken into account in the error series, $\{\epsilon_t\}$. In this case, we often model ϵ_t as an ARFIMA(p, d, q) process. This approach has been fairly successful in economic and climatology data sets [1], [9], [12], [13], and [14]. Unfortunately, the complicated autocovariance structure of ARFIMA(p, d, q) models presents a major difficulty with this technique. The inherent issues with analyzing a series that is defined using a non-convergent infinite sum make more traditional investigatory procedures challenging if not impossible. Furthermore, because a data set can show both long memory and short memory properties, more complicated models arise in research. One commonly used model of this type is a linear regression with a stochastic ARMA parameter and long memory errors, discussed in Chapter 2. However, treating long memory as part of the error series may under-emphasize the importance of the long memory factor in the response.

1.2 Literature Review

In 1970, Thomas Burnett and Donald Gurthrie [4] developed estimators for stochastic parameters. They found the best linear predictors that minimized the mean square error, but they assumed that observation and explanatory variables were stationary and that the stochastic parameter had a known covariance structure. In this paper, we also assume that our stochastic parameter has a known covariance structure, but relax the assumption that the observation and explanatory variables are stationary. Given that most data sets are not stationary, this is an important improvement.

In "A Survey of Stochastic Parameter Regression" [20], Barr Rosenberg explains

the importance of the stochastic parameter regression models. First, he makes the distinction between stochastic and systematic parameter processes. Systematic parameters vary with time but are deterministic and can thus be defined by a function. Stochastic parameters, however, are actually a realization of a process at a given time t. Rosenberg is slightly dismissive of stochastic parameter regression models in which the parameters are stationary time series processes, saying they are of theoretical value, but as they cannot, in general, be written in Markovian Canonical Form, the computational burden of these approaches is too high to be useful. A general comparison of the OLS approach to the stochastic parameter approach is made. In it, Rosenberg reports that inappropriate use of fixed-parameter regression results in unbiased but highly inefficient estimation of parameters. In fact, the OLS tends to estimate the average of the stochastic process, sometimes resulting in predictions that are merely an average. Methods of estimation for stochastic parameters are discussed, with the maximum likelihood estimation and Bayes Estimation being the most commonly applied, along with the Kalman Filter estimation. Rosenberg concludes with an example in economics illustrating how stochastic parameter models are important; if the stochastic process itself is a meaningful quantity, then changing relationships between predictor and response need to be more meticulously analyzed.

Newbold and Bos developed a stochastic regression model similar to the one discussed in this paper. In their paper, "Stochastic Parameter Regression Models" [17], they regressed the quarterly inflation rate on the quarterly interest rate with the stochastic parameter following an AR(1) process with mean b as follows:

$$y_t = \alpha + \beta_t z_t + \nu_t, \tag{1.3}$$

$$(\beta_t - b) = \phi(\beta_{t-1} - b) + \omega_t, \tag{1.4}$$

where ν_t are iid errors. Robinson and Hidalgo [19] use this approach but generalized it by allowing ν_t to have long memory. That approach is the most widely used in modeling long memory in the response variable and is discussed further in Section 2.3.

1.3 Our Approach

We present a different approach, regressing a series that shows long memory on an explanatory series with a stochastic parameter that shows long memory instead of confining long memory behavior into the errors. Specifically, we suppose the relationship

$$y_t = \mu + \beta_t z_t + \alpha a_t + \epsilon_t, \qquad \epsilon_t \sim \text{iid } \mathrm{N}(0, \sigma_\epsilon^2) \tag{1.5}$$

$$\beta_t = (1 - B)^{-d} \omega_t, \qquad \omega_t \sim \text{iid } \mathrm{N}(0, \sigma_\omega^2) \tag{1.6}$$

holds, where y_t is a long memory time series, μ is a constant location parameter, z_t is an explanatory series, α is a constant slope parameter, a_t is a series of known inputs, β_t is an ARFIMA$(0, d, 0)$ process with mean zero, $d \in (-1, .5)$ and $\{\epsilon_t\}$ and $\{\omega_t\}$ are uncorrelated. As mentioned previously, one benefit of the ARFIMA model is that it can be generalized to include long memory and some short memory behavior by letting $\beta_t \sim \text{ARFIMA}(p, d, q)$ [10]. We choose the ARFIMA$(0, d, 0)$ model because of the simple expressions involved in its ACVF and MA(∞) specification. However, our approach could be generalized to suit other situations in which short memory influences exist. Another benefit of this approach is that it can naturally account for heteroscedasticity in the response, since the variance changes with z_t. Perhaps the most important contribution of this model is that it allows for a more precise understanding of the relationship between y_t and z_t, providing more power in

analysis and prediction. We modify and use two previously established algorithms, the Innovations Algorithm and the Kalman Filter, to estimate the parameters of our model.

The rest of this paper proceeds as follows: Chapter 2 provides necessary background on time series for those less familiar with the topic, Chapter 3 explains the details of our model, Chapter 4 discusses the Kalman Filter estimation method, Chapter 5 discusses the Innovations Algorithm estimation method, Chapter 6 shows a summary of simulations we performed to test the practical accuracy of the model, and Chapter 7 contains concluding remarks.

CHAPTER 2

PRELIMINARY BACKGROUND

2.1 Autocorrelation

Given a series of random variables, $\{y_t\}$, it is common to assume in introductory statistics that the y_t are independent and identically distributed. While independence implies that two observations are not related at all, two observations being uncorrelated implies no linear relationship. If a series is correlated with itself, it is said to be autocorrelated. More formally, y_t and y_s, $s \neq t$, are autocorrelated if $E(y_t y_s) - E(y_t)E(y_s) \neq 0$. When y_t and y_s are correlated, $E(y_t|y_s) \neq E(y_t)$. Thus, if information about the underlying autocorrelation structure of $\{y_t\}$ can be obtained, then with observations through time t, we can use $E(y_{t+1}|y_t, y_{t-1}, ...y_1)$ to better predict y_{t+1}.

Instead of assuming a series is independent or uncorrelated, time series takes advantage of autocorrelation in a series when trying to understand how observations are related, providing powerful prediction capabilities. One widely implemented class of time series models is known as Autoregressive Moving-Average (ARMA) models. Let $\{w_t\}$, $t = 1, ..., n$, be an iid series. Then an ARMA(1,1) model for $\{y_t\}$ would be $y_t - \phi y_{t-1} = w_t + \theta w_{t-1}$ where ϕ and θ are undetermined parameters. Thus, the observation y_t depends on the observation and corresponding error from one time point ago and the error associated with time t. If an ARMA(2,1) model were more

appropriate, that would suggest the observation y_t depended on the observations from two time points ago, the error from one time point ago, and the error associated with time t. An ARMA(2,1) model reads $y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} = z_t + \theta z_{t-1}$. The number of parameters in an ARMA model is theoretically allowed to be any number, but in practice a large number of parameters is frowned upon for reasons of computation, understandability, and simplicity.

2.2 Long Memory Time Series

Long memory time series are implemented when the current observation, y_t, shows correlation with y_{t-k}, where k takes on arbitrarily large values. One widely used definition of a long memory time series mentioned above is

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| = \infty, \tag{2.1}$$

where $\gamma(h)$ denotes the autocovariance at lag h. This definition is a result of analyzing long memory time series in the time domain as opposed to the frequency domain. A popular definition of long memory used in the frequency domain treatment of time series is

$$f(\lambda) \sim |\lambda|^{-2d} l_2(1/|\lambda|), \tag{2.2}$$

where λ is in a neighborhood of zero and l_2 is a slowly varying function [18], p. 40. This paper focuses solely on the time domain. For more information on the frequency domain of a time series, see [2] and [18]. Qualitatively, a long memory time series is characterized by the idea that the current observation is related to *all past observations*. For example, monthly inflation rates might be expected to have a long memory,

as the expectation of inflation can cause inflation. In fact, monthly inflation rates have been examined in the past and found to have long memory characteristics, such as slowly decaying autocorrelation function values [9]. Data sets with long memory occur in a variety of other disciplines, including hydrology [11], [16] and geoscience [21].

2.2.1 ARFIMA Models

One way to model long memory time series is with an autoregressive fractionally integrated moving-average model, ARFIMA(p, d, q), model. The ARFIMA model is an extension of the ARMA model briefly discussed above, where p and q have the same meaning as in the ARMA model and d is the degree of fractional differencing, discussed below. An ARFIMA(p, d, q) model may be represented as:

$$\phi(B)y_t = \theta(B)(1-B)^{-d}\omega_t \tag{2.3}$$

where $\phi(B) = 1 - \phi_1 B - ... - \phi_p B^p$ is the pth order autoregressive polynomial and $\theta(B) = 1+\theta_1 B+...+\theta_q B^q$ is the qth order moving average polynomial. The parameter d is the long memory parameter, which describes *how long of a memory* the series y_t has, in some sense. We assume that $\phi(B)$ and $\theta(B)$ have no common roots. Now we can use the binomial expansion to rewrite

$$(1-B)^{-d} = \sum_{j=0}^{\infty} \eta_j B^j = \eta(B),$$

where

$$\eta_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} \tag{2.4}$$

and $\Gamma(\dot{)}$ denotes the gamma function defined as $\Gamma(y+1) = y\Gamma(y)$ for $y > 0$, [21] p. 269. The parameter d entirely determines the long memory properties of the model, so it is important to have good estimates for d. If $d = 0$, the stochastic process follows an ARMA(p, q) model. With a few additional assumptions, we have the existence a unique stationary solution to (2.3), which is is causal and invertible.

Theorem 2.2.1. *(Palma, [18]) Consider the ARFIMA process defined by* (2.3)*. Assume that the polynomials $\phi(z)$ and $\theta(z)$ have no common zeros and that $d \in (-1, \frac{1}{2})$. Then,*

(a) If the zeros of $\phi(z)$ lie outside the unit circle $\{z : |z| = 1\}$, then there is a unique stationary solution of (2.3) *given by*

$$y_t = \sum_{j=-\infty}^{\infty} \varphi_j \omega_{t-j} = \varphi(B)\omega_t, \tag{2.5}$$

where $\varphi(z) = (1 - z)^{-d}\theta(z)/\phi(z)$.

(b) If the zeros of $\phi(z)$ lie outside the closed unit disk, $\{z : |z| \leq 1\}$, then the solution, $\{y_t\}$, is causal.

(c) If the zeros of $\theta(z)$ lie outside the closed unit disk, $\{z : |z| \leq 1\}$, then the solution, $\{y_t\}$, is invertible.

For a proof of this theorem, see [18], p. 44. Note that if we had instead written the ARFIMA model as $\phi(B)(1 - B)^d y_t = \theta(B)\omega_t$, the solution may no longer be unique [18], p. 46.

From Theorem 2.2.1, we have a way to write infinite AR and infinite MA expansions of the stationary, causal, and invertible solution to an ARFIMA model. For such a solution, we have:

$$y_t = (1-B)^{-d}\phi(B)^{-1}\theta(B)\omega_t = \varphi(B)\omega_t \tag{2.6}$$

and

$$\omega_t = (1-B)^{d}\phi(B)\theta(B)^{-1}y_t = \pi(B)y_t. \tag{2.7}$$

The infinite MA expansion, (2.6), was of supreme importance in implementing our model. For an ARFIMA$(0,d,0)$ model, which is the process focused on in this thesis, we have explicit expressions for the infinite MA expansion's coefficients:

$$\varphi_j = \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)}, \qquad j \geq 1 \tag{2.8}$$

with initial condition $\psi_0 = 1$. Note that the φ_j come from equation (2.4) because $\phi(B) = \theta(B) = 1$ for an ARFIMA$(0,d,0)$ process. Here we stress that because of the hyperbolic decay characteristic of a long memory process, these φ_j decay very slowly to zero. One consequence of this is an inability to estimate the expression $y_t = \sum_{j=-\infty}^{\infty}\varphi_j\omega_{t-j}$ with a truncation at say, $m, \hat{y}_t = \sum_{j=-m}^{m}\varphi_j\omega_{t-j}, m \in \mathbb{Z}$. Even for very large m, this approximation would be off by a non-negligible amount [15], p. 1671. This fact will cause us to use alternate expressions in Chapters 4 and 5.

