ABSTRACT

Semiparametric Estimation and Forecasting
for Functional-Coefficient Autoregressive Models

Joshua D. Patrick, Ph.D.

Chairperson: Jane L. Harvill, Ph.D.

The functional-coefficient autoregressive (FCAR) model is a useful structure for reducing the size of the class of nonlinear time series models. Local linear regression has been shown to be an effective method for estimating the coefficient functions of these models. However, local linear regression suffers from the "curse of dimensionality" for high order models. We adapt a spline-backfitted kernel (SBK) method for estimating the coefficient functions. The SBK estimators are computationally expedient and theoretically reliable. We show the SBK estimator performs better than local linear regression in root mean square error through simulation results.

Three forecasting methods for FCAR models have been examined in the literature after fitting the model with local linear regression. We adapt the three methods to the SBK estimators. We also examine methods for constructing prediction intervals for the forecasts. The performance of the three forecasting methods and the prediction intervals are compared through simulation results.

Utility scale photovoltaic (PV) plants are becoming economically viable. Utility companies are interested in forecasting the short-term and long-term power generated from a PV plant. The power generated by the plant is correlated with the irradiance measured from the sun. We develop a spatio-temporal model for irradiance measurements from a 1.2 MW PV plant located in Lanai, Hawaii. We use the SBK method for estimating the time component of the model. We assume a separable covariance structure and find evidence that this assumption does not hold.

Semiparametric Estimation and Forecasting
for Functional-Coefficient Autoregressive Models

by

Joshua D. Patrick, B.S., M.S.

A Dissertation

Approved by the Department of Statistical Science

_____

Jack D. Tubbs, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of
Doctor of Philosophy

Approved by the Dissertation Committee

_____

Jane L. Harvill, Ph.D., Chairperson

_____

Eva I. Doyle, Ph.D.

_____

John W. Seaman Jr., Ph.D.

_____

Jack D. Tubbs, Ph.D.

_____

Dean M. Young, Ph.D.

Accepted by the Graduate School
May 2013

_____

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I would like to thank God for all the blessings in my life. My time in graduate school has not only been an academic journey, but also a spiritual journey. I pray that He guides me throughout the rest of my life.

I would like to thank Dr. Raid Amin and Dr. Morris Marx at the University of West Florida for encouraging me to continue my education and to get a Ph.D. Both of you showed me the joy of Statistics and you both were there to help me at any moment.

I would also like to thank the faculty in the Department of Statistical Science at Baylor University for your instruction throughout my graduate studies. I especially would like to thank my academic advisor, Dr. Harvill. You gave me the freedom to explore topics of my choosing while always being there to guide me down whatever path I was going. I look forward to working with you throughout my career. A special thanks to Dr. Young. You were not my academic advisor, but you were always checking up with me to see how I was doing. Your encouragement meant a lot to me. To Dr. Bratcher and Dr. Tubbs, thank you for taking a chance on me and bringing me to Baylor.

I am also appreciative of Dr. Cliff Hansen at Sandia National Laboratories for giving me a fascinating data set to work with. Thank you for allowing me to use this in my dissertation.

Finally, I would like to thank my family for all of the support. To my sister, brother-in-law, brothers, aunts, uncles, and cousins, thank you all for your support and encouragement. I would especially like to thank my parents, Rick and Ginnie Patrick, for all of the encouragement, support, and love.

DEDICATION

To my parents, Rick and Ginnie Patrick

CHAPTER ONE

Overview

1.1 Introduction

Nonlinear models are often more appropriate than linear models for fitting time series data. Some non-linear dynamic systems include limit cycles, jump phenomenon, time irreversibility, and amplitude-frequency models (Tong, 1990). The class of nonlinear time series models is extremely large and include models such as threshold autoregressive (TAR), self-exciting threshold autoregressive (SETAR), exponential autoregressive (EXPAR), bilinear, smooth threshold autoregressive (STAR), and generalized autoregressive conditional heteroscedasticity (GARCH) models. A useful structure that can reduce the size of the class of nonlinear models for modeling a data set is the functional-coefficient autoregressive (FCAR) model. The FCAR model is defined as

$$X_t = m_1\left(X_{t-d}\right) X_{t-1} + \cdots + m_p\left(X_{t-d}\right) X_{t-p} + \varepsilon_t, \tag{1.1}$$

where p and d are positive integers, $d \leq p$, $m_j\left(X_{t-d}\right)$ is a measurable function of the delay variable X_{t-d}, for $j = 1, \ldots, p$, and $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with mean 0 and variance σ^2. We will estimate these models by adapting a procedure known as spline-backfitted kernel (SBK) estimation and examine the forecasting performance using these estimators. In this chapter, we discuss the literature on FCAR models in Section 1.2 and the literature on forecasting methods for FCAR models in Section 1.3. We then discuss the SBK method in Section 1.4 and give an overview of Chapters Two, Three, and Four in Section 1.5.