We also have an explicit expression for the autocorrelation function (ACF) of an ARFIMA$(0,d,0)$ process:

$$\rho(h) = \frac{\Gamma(1-d)\Gamma(h+d)}{\Gamma(d)\Gamma(1+h-d)}. \tag{2.9}$$

The ACF for a general ARFIMA(p,d,q) process is quite complicated [22] and will not be discussed here. The PACF for a general ARFIMA(p,d,q) process is even more difficult to express. For the interested reader, please see [18] p. 48-50. When

investigating (2.9), we see the hyperbolic decay referenced above. In the sample ACF of a time series, a sample correlation between the bounds $\pm 1.96 n^{-1/2}$ is said not to be significantly different from zero. In a short memory process, the correlation between observations decays very quickly into the insignificant range. However, for a long memory time series, the correlations remain significant for very large time lags because of the hyperbolic decay, contributing to the name *long memory time series*.

Recall that any stationary short memory time series can be modeled arbitrarily well by an ARMA(p, q) process. Here we present the corresponding theorem for long memory time series.

Theorem 2.2.2. *(Palma, [18]) Let* $\{y_t : t \in \mathbb{Z}\}$ *be a linear regular process satisfying* $y_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} = \psi(B)\epsilon_t$, *where* $\psi_0 = 1$ *and* $\sum_{j=1}^{\infty} \psi_j^2 < \infty$ *with strictly positive spectral density* f_y *satisfying* $f(\lambda) \sim |\lambda|^{-2d} f_2(|\lambda|)$. *Then, there exists an ARFIMA process with spectral density* f *such that for any* $\epsilon > 0$,

$$\left| \frac{f_y(\lambda)}{f(\lambda)} - 1 \right| < \epsilon,$$

uniformly for $\lambda \in [-\pi, \pi]$.

Because this theorem is proved in the frequency domain, it requires more background for understanding than provided here, so interested readers may find the proof in [18], p. 55.

2.3 Linear Regression with Autocorrelated Errors

Recall that the classical univariate linear model is of the form (1.1). It is common to suppose that ϵ_t is iid. However, in many cases, this is an inappropriate assumption. Indeed, if there is autocorrelation in the error, we may be able come up with better predictions. Thus, models arose assuming ϵ_t was an ARMA and more recently an ARFIMA process. The estimation procedure usually uses the least-squares approximation for β and fits an ARMA or ARFIMA model to the sample residuals. Then, β can be re-estimated after subtracting the fitted values from y_t. This procedure is repeated until the estimates converge. As mentioned previously, this is the most common approach when long memory is apparent in y_t. However, note that in this case, $var(y_t) = var(\epsilon_t)$, implying that any change in the variance of the response series is due to fluctuation in the error. While this may be true in some cases, it also makes sense that variation in the response should be linked to variance in the explanatory series. One answer to this problem is to suppose that β is in fact a stochastic process instead of a constant. If the series shows both long memory and short memory characteristics, β_t might be modeled as an ARMA process with ARFIMA errors. A model supposing β_t is long memory does not appear to have occurred in the literature, and will be explained in the next chapter.

2.4 Maximum Likelihood Estimation

Maximum likelihood estimation, or MLE, is a widely used method of estimation in statistics. The idea behind the MLE is that given a family of distribution functions with a vector of parameters θ, we find the most likely value of θ given a set of observations, $\{y_1, ..., y_n\}$. An example taken from "Mathematical Statistics with

Applications," [23], is as follows: suppose we choose two balls from a box containing three balls with color red or white. We do not know how many balls are red and white, but we choose two red balls without replacement. We know that the number of red balls in the box must have been two or three. If there are only two red balls, the probability of choosing those two balls is

$$\frac{\binom{2}{2}\binom{1}{0}}{\binom{3}{2}} = \frac{1}{3}.$$

However, if the number of red balls is actually three, then the probability of choosing two red balls from three red balls is one! Since it is much more likely that there are three red balls, given the random sample, we would hypothesize that the number of red balls is three. This is an intuitive example of the maximum likelihood logic. A more formal definition of maximum likelihood is given below.

Definition 2.4.1. : Method of Maximum Likelihood *[23] p. 477*
Suppose that the likelihood function depends on k parameters $\theta_1, ..., \theta_k$. *Choose as estimates those values of the parameters that maximize the likelihood* $L(y_1, ..., y_n|\theta_1, ..., \theta_k)$.

If the random sample consists of continuous random variables, then the likelihood function $L(\theta)$ is the joint density function of those random variables. Since $\ln(L(\theta))$, known as the log-likelihood function, is a monotonically increasing function of $L(\theta)$, $\ln(L(\theta))$ is maximized when $L(\theta)$ is maximized. The log-likelihood function has some nice properties in regards to numerical stability and derivative calculation and is therefore often used in practice instead of the likelihood function.

CHAPTER 3

STOCHASTIC PARAMETER REGRESSION FOR LONG MEMORY TIME SERIES

3.1 Model Specification

We would like to develop a model where the long memory component is not treated as a random accident, but an integral component of the series in question. To that end, we suppose that given an explanatory series, $\{z_t\}$, and a long memory response series, $\{y_t\}$, the relationships (1.5) and (1.6) hold. Because we have specified β_t to follow an ARFIMA$(0, d, 0)$ process, we are only concerned with a series $\{\beta_t\}$ that is fractionally integrated white noise, but our approach could be generalized to an ARFIMA(p, d, q) process in order to model both long and short memory characteristics. Our approach allows $\{\epsilon_t\}$ to take its classic linear regression role as iid errors. Note that the variance of y_t depends on the values of z_t and the variance of β_t as follows:

$$var(y_t) = var(\mu) + var(z_t\beta_t) + var(\alpha a_t) + var(\epsilon_t) \tag{3.1}$$

$$= z_t^2 var(\beta_t) + \sigma_\epsilon^2 = z_t^2\sigma_\omega^2 + \sigma_\epsilon^2. \tag{3.2}$$

Thus, as z_t increases, the variance of y_t increases. This gives more explanatory power to z_t and contributes to a better understanding of the behavior of y_t.

One example where large z_t values might contribute to more variance in y_t is

large inflation values on nominal interest rates. In the late 1970s, inflation had increased so much that the Federal Reserve made a sharp increase in interest rates to encourage deflation. Doing so caused much unrest in the economy, and given the different opinions over the effects of high inflation inside the Federal Reserve, interest rates might have decreased much more quickly, causing even more fluctuation than was seen at the time. As it was, the Federal Reserve still adjusted interest rates a considerable amount until inflation broke. In times where the economy is not under such duress, the inflation rate is under control and the Federal Reserve keeps interest rates relatively constant. That is, when inflation is high, it causes interest rates to be more variable. We plot this data set in Figure 3.1, using rates on three month Treasury bills as a measure of interest rates. This data set is not appropriate for our model as interest rates do not seem to display long memory, but the historical data supports the hypothesis that large inflation rates contribute to more variable interest rates. A plot of this data is shown in Figure 3.1. The data came from [6] and [8]. This example demonstrates the need for stochastic parameter regression models because of the increased variability of y_t for different values of z_t.

A pictorial example of a data set that could be well-explained by our model is shown in Figure 3.2 and Figure 3.3. Figure 3.2 shows the y_t and z_t time series plots on the same graph. The y_t series has an intercept around ten and the z_t series has an intercept close to zero. Note that for peaks in z_t values, y_t tends to be more variant. Furthermore, y_t appears to have local changes in mean, something the ARFIMA$(0, d, 0)$ model accounts for. Figure 3.3 shows a scatterplot of y_t and z_t. Notice that the variance of y_t increases for large values of z_t. This is because the variance of y_t is dependent on the values of z_t. Because we take $E(\beta_t) = 0$, the scatterplot is centered at zero. This particular series was generated by taking z_t to

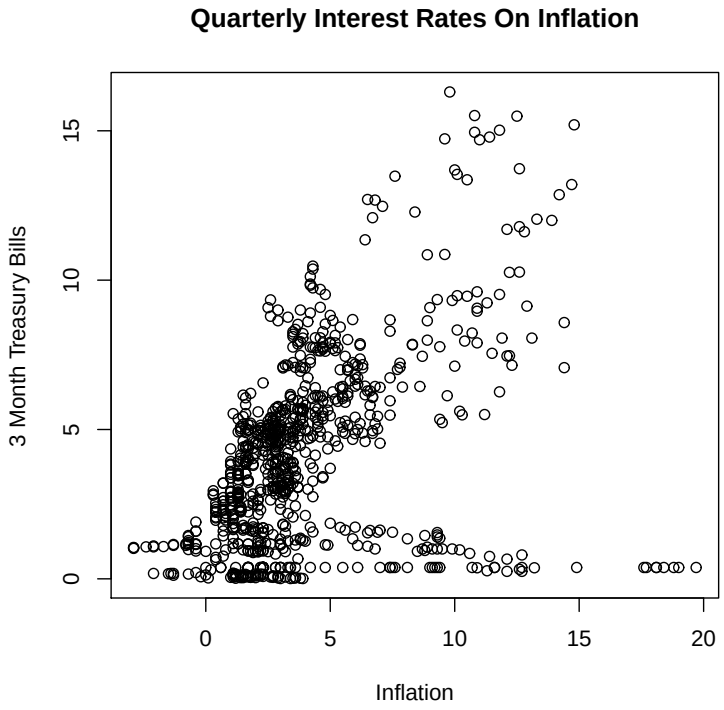

Figure 3.1: Interest Rates on Inflation

be a linear trend with AR(1) errors. That is, $z_t = \alpha_0 + \alpha_1 t + \xi_t$, where $\xi_t = \phi\xi_{t-1} + \eta_t$ and $\eta_t \sim$ iid $\mathrm{N}(0, \sigma_\eta^2)$. We let $\alpha_0 = 0, \alpha_1 = .05, \phi = .8$ and $\sigma_\eta^2 = 1$. We also set $\mu = 10$, $\alpha = 0$, $\sigma_\epsilon = 1.5$ and $\sigma_\omega = 1$. Here, β_t is generated from a zero-mean ARFIMA$(0, d, 0)$ process with $d = .4$. Notice that at first glance, it may not be obvious that y_t and z_t are linearly related. This means that data that should be theoretically related but do not appear to be so in practice may be explained by a model similar to ours. We definitely notice heteroscedasticity, with larger variance for larger values of z_t. In summary, our model has the flexibility to take into account long memory influences in a time series, like local changes in mean, and is able to explain increasing variance for larger values of the explanatory variable.

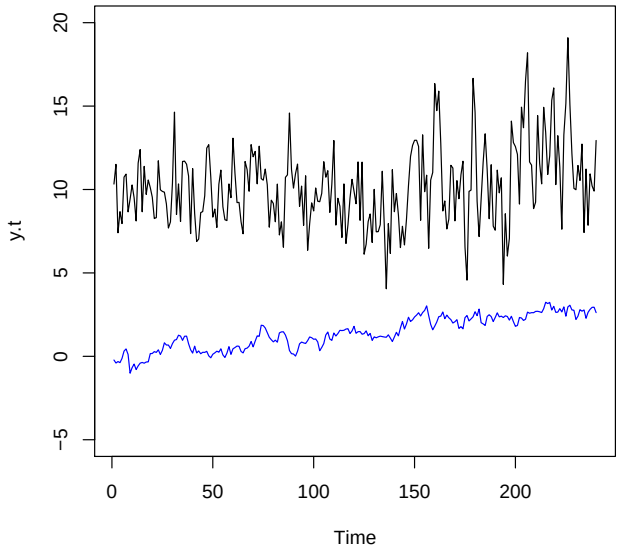

Figure 3.2: Simulated Time Series Plot

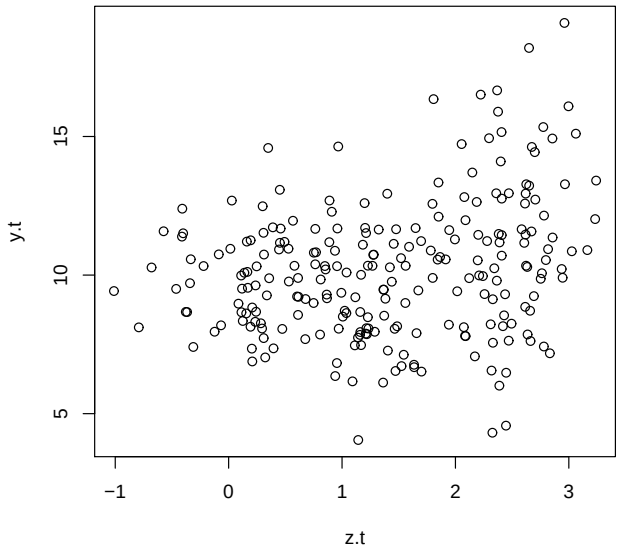

Figure 3.3: Simulated Scatterplot

3.2 Estimation Methods

In the next two chapters, we introduce our modifications to the Kalman Filter and Innovations Algorithm, which allow us to use maximum likelihood estimation for our parameters. We have four different techniques: a full Kalman Filter, a truncated Kalman Filter, a full Innovations Algorithm, and a truncated Innovations Algorithm. The maximum likelihood function has been investigated for these methods, so we can adapt maximum likelihood estimation to our purposes.

One extension of the likelihood function is Whittle Estimation, often used to estimate parameters when the observations show long memory and sometimes used by built-in estimation routines in statistical programs [2], p. 363. Whittle Estimation is a frequency domain approach and is a result of the fact that calculating the exact likelihood function of an ARFIMA process is computationally demanding. However, Whittle Estimation will not be used here as Chan and Palma [5] came up with another way around using the exact likelihood function for estimation:

Theorem 3.2.1. *(Chan and Palma, [5], Theorem 2.2)*
Let $\{y_1, ..., y_n\}$ *be a finite sample of an ARFIMA*(p, d, q) *process. If* Ω_1 *is the variance of the initial state* X_1 *for the infinite-dimensional representation, then the computation of the exact likelihood function depends only on the first* n *components of the Kalman equations.*

In other words, we can obtain the exact likelihood function for an ARFIMA process with a finite sample of n. For the proof, see [5]. This approach is what we use in our estimation methods. Because we have assumed that $\{\epsilon_t\}$ and $\{\omega_t\}$ are Gaussian processes, maximum likelihood estimation holds theoretical merits for estimating the parameters of our model. We then use the expressions derived below to obtain the

innovations for each method and use those innovations to compute the maximum likelihood function. Each likelihood function shown is equivalent, but the different forms are included for notation purposes. We then use a Newton-Raphson procedure with the "optim" function from R to obtain maximum likelihood estimates. For the rest of the paper, we will suppose that the response variable y_t and the covariate z_t have the relationship defined in (1.5) and (1.6).

CHAPTER 4

THE KALMAN FILTER

The Kalman Filter uses a state space representation to obtain the prediction expression with minimum mean square error in the class of linear estimators [21], p. 325. The Kalman Filter is a special application of state space models, so before presenting the Kalman Filter, it would be prudent to begin with a discussion of said state space models.

4.1 State Space Representation

State space models, also known as Dynamic Linear Models, are based on the premise that the random variable of interest, the state variable, is not directly observed; rather, we observe some linear combination of the state variable plus noise. The basic observation equation is

$$y_t = G_t X_t + \epsilon_t, \tag{4.1}$$

and the state equation is

$$X_t = F X_{t-1} + W_t. \tag{4.2}$$

We assume that $W_t \sim$ iid $N(0, Q)$ and $\epsilon_t \sim$ iid $N(0, R)$. While it is not a necessary assumption, we keep things simple by supposing that $\{W_t\}$ and $\{\epsilon_t\}$ are independent. We also assume that $X_0 \sim N(\mu_0, \Omega_0)$. Here, X_t and W_t are $p \times 1$ vectors, F_t is a $p \times p$

matrix, y_t and ϵ_t are $q \times 1$ vectors, G_t is $q \times p$ matrix, and q is allowed to be less than, equal to, or greater than p. For the purpose of broadening the model, we can also include inputs in the basic state and observation equation in the following way:

$$y_t = G_t X_t + \Gamma u_t + \epsilon_t, \tag{4.3}$$

$$X_t = F X_{t-1} + \zeta u_t + W_t. \tag{4.4}$$

At first glance, the requirement that the state variable be well-fitted by an AR(1) model may seem ill-suited to many situations. However, the state equation actually allows for great flexibility. For example, in the univariate case, suppose we wish to measure state β_t, which is actually better fitted by an AR(m) model. We may then write the state equation in the following way:

$$\begin{bmatrix} \beta_t \\ \beta_{t-1} \\ \vdots \\ \beta_{t-m+1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_m \\ 1 & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \beta_{t-1} \\ \beta_{t-2} \\ \vdots \\ \beta_{t-m} \end{bmatrix} + \begin{bmatrix} \omega_t \\ 0 \\ \vdots \\ 0, \end{bmatrix}, \tag{4.5}$$

where $X_t = (\beta_t, \beta_{t-1}, ..., \beta_{t+m-1})^T$. The observation equation becomes

$$y_t = \begin{bmatrix} G_t & 0 & \cdots & 0 \end{bmatrix} X_t + \epsilon_t. \tag{4.6}$$

In our case, we wish to measure the ARFIMA$(0, d, 0)$ state β_t, which follows an AR(∞) model, specified in Theorem 2.2.1. It may be apparent to careful readers that we would need an infinite state space model for our case. That intuition is correct, as confirmed by Chan and Palma in [5], but we will demonstrate a way

around this issue in Section 4.3. For more details on state space models, see [21]. With this groundwork laid out, we will consider the Kalman Prediction Equations.

4.2 Kalman Prediction

The Kalman Filter prediction equations are as follows:

Theorem 4.2.1. *(Shumway and Stoffer, [21]) For the state-space model specified in* (4.3) *and* (4.4)*, with initial conditions* $E(X_0) = \mu_0$ *and* $var(X_0) = \Omega_0$*, the Kalman prediction equations are:*

$$\begin{aligned} X_{t|t-1} &= F_t X_{t-1|t-1} + \zeta u_t, \\ \Omega_{t|t-1} &= F_t \Omega_{t-1|t-1} F_t^T + Q, \\ X_{t|t} &= X_{t|t-1} + K_t(y_t - G_t X_{t|t-1} - \Gamma u_t), \\ \Omega_{t|t} &= [I - K_t G_t]\Omega_{t|t-1}, \end{aligned}$$

where I *is the identity matrix and*

$$K_t = \Omega_{t|t-1} G_t^T [G_t \Omega_{t|t-1} G_t^T + R]^{-1}.$$

K_t is called the Kalman gain and is, in some sense, a measure of how well $X_{t|t-1}$ actually predicted $X_{t|t}$ [21]. In addition to these equations, we also calculate the corresponding innovations, which will be used to calculate the maximum likelihood estimate. The innovations are:

$$\upsilon_t = y_t - E(y_t|y_{t-1}, y_{t-2}, ...) = y_t - G_t X_{t|t-1} - \Gamma u_t, \tag{4.7}$$

and their corresponding variance-covariances matrices are:

$$\Sigma_t = var(y_t - G_t x_{t|t-1} - \Gamma u_t) = var(G_t X_t + \epsilon_t - G_t X_{t|t-1}) = G_t \Omega_{t|t-1} G_t^T + R. \quad (4.8)$$

These innovations are uncorrelated, which allows us to use them for the innovations form of the maximum likelihood function, discussed below. As previously mentioned, the Kalman estimates are those with the minimum mean square error in the class of linear estimators. Another advantage of the Kalman equations is that once we have calculated $X_{t|t-1}$, we can use that and y_t to calculate $X_{t+1|t}$ instead of having to recalculate all the estimates over again. This makes the Kalman equations very useful for real-time data. The final advantage is a bit more specific to this paper; the Kalman Prediction equations give us expressions for the conditional expectation of β_t, which is useful for prediction purposes.

4.3 Full Kalman Filter

We would like to find the Kalman Prediction expressions for our model, represented in (1.5) and (1.6). This presents a challenge; in 1998, Chan and Palma [5] proved that there is no finite dimensional state space model for a long memory ARFIMA process. However, we can extract the exact likelihood function of a long memory process from a finite dimensional expression following Chan and Palma's approach [5], by way of Theorem 3.2.1 from Chapter 3.

Note that, in our notation, Ω_1 is called $\Omega_{0|0}$ and $X_1 = X_{0|0}$. Furthermore, because our state space equations are different from those in [5], the likelihood actually depends on the first $n+1$ equations. Therefore, if our state space model is accurate for

the first $n+1$ rows, our likelihood function should be exact. To use Theorem 3.2.1, we will first state and prove a lemma used in the specification of our state space model.

Lemma 4.3.1. *Let $\{\beta_t\}$ be a zero-mean stationary ARFIMA$(0, d, 0)$ process with $d \in (-1, 0.5)$. For $i = 1, \ldots, m$, we have*

$$\beta_{t+i|t} = \sum_{k=i}^{\infty} \varphi_k \omega_{t+i-k}.$$

Proof. We will prove this for $i = 1$. Since $\beta_t = (1-B)^{-d}\omega_t = \sum_{k=0}^{\infty} \varphi_k \omega_{t-k}$ with $\varphi_k = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)}$, we have

$$\begin{aligned}
\beta_{t+1|t} &= E(\beta_{t+1}|\beta_t, \beta_{t-1}, \cdots) \\
&= E\left(\sum_{k=0}^{\infty} \varphi_k \omega_{t+1-k} \middle| \beta_t, \beta_{t-1}, \cdots\right) \\
&= E(\omega_{t+1}|\beta_t, \beta_{t-1}, \cdots) + E\left(\sum_{k=1}^{\infty} \varphi_k \omega_{t+1-k} \middle| \beta_t, \beta_{t-1}, \cdots\right) \\
&= \sum_{k=1}^{\infty} \varphi_k \omega_{t+1-k}
\end{aligned}$$

because $E(\omega_{t+1}|\beta_t, \beta_{t-1}, \cdots) = E(\omega_{t+1}) = 0$ and $E(\sum_{k=1}^{\infty} \varphi_k \omega_{t+1-k}|\beta_t, \beta_{t-1}, \cdots) = \sum_{k=1}^{\infty} \varphi_k \omega_{t+1-k}$, as $\sum_{k=1}^{\infty} \varphi_k \omega_{t+1-k}$ only involves data from the first t observations. □

Now we specify a new state space specification used for our model. We define the observation and state equations (4.3) and (4.4) as follows:

Theorem 4.3.2. *The state space representation* (4.3) *and* (4.4) *correctly models* (1.5) *and* (1.6)*, where*

$$X_t = \begin{bmatrix} \beta_t \\ \beta_{t+1|t} \\ \vdots \\ \beta_{t+n|t} \end{bmatrix}, F = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, W_t = \begin{bmatrix} 1 \\ \varphi_1 \\ \vdots \\ \varphi_n \end{bmatrix} \omega_t, G_t = \begin{bmatrix} z_t & 0 & \cdots & 0 \end{bmatrix}.$$

Here, $\beta_{t+i|t} = E(\beta_{t+i}|\beta_t, \beta_{t-1}, ...)$ *and* $\{W_t\} \sim \text{iid N}(0, Q)$ *where*

$$Q = \sigma_\omega^2 \begin{bmatrix} 1 & \varphi_1 & \cdots & \varphi_n \\ \varphi_1 & \varphi_1^2 & \cdots & \varphi_1\varphi_n \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_n & \varphi_n\varphi_1 & \cdots & \varphi_n^2 \end{bmatrix}$$