1.2 Functional Coefficient Autoregressive Models

Chen and Tsay (1993) introduced the FCAR model and proposed a procedure for building the model based on arranged local regression (ALR). The ALR procedure works by first selecting an interval length c to form a window and a minimum sample

size K to control the number of observations in the window. Let $x_{(1)}, \ldots, x_{(k)}$ be in the window $\left[x_{(1)}, x_{(1)} + c\right]$. Fit the linear regression

$$x_{t-d} = a_1 x_{t+d-1} + \cdots + a_p x_{t+d-p} + \epsilon_{t+d}$$

with $t = t_1, \ldots, t_k$ where ϵ_{t+d} is the residual at time $t+d$. The ordinary least squares (OLS) estimate of a_j is an estimate of $m_j\left(x_{(1)} + c\right)$. Move the window along the X_{t-d} axis until there is at least one new data point and at least K data points in the window. Refit the linear regression to obtain the estimate $\hat{m}_j\left(x_{(i)} + c\right)$ of $m_j\left(x_{(i)} + c\right)$ where $x_{(i)} + c$ is the right end point of the window. The scatterplots of the estimates of $\hat{m}_j(x)$ versus x can be used to infer the functional form of $m_j(\cdot)$. Chen and Tsay (1993) used simulated data and real data examples to compare the FCAR model and the proposed model building procedure to threshold and linear time series models through multistep forecasts. The FCAR model performed much better than the other two models in terms of bias. However, the FCAR model only performed better for short term forecasts in terms of mean square error (MSE). For long term forecasts, the linear model performed the best in MSE.

1.2.1 Local Linear Fitting of FCAR Models

The method of Chen and Tsay (1993) constructed estimators based on an iterative recursive formula that resembles local constant fitting. Cai, Fan, and Yao (2000) used a local linear fitting method to estimate $m_j(\cdot)$ in (1.1). In the local linear method, each $m_j(\cdot)$ is approximated locally at u_0 by a linear function $m_j(u_t) \approx a_j + b_j(u_t - u_0)$. In the time series context, U_t is the lagged variable X_{t-d}. The local linear estimate is $\hat{m}_j(u_0) = \hat{a}_j$, where $\left\{\left(\hat{a}_j, \hat{b}_j\right)\right\}$ minimize

$$\sum_{t=1}^{n}\left[X_t - \sum_{j=1}^{p}\left\{a_j + b_j(U_t - u_0)\right\} X_{t-j}\right]^2 K_h(U_t - u_0), \tag{1.2}$$

where $K_h(\cdot) = h^{-1}K(\cdot/h)$, $K(\cdot)$ is a kernel function, and $h > 0$ is a bandwidth. The least squares solution is

$$\hat{m}_j(u_0) = \sum_{t=1}^{n} K_{n,j}(U_t - u_0, \mathbf{X}_t)\, Y_t,$$

where

$$K_{n,j}(u, \mathbf{x}) = \mathbf{e}_{j,2p}^T \left(\tilde{\mathbf{X}}^T \mathbf{W} \tilde{\mathbf{X}} \right)^{-1} \tilde{\mathbf{X}}^T K_h(u). \tag{1.3}$$

In expression (1.3), $\mathbf{e}_{j,2p}$ is the $2p \times 1$ unit vector with 1 at the jth position, $\tilde{\mathbf{X}}$ denotes an $n \times 2p$ matrix with $\left(\mathbf{X}_t^T, \mathbf{X}_t^T (U_t - u_0)\right)$ as its tth row, and $\mathbf{W} = \text{diag}\{K_h(U_1 - u_0), \ldots, K_h(U_n - u_0)\}$.

Cai et al. (2000) used local linear fitting on simulated data from an EXPAR model. They assessed the fit by calculating the square root of average squared errors (RASE),

$$\text{RASE}_j = \left[n_{\text{grid}}^{-1} \sum_{k=1}^{n_{\text{grid}}} \{\hat{m}_j(u_k) - m_j(u_k)\}^2 \right]^{1/2},$$

where $\{u_k, k = 1, \ldots, n_{\text{grid}}\}$ are regular grid points. The performance of the method was gauged by comparing the RASE to the standard deviation of the time series denoted by σ_X. The mean of the estimate $\hat{\sigma}_X$ of σ_X for their 400 replications was 0.5389 with a standard deviation of 0.048. The RASE was well below that value with a median RASE of around 0.10.

Real data examples were used to assess the post sample forecasting performance of the local linear method. The two examples were the Canadian lynx data set (Tong, 1990, p. 377) and the Wolf's annual sunspot numbers data set (Tong, 1990, p. 420). The local linear method was compared with the linear AR model, the TAR model, and the ALR procedure of Chen and Tsay (1993) using a one-step ahead and a iterative two-step ahead forecast. In terms of average absolute predictive errors (AAPE), the local linear method had much better performance than both the linear AR model and the TAR model in the Canadian Lynx example and performed just as well as the other two models in the sunspot numbers example.

Chen and Liu (2001) establish the asymptotic properties of the local linear estimator. They summarize these properties in a theorem which states that for $\boldsymbol{\beta} = \left(\hat{a}, \hat{b}\right)$ that minimizes (1.2), $\mu_2 = \int u^2 K(u)\, du$, $K_2^2 = \int K^2(u)\, du$, $p_{i,j,d}$ is the joint stationary density of the triplet $(X_{t-i}, X_{t-j}, X_{t-d})$, and $p(x)$ is the marginal

density of X_t, then

$$n^{2/5}\left(\hat{\mathbf{m}}\left(x\right)-\mathbf{m}\left(x\right)-\boldsymbol{\beta}^2\mathbf{b}\left(x\right)\right)\xrightarrow{D}\mathrm{N}_p\left(0,\beta^{-1}\sigma^2K_2^2\mathbf{A}^{-1}\left(x\right)\right),$$

where $\mathbf{A}\left(x\right)=p\left(x\right)E\left[\mathbf{X}_t\mathbf{X}_t^T|X_{t-d}=x\right]$ and $\mathbf{b}\left(x\right)=\mu_2\mathbf{A}^{-1}\left(x\right)\mathbf{B}\left(x\right)$ where $\mathbf{B}\left(x\right)$ is a vector with ith element being

$$\sum_{j=1}^{p}\int uv\left\{\frac{1}{2}f_j''\left(x\right)p_{i,j,d}\left(u,v,x\right)+m_j'\left(x\right)p_{i,j,d}'\left(u,v,x\right)\right\}dudv$$

with $m_j'\left(x\right)$ and $m_j''\left(x\right)$ being the first and second derivative of $m_j\left(x\right)$ respectively, and $p_{i,j,d}'$ being the partial derivative with respect to the third argument (see Theorem 1 in Chen and Liu, 2001, p. 154). This result is only applicable when x is a continuity point. Chen and Liu (2001) use the theorem to construct pointwise confidence bands for the function estimates of an EXPAR model. They show that the local linear procedure is reasonably robust for this type of model.

Now consider the TAR model

$$x_t=\phi_1^{(i)}x_{t-1}+\cdots+\phi_p^{(i)}x_{t-p}+\varepsilon_t^{(i)}\quad\text{if } x_{t-d}\in\Omega_i,\quad i=1,\ldots,k,$$

where $\{\Omega_i\}$ form a partition of $\Re$. Clearly, this model is a special case of the FCAR model where $m_j\left(x_{t-d}\right)=\phi_j^{(i)}$. However, $m_j\left(x_{t-d}\right)$ is not continuous at all points in this model. Chen and Liu (2001) note that estimating the functional coefficients using the local linear procedure result in large bias in the neighborhood of points of discontinuity. The authors propose a procedure similar to the local linear estimator but using one-sided kernels. They also establish the asymptotic properties using the one-sided kernels. However, the convergence rate of the one-sided estimator is slower than that of the two-sided estimator. The simulations of Chen and Tsay (1993) are repeated using the two-sided local linear estimator and they show the same results even with the TAR model.