If there is a drift in the data of α*, then we set* $\Gamma = \begin{bmatrix} \mu & \alpha \end{bmatrix}$ *and* $u_t = \begin{bmatrix} 1 \\ t \end{bmatrix}$*. Otherwise,* $\Gamma = \mu$ *and* $u_t = 1$*. Now, assume* $X_0 \sim \text{N}(X_{0|0}, \Omega_{0|0})$*. We set* $X_{0|0} = 0$ *since* $E(\beta_t) = 0$ *and have*

$$\Omega_{0|0} = \begin{bmatrix} \gamma_\beta(0) & \gamma_\beta(1) & \cdots & \gamma_\beta(n) \\ \gamma_\beta(1) & & & \\ \vdots & & \mathcal{T} & \\ \gamma_\beta(n) & & & \end{bmatrix} \tag{4.9}$$

where $\gamma_\beta(h)$ *is the ACVF of* $\{\beta_t\}$*, and the* $n \times n$ *matrix* $\mathcal{T}$ *has*

$$\mathcal{T}_{ij} = \begin{cases} \gamma_\beta(j-i) - \sigma_\omega^2 \sum_{k=0}^{i-1} \varphi_k\varphi_{k+(j-i)}, & i \le j \\ \mathcal{T}_{ji}, & i > j \end{cases} \tag{4.10}$$

Proof. First, we will show that (4.3) and (4.4) are equivalent to (1.5) and (1.6). Showing the equivalence of (1.5) and (4.3) is straightforward as $G_t X_t = \beta_t z_t$. Now we focus on the state equation, (4.4). For this, note that the state equation (4.4) is equivalent to the following system of equations:

$$\begin{aligned}
\beta_t &= \beta_{t-1+1|t-1} + \omega_t, \\
\beta_{t+1|t} &= \beta_{t-1+2|t-1} + \varphi_1 \omega_t, \\
&\vdots \\
\beta_{t+n-1|t} &= \beta_{t-1+n|t-1} + \varphi_{n-1} \omega_t, \\
\beta_{t+n|t} &= \varphi_n \omega_t.
\end{aligned}$$

We will show the first n equations hold true. For this, we use Lemma 4.3.1 to obtain:

$$\beta_t = \sum_{k=0}^{\infty} \varphi_k \omega_{t-k} = \omega_t + \sum_{k=1}^{\infty} \varphi_k \omega_{t-k} = \omega_t + \sum_{k=1}^{\infty} \varphi_k \omega_{t-1+1-k} = \omega_t + \beta_{t-1+1|t-1},$$

which is equivalent to the first equation above. Next, Lemma 4.3.1 gives, for $i = 1, \ldots, n-1$,

$$\begin{aligned}
\beta_{t+i|t} &= \sum_{k=i}^{\infty} \varphi_k \omega_{t+i-k} \\
&= \varphi_i \omega_t + \sum_{k=i+1}^{\infty} \varphi_k \omega_{t-1+i+1-k} \\
&= \varphi_i \omega_t + \beta_{t-1+i+1|t-1},
\end{aligned}$$

which corresponds to the second through nth equations. The last equation is not true, but for large n, this would be appreciably negligible as shown in [5].

Now, for the initial condition on $\Omega_{0|0}$ in (4.9), note that the stationarity of $\{X_t\}$ gives

$$\begin{aligned}
\Omega_{0|0} &= Var(X_0) \\
&= Var(X_t) \\
&= \begin{pmatrix} Var(\beta_t) & Cov(\beta_t, \beta_{t+1|t}) & \cdots & Cov(\beta_t, \beta_{t+n|t}) \\ Cov(\beta_{t+1|t}, \beta_t) & Var(\beta_{t+1|t}) & \cdots & Cov(\beta_{t+1|t}, \beta_{t+n|t}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\beta_{t+n|t}, \beta_t) & Cov(\beta_{t+n|t}, \beta_{t+1|t}) & \cdots & Var(\beta_{t+n|t}) \end{pmatrix}.
\end{aligned}$$

First, we consider the following covariances: for $j = 1, \ldots, n$,

$$\begin{aligned}
Cov(\beta_t, \beta_{t+j|t}) &= Cov\left(\sum_{k=0}^{\infty} \varphi_k \omega_{t-k}, \sum_{k=j}^{\infty} \varphi_k \omega_{t+j-k}\right) \\
&= \sigma_\omega^2 \sum_{k=0}^{\infty} \varphi_k \varphi_{k+j} \\
&= \sigma_\omega^2 \sum_{k=0}^{\infty} \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)} \frac{\Gamma(k+j+d)}{\Gamma(d)\Gamma(k+j+1)} \\
&= \sigma_\omega^2 \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} \sum_{k=0}^{\infty} \frac{\Gamma(k+d)\Gamma(k+j+d)\Gamma(j+1)}{\Gamma(k+1)\Gamma(d)\Gamma(j+d)\Gamma(k+j+1)} \\
&= \sigma_\omega^2 \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} F(d, j+d; j+1; 1) \\
&= \sigma_\omega^2 \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} \frac{\Gamma(j+1)\Gamma(1-2d)}{\Gamma(j+1-d)\Gamma(1-d)} \\
&= \gamma_\beta(j).
\end{aligned}$$

Here, $F(x, y; v; w)$ is the hypergeometric function. Second, consider, for $i \leq j$, $i, j = 1, \ldots, n$,

$$\begin{aligned}
\tau_{i,j} &= Cov(\beta_{t+i|t}, \beta_{t+j|t}) \\
&= Cov\left(\sum_{k=i}^{\infty} \varphi_k \omega_{t+i-k}, \sum_{k=j}^{\infty} \varphi_k \omega_{t+j-k}\right) \\
&= \sigma_\omega^2 \sum_{k=i}^{\infty} \varphi_k \varphi_{k+(j-i)} \\
&= \sigma_\omega^2 \left(\sum_{k=0}^{\infty} \varphi_k \varphi_{k+(j-i)} - \sum_{k=0}^{i-1} \varphi_k \varphi_{k+(j-i)}\right) \\
&= \gamma_\beta(j-i) - \sigma_\omega^2 \sum_{k=0}^{i-1} \varphi_k \varphi_{k+(j-i)}.
\end{aligned}$$

These results verify our expression of $\Omega_{0|0}$ in (4.9). □

Now that we have expressions for the innovations from the Kalman Filter, we use these innovations to compute the likelihood function:

$$L(\mu, \alpha, \sigma_\epsilon, d, \sigma_\omega) = (2\pi)^{-n/2} \left(\prod_{j=1}^{n} \det \Sigma_t\right)^{-1/2} \exp\left[-\frac{1}{2}\sum_{j=1}^{n} \upsilon_t^T \Sigma_t^{-1} \upsilon_t\right]. \tag{4.11}$$

Since, in our case, Σ_t and ν_t are scalars, this makes the log-likelihood function:

$$\ell(\mu, \alpha, \sigma_\epsilon, d, \sigma_\omega) = L(\mu, \alpha, \sigma_\epsilon, d, \sigma_\omega) = -\frac{1}{2}\sum_{j=1}^{n} \ln \Sigma_t - \frac{1}{2}\sum_{j=1}^{n} \frac{\upsilon_t^2}{\Sigma_t}. \tag{4.12}$$

when the constant term in (4.11) is ignored. This is also the likelihood function expression for the Truncated Kalman Filter, detailed in the next section.

4.4 Truncated Kalman Filter

Thus far, we have developed a state space model of dimension $(n+1) \times (n+1)$. In comparison to an infinite dimensional state space model, this is good news for

computer estimation. However, using the Full Kalman Filter method described above can be cumbersome for large n. One might be tempted to simply cut off the Full Kalman Filter expressions, but because of the nature of a long memory time series, simply cutting off the MA(∞) expression leads to non-negligible errors, as discussed in Chapter 2. To that end, we would like to introduce a Truncated Kalman Filter of dimension $m \times m$ where m is considerably less than n. For that purpose, we turn our attention to the series $\delta_t = \beta_t - \beta_{t-1}$.

Lemma 4.4.1. *The MA(∞) representation for the differenced ARFIMA$(0, d, 0)$ process is*

$$\delta_{t+i} = \sum_{k=0}^{\infty} \psi_k \omega_{t+i-k}, \tag{4.13}$$

where

$$\psi_k = \varphi_k - \varphi_{k-1}$$

with initial condition $\varphi_{-1} = 0$.

Proof. We re-express β_t as an MA(∞) process: for $d \in (-1, 0.5)$,

$$\beta_t = (1 - B)^{-d} \omega_t = \sum_{k=0}^{\infty} \varphi_k \omega_{t-k},$$

with φ_j as in (2.8) [18], p.47. The differenced process, δ_t, is expressed as follows:

$$\delta_t = \beta_t - \beta_{t-1} = \sum_{k=0}^{\infty} \varphi_k \omega_{t-k} - \sum_{\ell=0}^{\infty} \varphi_\ell \omega_{t-1-\ell} = \sum_{k=0}^{\infty} \psi_k \omega_{t-k},$$
$$\delta_{t+1} = \beta_{t+1} - \beta_t = \sum_{k=0}^{\infty} \varphi_k \omega_{t+1-k} - \sum_{\ell=0}^{\infty} \varphi_\ell \omega_{t-\ell} = \sum_{k=0}^{\infty} \psi_k \omega_{t+1-k},$$

where $\psi_k = \varphi_k - \varphi_{k-1}$ with the initial condition $\varphi_{-1} = 0$. In general, we can deduce (4.13): for $i = 0, 1, \ldots,$

$$\delta_{t+i} = \sum_{k=0}^{\infty} \psi_k \omega_{t+i-k},$$

which is an MA(∞) representation for the differenced process $\delta_{t+i} = \beta_{t+i} - \beta_{t+i-1}$. □

One of the solutions Chan and Palma [5] came up with for the problem of being unable to write a finite state space model was to consider the differenced series. Indeed, since $\psi_k \to 0$ faster than $\varphi_k \to 0$, Chan and Palma truncate this expression after m. That is,

$$\delta_{t+i} \approx \sum_{k=0}^{m} \psi_k \omega_{t+i-k} \tag{4.14}$$

Lemma 4.4.2. *The conditional expectation of δ_{t+i} given $\beta_t, ..., \beta_1$ has the expression*

$$\begin{aligned} \delta_{t+i|t} &= E(\delta_{t+i} | \beta_t, ..., \beta_1) \\ &\approx \sum_{k=i}^{m} \psi_k \omega_{t+i-k} \end{aligned}$$

Proof. We consider the conditional expectation of the differenced process $\{\delta_t\}$ as follows. Using the MA(∞) expression (4.13) and the MA(m) truncation expression (4.14), we obtain:

$$
\begin{aligned}
\delta_{t+1|t} &= E\left(\sum_{k=0}^{\infty}\psi_k\omega_{t+1-k}\middle|\beta_t,\beta_{t-1},\cdots\right)\\
&\approx E\left(\sum_{k=0}^{m}\psi_k\omega_{t+1-k}\middle|\beta_t,\beta_{t-1},\cdots\right)\\
&= E(\omega_{t+1}|\beta_t,\beta_{t-1},\cdots)+E\left(\sum_{k=1}^{m}\psi_k\omega_{t+1-k}\middle|\beta_t,\beta_{t-1},\cdots\right)\\
&= \sum_{k=1}^{m}\psi_k\omega_{t+1-k},\\
\delta_{t+2|t} &= E\left(\sum_{k=0}^{\infty}\psi_k\omega_{t+2-k}\middle|\beta_t,\beta_{t-1},\cdots\right)\\
&\approx E\left(\sum_{k=0}^{m}\psi_k\omega_{t+2-k}\middle|\beta_t,\beta_{t-1},\cdots\right)\\
&= E(\omega_{t+2})+\psi_1E(\omega_{t+1})+E\left(\sum_{k=2}^{m}\psi_k\omega_{t+2-k}\middle|\beta_t,\beta_{t-1},\cdots\right)\\
&= \sum_{k=2}^{m}\psi_k\omega_{t+2-k},
\end{aligned}
$$

which inductively results in the truncated expression in Lemma 4.4.2. □

If we suppose that m in the approximation equation is such that (4.14) is within a reasonable margin of error, we will take

$$\delta_{t+i}=\sum_{k=0}^{m}\psi_k\omega_{t+i-k}$$

From Lemma 4.4.2, we get:

$$\delta_{t-1+i|t-1}\approx\sum_{k=i}^{m}\psi_k\omega_{t-1+i-k} \tag{4.15}$$

Now we will present our long memory SPR model via the following truncated

state space model representation.