1.2.2 The Curse of Dimensionality

A general review of nonparametric methods for time series models can be found in Härdle, Lütkepohl, and Chen (1997). They note that using nonparametric methods has the advantage of letting the data speak for itself but at the cost of poor performance in high dimensions. This disadvantage is known as the "curse of

dimensionality." The authors suggest the nonparametric approach should be used as a guidance for choosing appropriate parametric models with lower dimensions and for deciding between competing classes of models. They state that one way to ease the curse of dimensionality is to impose some restrictions on the model. For example, we could impose the nonlinear additive autoregressive (NAAR) model defined as

$$X_t = c + f_1(X_{t-1}) + f_2(X_{t-2}) + \cdots + f_p(X_{t-p}) + \varepsilon_t, \tag{1.4}$$

where c is a constant. The Nadaraya-Watson estimator can be used to estimate $f_j(X_{t-j})$. This estimator has a rate of convergence of $n^{2/5}$ which is the same as one-dimensional smoothing. Hence, the curse of dimensionality is not as severe for this estimator.

Chen and Liu (2001) also note that the local linear estimator for FCAR models has a rate of convergence of $n^{2/5}$. Thus, this estimator does not suffer from the curse of dimensionality as much as direct p-dimensional estimation. Cai et al. (2000) give more detail as to why this estimator has the same convergence as one-dimensional smoothing. They note that the local linear method smooths the delay variable X_{t-d} only. So, unless the delay variable is a vector, the estimator does not suffer the difficulties of higher dimensional models.

1.2.3 Estimation Using Polynomial Splines

Most of the literature on nonlinear times series estimation involves using local polynomial smoothing with kernel weights. For FCAR models, the local polynomial procedure was the only method investigated until the work of Huang and Shen (2004). In that paper, the authors propose a global smoothing procedure based on polynomial splines for estimating FCAR models. They also prove consistency and find the rate of convergence for their method. The authors note that the spline method yields a fitted model with a parsimonious explicit expression which is an advantage over the local polynomial method. This feature allows us to produce multi-step ahead forecasts conveniently. Additionally, their spline method is less computationally intensive than the local polynomial method.

We now define the spline method with degree 3 and knot sequence $\kappa_0, \ldots, \kappa_{N+1}$ where κ_0 and κ_{N+1} are the exterior knots. Huang and Shen (2004) use the truncated power basis which are $1, x, x^2, x^3, (x-\kappa_1)_+^3, \ldots, (x-\kappa_N)_+^3$. Suppose the coefficient function $f_j(X_{t-d})$ in (1.1) is smooth. Then there is a set of basis functions $B_{js}(\cdot)$ and constants β_{js}^*, $s = 1, \ldots, K_i$, such that

$$m_j(u) \approx m_j^*(u) = \sum_{s=1}^{K_i} \beta_{js}^* B_j(u). \tag{1.5}$$

The β_{js}^*s are estimated by minimizing

$$\sum_{t=1}^{n} \left(X_t - \sum_{j=1}^{p} \left\{ \sum_{s=1}^{K_i} \beta_{js} B(U_t) \right\} X_{t-i} \right)^2$$

with respect to β. The estimate is denoted as $\hat{\beta}_{js}$. The least squares estimate of the coefficient function $m_j(u_t)$ is

$$\hat{m}_j(u_t) = \sum_{s=1}^{K_j} \hat{\beta}_{js} B_{js}(u_t).$$

The number of terms in (1.5) is determined by the number of knots N and the power of the splines. Huang and Shen (2004) applied the spline method to simulated examples and compared their results to the results of Cai et al. (2000). They found that the local linear estimates tended to have larger bias than the estimates of the spline method. They also found that the spline fits are usually smoother than the local linear fits. One disadvantage of using the spline method is the lack of asymptotic theory. Recent literature has provided substantial development of asymptotic theory for i.i.d. data but not for time series data (see e.g. Stone (1994) and Huang (2001)).

1.3 Forecasting Methods for FCAR Models

Early literature for forecasting nonlinear models focused primarily on using Monte Carlo (MC) or bootstrap methods with parametric models. For example, Clements and Smith (1997) discuss methods for forecasting with SETAR models. In their paper, the authors compare different forecasting methods against the forecast from a linear AR model. They assume all the parameters are unknown except for the lag orders and the delay lag. Through simulation results, the authors show that

the MC method is at least as good as the other methods they compared, but it is generally not preferable because it is computationally expensive. The authors conclude that the bootstrap method is preferred over the MC method, but the differences in the forecast accuracy are relatively small. For comparison to the linear AR forecasts, the authors found the linear AR model provides a good approximation to the SETAR model. However, this approximation is dependent on the sign and magnitude of the intercepts of the model.