Theorem 4.4.3. *The state space representation* (4.3) *and* (4.4) *correctly models* (1.5) *and* (1.6)*, where*

$$X_t = \begin{bmatrix} \beta_t \\ \delta_{t+1|t} \\ \vdots \\ \delta_{t+m|t} \end{bmatrix}, F = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, W_t = \begin{bmatrix} 1 \\ \psi_1 \\ \vdots \\ \psi_m \end{bmatrix}, G_t = \begin{bmatrix} z_t & 0 & \cdots & 0 \end{bmatrix}.$$

Here, $W_t \sim$ iid $\mathrm{N}(0, Q)$ *with*

$$Q = \sigma_\omega^2 \begin{bmatrix} 1 & \psi_1 & \cdots & \psi_m \\ \psi_1 & \psi_1^2 & \cdots & \psi_1\psi_m \\ \vdots & \vdots & \ddots & \vdots \\ \psi_m & \psi_m\psi_1 & \cdots & \psi_m^2 \end{bmatrix}.$$

We now assume that $X_0 \sim \mathrm{N}(x_{0|0}, \Omega_{0|0})$. Since $E(\beta_t) = 0$, we find $X_{0|0} = 0$ and

$$\Omega_{0|0} = \begin{bmatrix} \sigma_\omega^2 \frac{\Gamma(1-2d)}{\Gamma(1-d)^2} & \sigma_\omega^2 \sum_{k=0}^{m-1} \varphi_k\psi_{k+1} & \cdots & \sigma_\omega^2\varphi_0\psi_m \\ \sigma_\omega^2 \sum_{k=0}^{m-1} \varphi_k\psi_{k+1} & & & \\ \vdots & & \mathcal{T} & \\ \sigma_\omega^2\varphi_0\psi_m & & & \end{bmatrix},$$

where $\mathcal{T}$ is an $m \times m$ matrix with

$$\mathcal{T}_{ij} = \begin{cases} \sigma_\omega^2 \sum_{k=i}^{m-1} \psi_k \psi_{k+(j-i)}, & i \leq j; \\ \mathcal{T}_{ji}, & i > j. \end{cases}$$

Proof. First we will show that (4.3) and (4.4) are equivalent to (1.5) and (1.6). Showing the equivalence of (1.5) and (4.3) is straightforward as $G_t X_t = \beta_t z_t$. Now we focus on the state equation, (4.4). For this, note that the state equation (4.4) is equivalent to the following system of equations:

$$\begin{aligned} \beta_t &= \beta_{t-1} + \delta_{t|t-1} + \omega_t, \\ \delta_{t+1|t} &= \delta_{t-1+2|t-1} + \psi_1 \omega_t, \\ &\vdots \\ \delta_{t+m-1|t} &= \delta_{t-1+m|t-1} + \psi_{m-1} \omega_t, \\ \delta_{t+m|t} &= \psi_m \omega_t. \end{aligned}$$

We will show these $m+1$ equations hold true (within an approximation of m). First, we use (4.13) and (4.15) as follows:

$$\beta_t - \beta_{t-1} = \delta_t \approx \sum_{k=0}^{m} \psi_k \omega_{t-k} = \omega_t + \sum_{k=1}^{m} \psi_k \omega_{t-1+1-k} = \omega_t + \delta_{t-1+1|t-1},$$

which is equivalent to the first equation above. Second, we use Lemma 4.4.2 to obtain, for $i = 1, \ldots, m-1$,

$$\delta_{t+i|t} \approx \sum_{k=i}^{m} \psi_k \omega_{t+i-k} = \psi_i \omega_t + \sum_{k=i+1}^{m} \psi_k \omega_{t-1+i+1-k} = \psi_i \omega_t + \delta_{t-1+i+1|t-1},$$

which corresponds to the second through mth equations. Last, using Lemma 4.4.2

gives

$$\delta_{t+m|t} \approx \sum_{k=m}^{m} \psi_k \omega_{t+m-k} = \psi_m \omega_t$$

as required in the $(m+1)$th equation of (4.4).

Now we will show that the expression of $\Omega_{0|0}$ is correct. Because $\{X_t\}$ is stationary, we have that

$$\begin{aligned}
\Omega_{0|0} &= Var(X_0) \\
&= Var(X_t) \\
&= \begin{pmatrix}
Var(\beta_t) & Cov(\beta_t, \delta_{t+1|t}) & \cdots & Cov(\beta_t, \delta_{t+m|t}) \\
Cov(\delta_{t+1|t}, \beta_t) & Var(\delta_{t+1|t}) & \cdots & Cov(\delta_{t+1|t}, \delta_{t+m|t}) \\
\vdots & \vdots & \ddots & \vdots \\
Cov(\delta_{t+m|t}, \beta_t) & Cov(\delta_{t+m|t}, \delta_{t+1|t}) & \cdots & Var(\delta_{t+m|t})
\end{pmatrix}.
\end{aligned}$$

We also know that the stationary ARFIMA$(0, d, 0)$ process $\{\beta_t\}$ has variance

$$Var(\beta_t) = \gamma_\beta(0) = \sigma_\omega^2 \frac{\Gamma(1-2d)}{\Gamma(1-d)^2},$$

which verifies the $(1,1)$th element of $\Omega_{0|0}$. From Lemma 4.4.2 and the infinite MA expansion of β_t, we have

$$\begin{aligned}
Cov(\beta_t, \delta_{t+j|t}) &\approx Cov\left(\sum_{k=0}^{\infty} \varphi_k \omega_{t-k}, \sum_{k=j}^{m} \psi_k \omega_{t+j-k}\right) \\
&= \sigma_\omega^2 \sum_{k=0}^{m-j} \varphi_k \psi_{k+j},
\end{aligned}$$

for $j = 1, \ldots, m$, which verifies the remaining elements in the first row of $\Omega_{0|0}$. Finally, for $i \leq j$, we have

$$
\begin{aligned}
\mathcal{T}_{i,j} &= Cov(\delta_{t+i|t}, \delta_{t+j|t}) \\
&\approx Cov\left(\sum_{k=i}^{m} \psi_k \omega_{t+i-k}, \sum_{k=j}^{m} \psi_k \omega_{t+j-k}\right) \\
&= \sigma_\omega^2(\psi_i\psi_{i+(j-i)} + \psi_{i+1}\psi_{i+1+(j-i)} + \cdots + \psi_{m-(j-i)}\psi_{m-(j-i)+(j-i)}) \\
&= \sigma_\omega^2 \sum_{k=i}^{m-(j-i)} \psi_k\psi_{k+(j-i)},
\end{aligned}
$$

which verifies the expression for $\mathcal{T}$. □

CHAPTER 5

THE INNOVATIONS ALGORITHM

The Innovations Algorithm and the Durbin-Levinson Algorithm are both one-step prediction recursion schemes. The Durbin-Levinson Algorithm can have smaller variance, but has the requirement that the series it is predicting be stationary. The Innovations Algorithm, however, only requires the series have a finite second moment, meaning it can be applied directly to a non-stationary series. This is important because one of the advantages of our model is the ability to take into account dynamic variance patterns. Thus, the technique used to predict the series must also not require stationarity.

5.1 Best Linear Predictor

The Innovations Algorithm stems from applications of the best linear predictor. The best linear predictor of random variable is defined as the linear combination of past values that has minimum variance [2], p. 63. If we denote the best linear predictor of X_{n+h} given n past values as P_nX_{n+h}, then the best linear predictor of a stationary time series given $X_1, X_2, ..., X_n$ is

$$P_nX_{n+h} = \mu + \sum_{i=1}^{n} a_i(X_{n+1-i} - \mu), \tag{5.1}$$

where $a = (a_1, ..., a_n)^T$ satisfies

$$\Gamma_n a = \gamma_n(h), \tag{5.2}$$

$\Gamma_n = [\gamma_X(i-j)]_{i,j=1}^n$, and $\gamma_n(h) = (\gamma_X(h), \gamma_X(h+1), ..., \gamma_X(h+n-1))^T$.

The following properties of the best linear predictor are useful [2], p. 65:

1. $E(X_{n+h} - P_n X_{n+h})^2 = \gamma_X(0) - a^T \gamma_n(h)$
2. $E(X_{n+h} - P_n X_{n+h}) = 0$
3. $E[(X_{n+h} - P_n X_{n+h}) X_j] = 0$, for $j = 1, 2, ..., n$

Thus, the prediction error is uncorrelated with the observations. This result will be used for obtaining the likelihood function via the Innovations Algorithm.

5.2 Innovations Algorithm

To introduce the Innovations Algorithm, we will follow Brockwell and Davis in their approach and notation [2], p. 71-73. Now, suppose X_t is a zero-mean time series with finite second moment and set $\kappa(i,j) = E(X_i X_j)$. Let

$$\hat{X}_t = \begin{cases} 0, & t = 1 \\ P_{t-1} X_t, & t = 2, ..., n \end{cases}$$

and

$$\nu_t = E(X_{t+1} - P_t X_{t+1})^2.$$

We will define the *innovations* as $U_t = X_t - \hat{X}_t$. Notice that the ν_t are the corresponding mean square prediction errors. With this notation, we can rewrite the above as $U = AX$ where $U = (U_1, ...U_n)^T, X = (X_1, ..., X_n)^T$ and

$$A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{11} & 1 & 0 & \cdots & 0 \\ a_{22} & a_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ a_{n-1,n-1} & a_{n-1,n-2} & a_{n-1,n-3} & \cdots & 1 \end{bmatrix}.$$

If $\{X_t\}$ is stationary, this simplifies with $a_{ij} = a_j$ from (5.2) with $h = 1$. A is invertible as its determinant is one. Because A is lower triangular, $A^{-1} := C$ has the form:

$$C = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \theta_{11} & 1 & 0 & \cdots & 0 \\ \theta_{22} & \theta_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \cdots & 1 \end{bmatrix}$$

Now, since $U_t = X_t - \hat{X}_t$, we have $\hat{X} = X - U, (\hat{X} = (\hat{X}_1, ..., \hat{X}_n)^T)$ and because $U = AX$ we have, $\hat{X} = CU - U = \Theta(X - \hat{X})$ where

$$\Theta = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ \theta_{11} & 0 & 0 & \cdots & 0 \\ \theta_{22} & \theta_{21} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \cdots & 0 \end{bmatrix},$$

This relationship can be rewritten as

$$\hat{X}_{t+1} = \begin{cases} 0, & t = 0 \\ \sum_{j=1}^{n} \theta_{tj}(X_{t+1-j} - \hat{X}_{t+1-j}), & t = 1, ..., n \end{cases}$$

Once we find the θ_{ij}, then we can recursively compute the one-step ahead prediction. The Innovations Algorithm below is a recursive scheme to find the θ_{ij}, the one-step ahead prediction, and the corresponding prediction error.

Definition 5.2.1. The Innovations Algorithm *(Brockwell and Davis, [2])*

The coefficients $\theta_{n1}, .., \theta_{nn}$ *can be computed recursively from the equations*

$\nu_0 = \kappa(1, 1)$,

$\theta_{n,n-k} = \nu_k^{-1}(\kappa(n + 1, k + 1) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}v_j), \quad 0 \leq k < n$

and

$\nu_n = \kappa(n + 1, n + 1) - \sum_{j=1}^{n-1} \theta_{n,n-j}^2 \nu_j$.

5.3 Full Innovations Algorithm

Recall that, for this paper, we have assumed $E(\beta_t) = 0$. Therefore, the quantity $V_t = y_t - E(y_t) = \beta_t z_t + \epsilon_t$ has $E(V_t) = 0$. These V_t will serve as our X_t. Let us examine the covariance structure of V_t.

$$Cov(V_t, V_s) = Cov(\beta_t z_t + \epsilon_t, \beta_s z_s + \epsilon_s) \tag{5.3}$$

$$= Cov(\beta_t z_t, \beta_s z_s) + Cov(\beta_t z_t, \epsilon_s) + Cov(\epsilon_t, \beta_s z_s) + Cov(\epsilon_t, \epsilon_s) \tag{5.4}$$

$$= \begin{cases} Cov(\beta_t z_t, \beta_t z_t) + 2Cov(\beta_t z_t, \epsilon_t) + Cov(\epsilon_t, \epsilon_t), & t = s \\ Cov(\beta_t z_t, \beta_s z_s) & t \neq s \end{cases} \tag{5.5}$$

$$= \begin{cases} z_t^2 \gamma_\beta(0) + \sigma_\epsilon^2, & t = s \\ z_t z_s \gamma_\beta(|t - s|), & t \neq s \end{cases} \tag{5.6}$$

since β_t and ϵ_t are assumed uncorrelated. This implies that V_t is not stationary, since the ACVF of V_t depends on t in the way of z_t. However, as discussed above, the Innovations Algorithm does not require stationarity of the process of interest. Rather, we only need that $E(V_t^2) < \infty$. Under the mild assumptions that β_t and ϵ_t have finite variance, which is assumed throughout this paper, we have that condition.