Before Harvill and Ray (2005), little work had been done in comparing different forecasting methods for FCAR models. Chen and Tsay (1993) used the ALR procedure to estimate the FCAR model (see Section 1.2). They also obtained multi-step forecasts but they did specify how this is done. Chen (1996) proposed a multistage kernel smoother for a general non-linear AR model and derived the asymptotic mean squared error and integrated mean squared error. Through simulation results, the author was able to show the multistage smoother has smaller mean squared error than the direct kernel smoother. Fan and Yao (2003) showed that a direct and iterative "naive" method performs well in average absolute prediction error when compared to forecasting using a linear AR model. They use the local smoothing method of Cai et al. (2000) to estimate the coefficient functions in both methods. The direct method forecasts Y_{t+M} as a function of Y_t while ignoring the relationship between Y_{t+M} and Y_{t+m-j}, $j = 1, \ldots, M-1$. The iterative method forecasts Y_{t+M} by substituting the forecasted values $\hat{Y}_{t+M-j}$, $j = 1, \ldots, M-1$ into the FCAR mean function. The authors' simulation results were only for a TAR model of order two and show that the iterative method performs better than the direct method. Huang and Shen (2004) proposed a simulation based method using the fitted model recursively (see Section 1.2.3). They use bootstrapping to sample the residuals of the fitted models to generate an error term for the simulated series.

Harvill and Ray (2005) adapt the bootstrapping method of Huang and Shen (2004) to forecasting FCAR models fitted using local linear smoothing. They also adapt the multistage method of Chen (1996) to FCAR models. The authors compare

the "naive" method of Fan and Yao (2003) to the bootstrap and multistage methods. This comparison is done for both univariate and multivariate FCAR models. The authors find that the bootstrap method out performs the other two methods for non-linear forecasting and performs well for forecasting a linear process.

1.4 Spline-Backfitted Kernel Smoothing

The nonparametric and semiparametric methods discussed so far either suffer from the curse of dimensionality, are not computationally expedient, or are not theoretically reliable. The spline-backfitted kernel (SBK) estimation method is both computationally expedient and theoretically reliable. These estimators are also reliable for handling high-dimensional data, thus overcoming the curse of dimensionality. The SBK method was first proposed by Wang and Yang (2007). They developed the method for estimating NAAR models. The method finds "pilot estimates" $\hat{f}_j(X_{t-j})$ for each $f_i(X_{t-i})$, $i = 1, \ldots, p$, in (1.4) through an under-smoothed centered standard spline procedure. These estimates are then used to find pseudo-responses $\hat{Y}_{t,\alpha}$ through a backfitting procedure given by

$$\hat{Y}_{t,\alpha} = Y_t - \hat{c} - \sum_{1 \leq j \leq p, j \neq \alpha} \hat{f}_j(X_{t-j}),$$

where

$$\hat{c} = \frac{1}{n} \sum_{t=1}^{n} Y_t.$$

This procedure is similar to a dimension reduction method since now we can estimate $f_\alpha(X_{t-\alpha})$ by its Nadaraya-Watson estimator

$$\tilde{f}_{\mathrm{SBK},\alpha}(X_{t-\alpha}) = \frac{\sum_{t=1}^{n} K_h(X_{t-\alpha} - x_\alpha) \hat{Y}_{t,\alpha}}{\sum_{t=1}^{n} K_h(X_{t-\alpha} - x_\alpha)},$$

where $K_h(\cdot)$ is defined as in (1.2).

The SBK method is adapted for i.i.d. data in Wang and Yang (2009), to generalized additive models in Liu, Yang, and Härdle (2011), and to partially linear additive models in Ma and Yang (2011). In Song and Yang (2010), a spline-backfitted spline (SBS) procedure is proposed. The SBS procedure is similar to the SBK procedure except that it uses splines to estimate $f_\alpha(X_{t-\alpha})$ with the pseudo-responses

$\hat{Y}_{t,\alpha}$ instead of kernels. This method is even more computationally expedient but lacks point-wise confidence bands. Liu and Yang (2010) propose the SBK method for additive coefficient models which are generalized forms of FCAR models.