Now we can apply the Innovations Algorithm to V_t. Once we obtain the innovations and their mean square prediction errors, ν_t, the likelihood function of $\vec{V} = (V_1, ..., V_n)^T$ can be expressed as

$$L(\mu, \alpha, \sigma_\epsilon, d, \sigma_\omega) = \frac{1}{\sqrt{(2\pi)^n \prod_{t=1}^{n} \nu_{t-1}}} \exp\left[-\frac{1}{2}\sum_{t=1}^{n} \frac{(V_t - \hat{V}_t)^2}{\nu_{t-1}}\right]. \tag{5.7}$$

In the interest of numerical stability, we use the log-likelihood form of the above.

$$\ell(\mu, \alpha, \sigma_\epsilon, d, \sigma_\omega) = \ln L(\mu, \alpha, \sigma_\epsilon, d, \sigma_\omega) = -\frac{1}{2}\sum_{t=1}^{n} \ln(\nu_{t-1}) - \frac{1}{2}\sum_{t=1}^{n} \frac{(V_t - \hat{V}_t)^2}{\nu_{t-1}}. \tag{5.8}$$

With this function, we run a Newton-Raphson algorithm using the optim function in R to obtain the maximum likelihood estimates of $\mu, \alpha, \sigma_\epsilon, d$ and σ_ω.

5.4 Truncated Innovations Algorithm

Recall from the discussion in Section 4.4 that if we use the differenced process, $\delta_t = \beta_t - \beta_{t-1}$, instead of β_t, we can use a truncated expression to model our process within a reasonable margin of error and with the benefit of increased computation speed. To obtain innovations for the truncated Innovations Algorithm, we need to perform some algebra to get the innovations in terms of δ_t and make sure the expected value of those innovations is zero.

First, we have

$$y_{t-1} = \mu + \alpha a_{t-1} + \beta_{t-1} z_{t-1} + \epsilon_{t-1} \tag{5.9}$$

from (1.5). Then, define $r_t = z_t / z_{t-1}, t = 2, ..., n$ and multiply both sides of (5.9) by r_t. Doing so, we get

$$r_t y_{t-1} = r_t \mu + \alpha r_t a_{t-1} + \beta_{t-1} z_t + r_t \epsilon_{t-1}. \tag{5.10}$$

Now we denote $U_t = y_t - r_t y_{t-1}$ and obtain

$$\begin{aligned} U_t &= (\mu + \alpha a_{t-1} + \beta_{t-1} z_{t-1} + \epsilon_{t-1}) - (r_t \mu + \alpha r_t a_{t-1} + \beta_{t-1} z_t + r_t \epsilon_{t-1}) \\ &= (1 - r_t)\mu + (a_t - r_t a_{t-1})\alpha + \delta_t z_t + (\epsilon_t - r_t \epsilon_{t-1}). \end{aligned} \tag{5.11}$$

Taking the expected value of U_t yields $E(U_t) = (1 - r_t)\mu + (a_t - r_t a_{t-1})\alpha$. Finally, we define X_t for our truncated model as $V_t = U_t - E(U_t) = \delta_t z_t + \epsilon_t - r_t \epsilon_t,\ t = 2, ..., n$. Therefore, $E(V_t) = 0$. Let us examine the covariance structure of V_t; WLOG, $s > t$.

$$\begin{aligned} Cov(V_t, V_s) &= Cov(\delta_t z_t, \delta_s z_s) + Cov(\varepsilon_t - r_t \varepsilon_{t-1}, \varepsilon_s - r_s \varepsilon_{s-1}) \\ &= \begin{cases} z_t^2 Var(\delta_t) + Var(\varepsilon_t) + r_t^2 Var(\varepsilon_{t-1}), & \text{if } s - t = 0; \\ z_t z_s Cov(\delta_t, \delta_{t+1}) + Cov(\varepsilon_t - r_t \varepsilon_{t-1}, \varepsilon_{t+1} - r_{t+1} \varepsilon_t), & \text{if } s - t = 1; \\ z_t z_s Cov(\delta_t, \delta_s), & \text{if } s - t = 2, \ldots, n-2, \end{cases} \\ &\approx \begin{cases} z_t^2 \gamma_\delta(0) + (1 + r_t^2)\sigma_\varepsilon^2, & \text{if } s - t = 0; \\ z_t z_s \gamma_\delta(1) - r_s \sigma_\varepsilon^2, & \text{if } s - t = 1; \\ z_t z_s \gamma_\delta(s - t), & \text{if } s - t = 2, \ldots, m; \\ 0, & \text{if } s - t = m + 1, \ldots, n - 2, \end{cases} \end{aligned}$$

because of the truncation of the model at m, and where $\gamma_\delta(h)$ is the ACVF of δ_t.

Clearly, V_t is not stationary, but that is not a problem for the Innovations Algorithm. Again, the requirement that $E(V_t^2) < \infty$ is satisfied as long as the variance of β_t and ϵ_t are finite. Now, WLOG $V_{t-1}^* = V_t, t = 2, .., n$ for indexing purposes, and we apply the Innovations Algorithm, Definition 5.2.1, to V_t^*. Then, the log-likelihood function for $(V_1^*, ..., V_{n-1}^*)^T$ is

$$\ell(\mu, \alpha, \sigma_\epsilon, d, \sigma_\omega) = \ln L(\mu, \alpha, \sigma_\epsilon, d, \sigma_\omega) = -\frac{1}{2}\sum_{t=1}^{n-1} \ln(\nu_{t-1}) - \frac{1}{2}\sum_{t=1}^{n-1} \frac{(V_t^* - \hat{V}_t^*)^2}{\nu_{t-1}}. \quad (5.12)$$

CHAPTER 6

SIMULATION RESULTS

We simulated covariates from five different models to test our estimation methods described above. The models we used simulate common characteristics of data and are described below. When estimating the parameters of our SPR model, we standardized the z_t in the following way: $z_t = z_{t,\text{unstandard}}/sd(z_{t,\text{unstandard}})$. This helps us to obtain more stable convergence. The covariates used in our simulation findings are summarized in Table 6.1.

	Process	Expression	Parameter Choices
Model 1	AR(1)	$z_t = \phi z_{t-1} + \xi_t$	$\phi = .8$, $\sigma_\xi^2 = 1$
Model 2	Random Walk with Drift	$z_t = z_{t-1} + \xi_t$	$\xi_t \sim$ iid N$(0, \sigma_\xi^2)$, $\sigma_\xi^2 = 1$
Model 3	Linear Trend with AR(1) Errors	$z_t = \alpha_0 + \alpha_1 t + \xi_t$	$\xi_t = \phi\xi_{t-1} + \eta_t$, $\eta_t \sim$ iid N$(0, \sigma_\eta^2)$, $\alpha_0 = 0, \alpha_1 = .05$, $\phi = .8, \sigma_\eta^2 = 1$
Model 4	Periodic with Random Errors	$z_t = \alpha_0 + \alpha_1 \cos(2\pi/T) + \alpha_2 \sin(2\pi/T) + \xi_t$	$\alpha_0 = 4, \alpha_1 = 1$ $\alpha_2 = 1, T = 12$, $\sigma_\xi^2 = 1$
Model 5	Periodic with Linear Trend and Random Errors	$z_t = \alpha_0 + \alpha_1 \cos(2\pi/T) + \alpha_2 \sin(2\pi/T) + \alpha_3 t + \xi_t$	$\xi_t \sim$ iid N$(0, \sigma_\xi^2)$, $\alpha_0 = 4, \alpha_1 = 1$, $\alpha_2 = 2, \alpha_3 = .5$, $T = 12, \sigma_\xi^2 = 1$

Table 6.1: Choices for Covariate Models

Notice that a period of $T = 12$ suggests monthly data.

We ran 1000 simulations from each model, using FIA,TIA, and TKF. Because FIA and FKF produce the same results both theoretically and in practice, we will not include the FKF results. We obtained OLS estimates by fitting a classical constant parameter regression model and compare those estimates to ours. The OLS approach does not have a natural estimate for σ_ϵ and σ_ω, so OLS estimates for those results are not included. Finally, for presentability, we did not include estimates for μ, as μ is not generally a parameter of interest in many practices. In mean and variance, the estimates for μ behave much the same as for α. The tables below show the results for $\alpha = .05, d = .4, \sigma_\epsilon = 1.5$ and $\sigma_\omega = 1$. The figure following contains the boxplots for the parameters in Model 1. For the remaining boxplots, please see Appendix A.

	α_{OLS}	α_{FIA}	α_{TIA}	α_{TKF}	d_{OLS}	d_{FIA}	d_{TIA}	d_{TKF}
Model 1	.0496	.0499	.0499	.0498	.3056	.3801	.3976	.3842
Model 2	.0497	.0498	.0499	.0498	.2696	.3788	.3877	.3757
Model 3	.0497	.0498	.0497	.0498	.2312	.3671	.3823	.3662
Model 4	.0496	.0499	.0499	.0499	.2906	.3810	.3980	.3836
Model 5	.0498	.0500	.0500	.0500	.2641	.3801	.3925	.3793

Table 6.2: Simulation Results: α and d

	$\sigma_{\epsilon,FIA}$	$\sigma_{\epsilon,TIA}$	$\sigma_{\epsilon,TKF}$	$\sigma_{\omega,FIA}$	$\sigma_{\omega,TIA}$	$\sigma_{\omega,TKF}$
Model 1	1.4043	1.4015	1.3846	.9999	1.0016	1.0068
Model 2	1.4409	1.4385	1.4324	1.0118	1.0145	1.0197
Model 3	1.4811	1.4793	1.4784	1.0032	1.0034	1.0094
Model 4	1.4306	1.4330	1.4117	1.0003	1.0010	1.0089
Model 5	1.4660	1.4630	1.4563	1.0014	1.0040	1.0102

Table 6.3: Simulation Results: σ_ϵ and σ_ω

The simulation results show that each of the estimation methods proposed in this paper provide very good point estimates for each of the unknown parameters. The estimate for d in particular is very promising, given that the OLS estimate is quite far from the true value, 0.4. As seen in the boxplots following in Figure 6.1, the standard

error for the OLS estimate of α is quite high compared to FIA, TIA, and TKF. Thus, our proposed estimates seem to outperform the OLS estimates. The standard errors of our estimates are quite reasonable, except perhaps for the estimates of σ_ϵ. This issue is addressed in Chapter 7.