1.4.1 Asymptotic Results

The most appealing aspect of SBK estimators is that they overcome the curse of dimensionality while having reliable asymptotic results. Wang and Yang (2007) first present the convergence and asymptotic results of the oracle estimator and then develop the results for the SBK estimator for NAAR models. The oracle estimator assumes the $f(X_{t-j})$, $1 \leq j \leq p$, $j \neq \alpha$, are known by "oracle" and derives the pseudo-responses

$$Y_{t,\alpha} = Y_t - c - \sum_{1 \leq j \leq p, j \neq \alpha}^{p} f_j(X_{t-j}).$$

Using the Nadaraya-Watson kernel smoother with these pseudo-responses gives the oracle estimate $\hat{f}_{\text{oracle},\alpha}(X_{t-\alpha})$. The convergence and asymptotic results for the oracle estimator were first found by Bosq (1998) to be

$$\sup_{x \in [h, 1-h]} \left| \hat{f}_{\text{oracle},\alpha}(x) - f_\alpha(x) \right| = o_p\left(n^{-2/5} \log n\right)$$

and

$$\sqrt{nh}\left\{ \tilde{f}_{\text{oracle},\alpha}(x) - f_\alpha(x) - b_\alpha(x) h^2 \right\} \xrightarrow{D} N\left\{0, v_\alpha^2(x)\right\},$$

where $\mathbf{X}$ has density function $f_{X,\alpha}(x)$, $-1 \leq x \leq 1$,

$$b_\alpha(x) = \int u^2 K(u)\, du \left\{ f''_\alpha(x) f_{X,\alpha}(x)/2 + f'_\alpha(x) f'_{X,\alpha}(x) \right\} f_{X,\alpha}^{-1}(x),$$

and

$$v_\alpha^2(x) = \int K^2(u)\, du E\left[\sigma^2(X_1, \ldots, X_d) | X_\alpha = x\right] f_{X,\alpha}^{-1}(x).$$

Wang and Yang (2007) found the difference between $\hat{f}_{\text{SBK},\alpha}$ and $\hat{f}_{\text{oracle},\alpha}$ is of order $o_p\left(n^{-2/5}\right)$ and that $\hat{f}_{\text{SBK},\alpha}$ has the same asymptotic distribution as $\hat{f}_{\text{oracle},\alpha}$. Liu and Yang (2010) extend these results for additive coefficient models and found similar results for using local linear smoothing instead of Nadaraya-Watson smoothing.

1.5 Overview of Dissertation

This dissertation is organized as follows. In Chapter Two, we will adapt the SBK estimation method to FCAR models and compare the fit to the local linear estimation method. The SBK method has been adapted to additive coefficient models in Liu and Yang (2010) but the method has not been adapted to the specific case of FCAR models. We provide the convergence and asymptotic results for the SBK method for the FCAR model. The comparison of the two methods is shown through simulation results.

In Chapter Three, we will adapt the naive, multistage, and bootstrap forecasting methods to the SBK method and examine prediction intervals for the forecasts. Forecasting performance of the SBK method has not been explored in the literature yet for any type of model. We compare the three forecasting methods through simulation results.

Chapter Four presents a data set consisting of solar irradiance measurements obtained at the La Ola photovoltaic plant in Lanai, Hawaii. Spatio-temporal models are explored for this data set with the SBK method used to model the time structure. This analysis of the Lanai data set is part of an ongoing project aimed at forecasting the irradiance measurements of photovoltaic systems of different sizes and designs.

CHAPTER TWO

Spline-Backfitted Kernel Estimation for FCAR Models

2.1 Introduction

In this chapter, we present the spline-backfitted kernel (SBK) estimators for functional-coefficient autoregressive (FCAR) models. Using notation similar to Liu and Yang (2010), define the FCAR model as

$$X_t = c + \sum_{l=1}^{p} m_l(U_t) X_{t-l} + \varepsilon_t. \tag{2.1}$$

This notation is analogous to (1.1) with $U_t = X_{t-d}$ is the delay variable, and $\{\varepsilon_t\}_{t=1}^n$ is a sequence of i.i.d. random variables with mean 0 and variance σ^2. We denote a model taking the form of (2.1) as FCAR(p, d) where p is the order of the model and d is the delay.