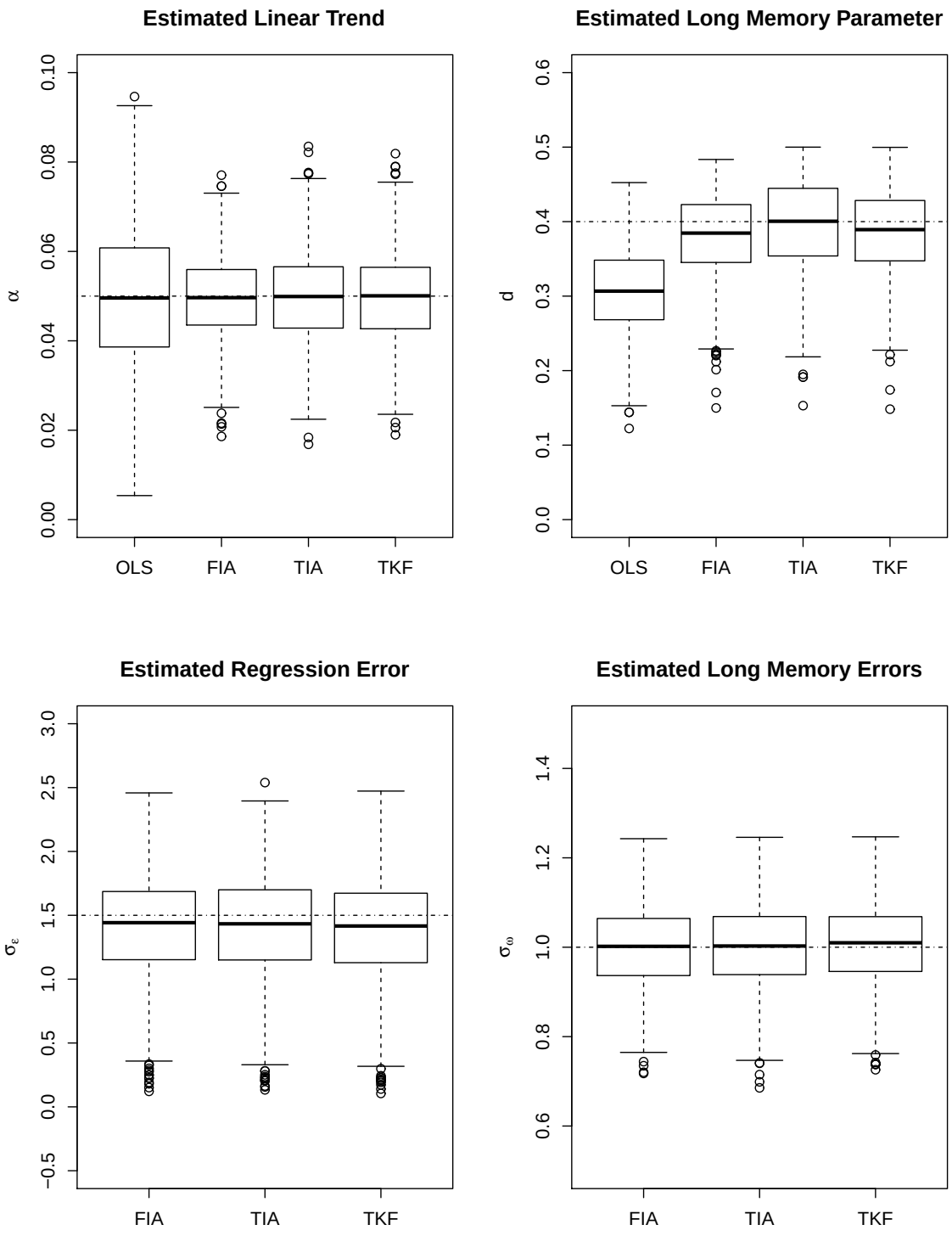

Figure 6.1: Model 1 Boxplots

CHAPTER 7

CONCLUSION

In this paper, we established a model for taking into account changes in the relationship between predictor and response variables when the response shows long memory behavior. This model takes heteroscedasticity into account because the variance of the response variable is related to the value of the predictor variable, adding to the time series analysis toolbox. Using this approach also gives more explanatory power to the predictor variable by allowing for the relationship between the predictor and response to have its own meaning as a stochastic process. We also developed estimation methods using the Kalman Filter and Innovations Algorithm, including a truncated version of each to decrease computational burden, and then extracted the maximum likelihood estimates using a Newton-Raphson procedure. The specialized model in this paper is most applicable to data sets where the response shows long memory and the coefficients describing the response's linear relationship to the predictor are zero on average, since we focused on the case where $E(\beta_t) = 0$ for computational reasons. If our approach is generalized, it could also be applied when the overall relationship between the predictor and response is nonzero.

Our simulation results provide numerical evidence for the usefulness of our model and estimation methods. We can see from the simulations that the OLS estimate performs well on average for estimates of α and μ (not pictured), but it is significantly

low for the d estimate. The OLS estimate also has considerably larger variance in the case of estimating α. The estimation methods we developed appear to be quite accurate for the simulated covariate models described. We also notice that the parameter estimates do not appear to be greatly affected by the chosen estimation method (aside from the OLS estimate), which adds credence to the truncated methods, as they cause much less computational burden. One topic of further investigation might be the estimates for σ_ϵ, which have a higher variance than the other parameter estimates for certain models (see Appendix A). This could be due to the relationship of σ_ϵ, d, and σ_ω; if σ_ϵ is over-estimated, it takes away explanatory power from the β_t series, and thus d and σ_ω. It could be possible that, if σ_ϵ is large, it overwhelms the estimation of the long memory parameters, making our estimates less accurate. Thus, different parameter values for σ_ϵ, d, and σ_ω could make our estimates slightly more or less variant.

Extensions of this paper have been alluded to throughout. The simplest modification would be allowing for $E(\beta_t) = \beta$ to be nonzero. The theoretical changes to the model are uncomplicated, but when we attempted to estimate a nonzero β, we ran into computation issues. The second, and much more complicated modification to our model would be to allow β_t to take on an arbitrary ARFIMA(p, d, q) process. Again, this should not pose any major theoretical issues, but the ACF and infinite MA coefficients for the ARFIMA(p, d, q) model are far more difficult to work with. At the very least, this could introduce more room for computational error and increase computational demands. An alternative would be to take short memory influences into account by adding a different predictor variable and supposing that the long memory and short memory components come from different sources. In any case, the model discussed in this paper should act as a useful base for long memory time series

regression analysis.

REFERENCES

[1] Bloomfield, P. and Nychka, D. (1992), "Climate Spectra and Detecting Climate Change," *Climate Change*, 21, 275-287.

[2] Brockwell, P. J. and Davis, R. A. (2002), *Introduction to Time Series and Forecasting*, New York: Springer-Verlag.

[3] Brown, J. P., Song, H. and McGillivray, A. (1997), "Forecasting UK House Prices: a Time-varying Coefficient Approach," *Economic Modeling*, 14, 529-548.

[4] Burnett, T. D. and Guthrie, D. (1970), "Estimation of Stationary Stochastic Regression Parameters," *Journal of the American Statistical Association*, 65(332), 1547-1553.

[5] Chan, N. H. and Palma, W. (1998), "State Space Modeling of Long-Memory Processes," *The Annals of Statistics*, 26(2), 719-740.

[6] CoinNews Media Group LLC (updated January 16, 2014), "US Inflation Calculator," http://www.usinflationcalculator.com/inflation/historical-inflation-rates/.

[7] Faraway, J. (2005), *Linear Models in R*, New York:Chapman & Hall/CRC.

[8] Federal Reserve Bank of St. Louis (updated February 3, 2014), "3-Month Treasury Bill: Secondary Market Rate (TB3MS), Percent, Monthly, Not Seasonally Adjusted," http://research.stlouisfed.org/fred2/series/TB3MS.

[9] Hassler, U. and Wolters, J. (1995), "Long Memory in Inflation Rates: International Evidence," *Journal of Business & Economic Statistics*, 13, 37-45.

[10] Hosking, J. R. M. , (1981), "Fractional Differencing," *Biometrika*, 68, 165-176.

[11] Hurst, H. E. (1951), "Long-term Storage Capacity of Reservoirs," *Transactions of the American Society of Civil Engineers*, 73, 261-284.

[12] Koopman, S. J., Ooms, M. and Carnero, M. A. (2007), "Periodic Seasonal Reg-ARFIMA-GARCH Models for Daily Electricity Spot Prices," *Journal of the American Statistical Association*, 102, 16-27.

[13] Lee, J. and Ko, K. (2007), "One-way Analysis of Variance with Long Memory Errors and its Application to Stock Returns Data," *Applied Stochastic Models in Business and Industry*, 23, 493-502.

[14] Lee, J. and Ko, K. (2009), "First-order Bias Correction for Fractionally Integrated Time Series," *The Canadian Journal of Statistics*, 37(3), 476-493.

[15] McElroy, T. S. and Holan, S. H. (2012), "On The Computation of Autocovariances for Generalized Gegenbauer Processes," *Statistica Sinica*, 22, 1661-1687.

[16] McLeod, A.I. and Hipel, K. W. (1978), "Preservation of the Rescaled Adjusted Range, I. A Reassessment of the Hurst Phenomenon," *Water Resources Research*, 101, 133-443.

[17] Newbold, P. and Bos, T. (1985), *Stochastic Parameter Regression Models*, Beverly Hillls: Sage.

[18] Palma, W. (2007), *Long-Memory Time Series Theory and Methods*, New Jersey: John Wiley & Sons.

[19] Robinson, P. M. and Hidalgo, F. J. (1997), "Time Series Regression with Long-range Dependence," *The Annals of Statistics* 25(1), 77-104.

[20] Rosenberg, B. (1973), "A Survey of Stochastic Parameter Regression," *Annals of Economic and Social Measurement*, 2(4), 380-396.

[21] Shumway, R.H. and Stoffer, D.S. (2006), *Time Series Analysis and Its Applications with R Examples* (2nd ed.), New York: Springer.

[22] Sowell, F. (1992), "Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models," *Journal of Econometrics*, 29, 477-488.

[23] Wackerly, D.D., Mendenhall III, W., and Scheaffer, R. L. (2008), *Mathematical Statistics with Applications*, Canada: Brooks/Cole.

APPENDIX A

OTHER SIMULATION BOXPLOTS

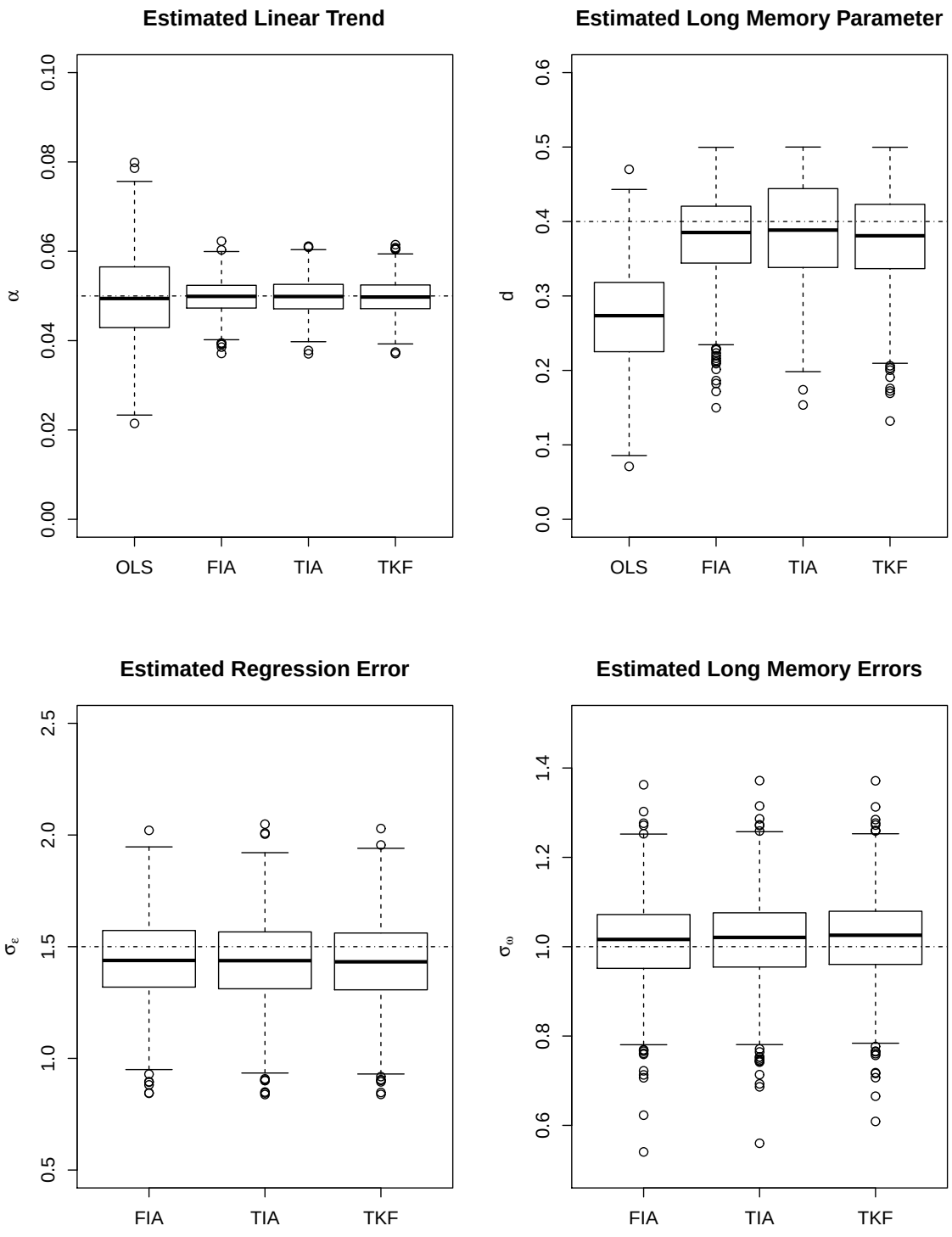

Figure A.1: Model 2 Boxplots

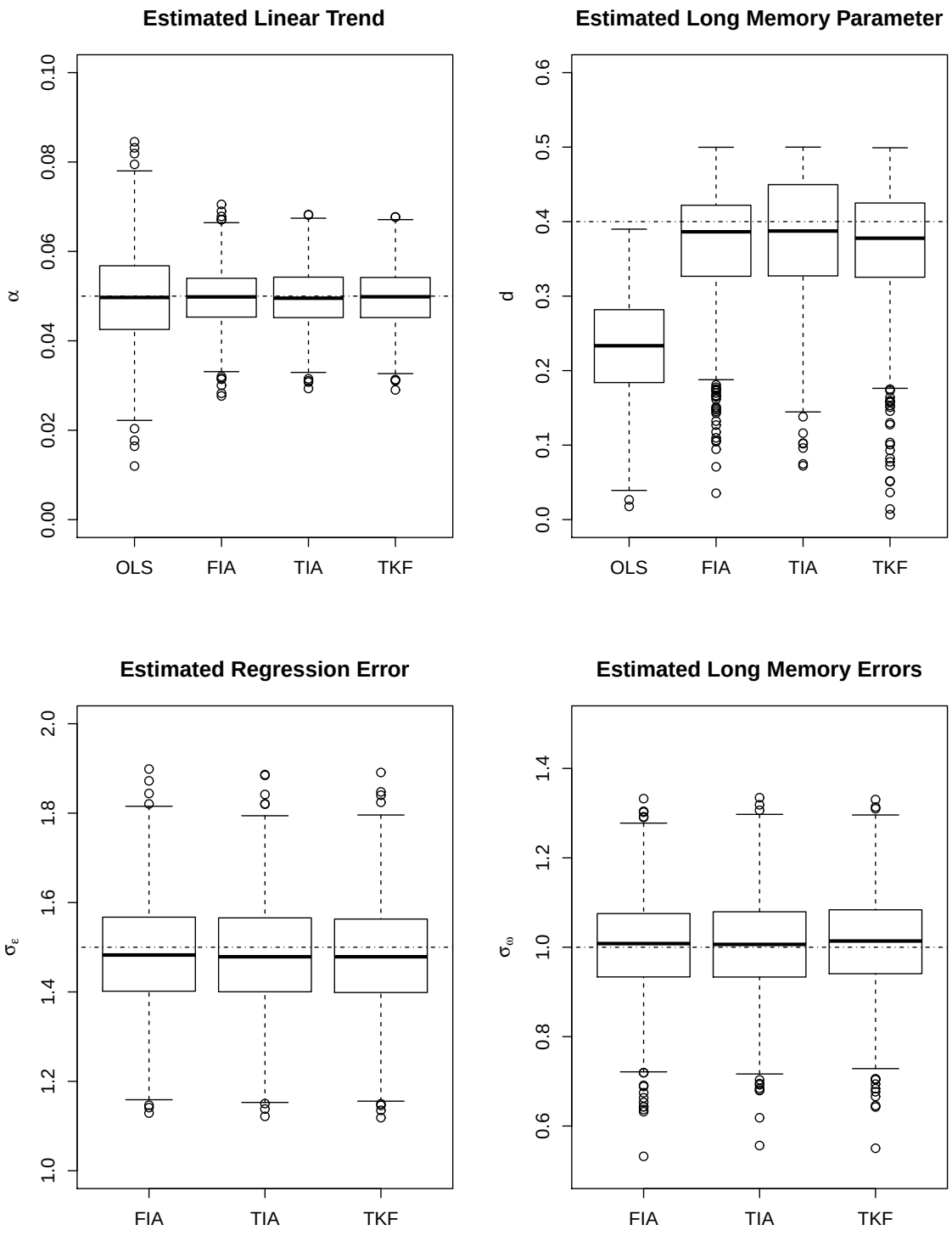

Figure A.2: Model 3 Boxplots

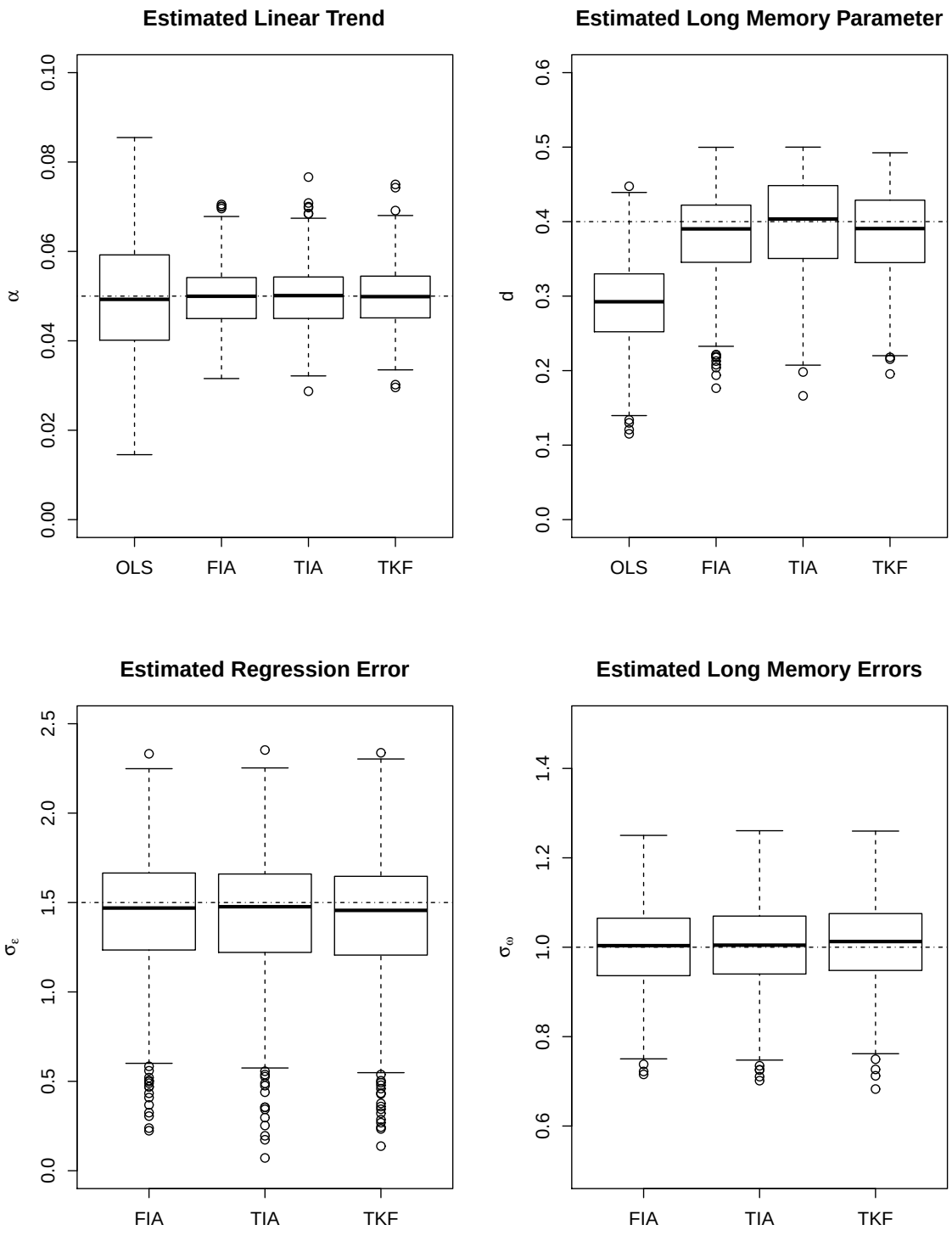

Figure A.3: Model 4 Boxplots

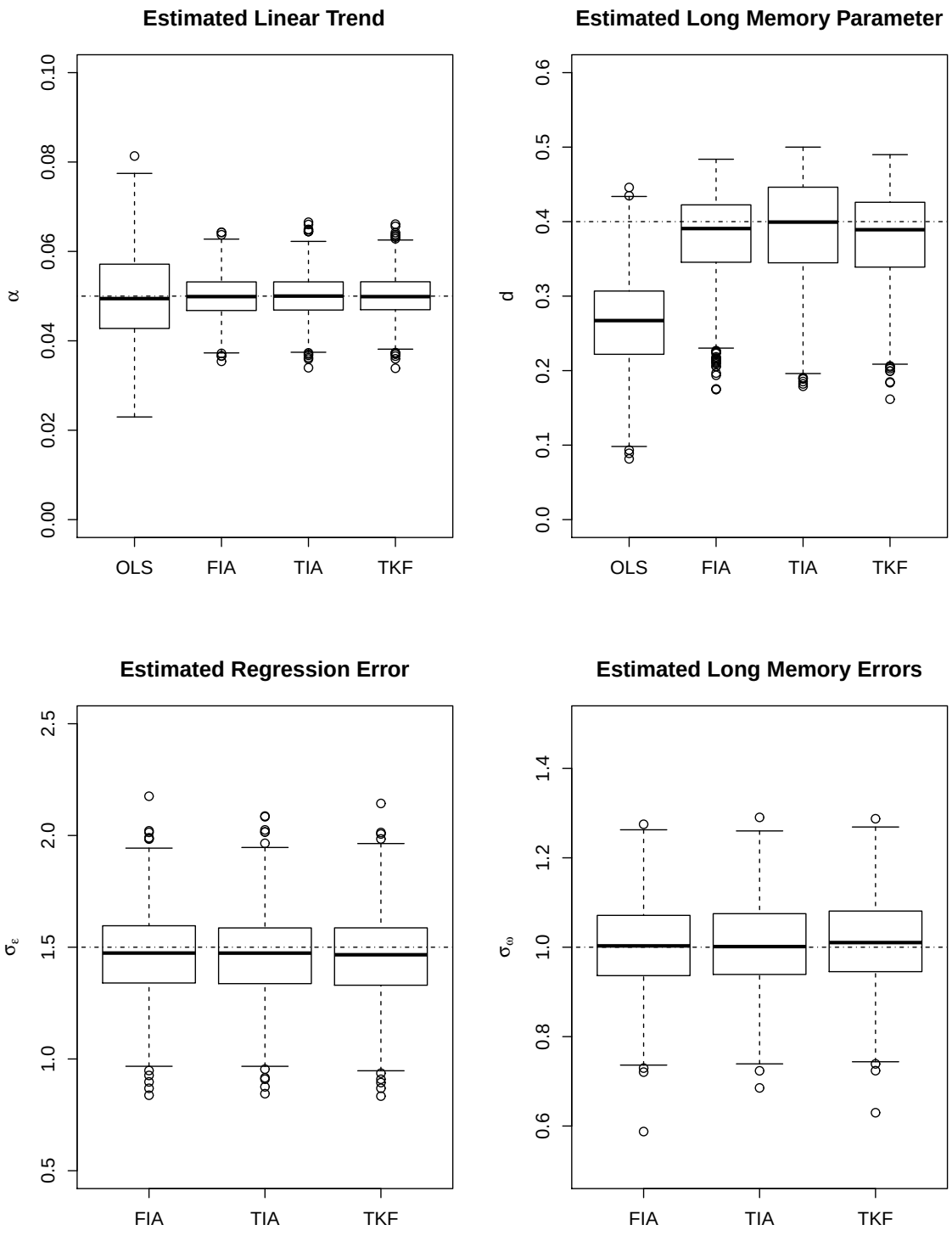

Figure A.4: Model 5 Boxplots