The SBK estimation method was first developed by Wang and Yang (2007) for nonlinear additive autoregressive (NAAR) models. The NAAR model is defined as

$$X_t = c + \sum_{l=1}^{p} f_l(X_{t-l}) + \varepsilon_t.$$

In this model, we are interested in estimating $f_l(X_{t-l})$, whereas in the FCAR model, we are interested in estimating $m_l(U_t)$. The SBK method is loosely based on the backfitting algorithm of Hastie and Tibshirani (1990) and uses the idea of oracle smoothing with pre-estimated component functions. The SBK method combines the computational speed of spline estimators and the asymptotic properties of kernel smoothing. Wang and Yang (2007) developed the method using a constant spline procedure to pre-estimate the component functions and a Nadaraya-Watson estimator to estimate the function of interest. Wang and Yang (2009) adapted the SBK method for additive regression models but used local linear kernel smoothing instead of the Nadaraya-Watson estimator. Ma and Yang (2011) used linear splines and Nadaraya-Watson estimators for partially linear additive models and Liu, Yang, and Härdle (2011) used the same for generalized linear models. The choice of the

degree of the spline and the kernel procedures does not make a difference in the performance of the method but does play a role in the number of knots used in the spline procedure. In this dissertation, we will use a linear spline with a local linear kernel fit in both the theoretical and simulation results.

The remainder of this chapter is organized as follows. We describe the oracle estimator in Section 2.2 and adapt the SBK method to FCAR models in Section 2.4. We present the asymptotic results in Section 2.4 and simulation results in Section 2.5. We conclude this chapter in Section 2.6.

2.2 The Oracle Estimator

As a benchmark for evaluating the SBK estimator, we introduce the oracle estimator of Wang and Yang (2007). Define $\mathbf{U} = (U_1, \ldots, U_n)^T$,

$$\mathbf{X}_j = \left(X_{\max\{p,d\}-j}, \ldots, X_{n-j}\right)^T,$$

and $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_p)$. Suppose we are estimating $m_1(\mathbf{U})$ in (2.1). If we know $m_\alpha(\mathbf{U})$, $\alpha = 2, \ldots, p$, by "oracle," then we can construct pseudo-responses $Y_{t,1}$ as

$$Y_{t,1} = X_t - c - \sum_{l=2}^{p} m_l(U_t) X_{t-l}. \tag{2.2}$$

Denote the vector of pseudo-responses $\mathbf{Y}_1 = (Y_{1,1}, \cdots, Y_{n,1})^T$. We can use these pseudo-responses to estimate $m_1(\mathbf{U})$ by solving a kernel weighted least squares problem

$$\tilde{m}_1(u) = \underset{\lambda=(\lambda_l)_{1\le l\le p}}{\operatorname{argmin}} L_{m_1}(\boldsymbol{\lambda}, u),$$

in which

$$L_{m_1}(\boldsymbol{\lambda}, u) = \sum_{t=1}^{n} \left(Y_{t,1} - \sum_{l=2}^{p} \lambda_l X_{tl}\right)^2 K_h(U_t - u),$$

where $K_h(u) = K(u/h)/h$ for a kernel function K and bandwidth h that satisfy Assumption (vi) in Appendix A. We can rewrite the oracle smoother in matrix form as

$$\tilde{m}_1(u) = (\mathbf{I}_{p\times p}, \mathbf{0}_{p\times p}) \left(\frac{1}{n}\mathbf{C}_1^T \mathbf{W}_1 \mathbf{C}_1\right)^{-1} \frac{1}{n}\mathbf{C}_1 \mathbf{W}_1 \mathbf{Y}_1,$$

where

$$\mathbf{C}_1 = (\mathbf{X}_1, \cdots, \mathbf{X}_p, \mathbf{X}_1(\mathbf{U} - u), \cdots, \mathbf{X}_p(\mathbf{U} - u)), \tag{2.3}$$

and

$$\mathbf{W}_1 = \text{diag}\left\{K_h\left(U_1 - u\right), \ldots, K_h\left(U_n - u\right)\right\}.$$

Clearly, these oracle estimators are not useful since they are computed based on the knowledge of the unknown functions $m_\alpha(\mathbf{U})$, $\alpha = 2, \ldots, p$. However, they do provide motivation for the SBK estimators.

2.3 Spline-Backfitted Kernel Estimators

In this section, we describe how the unknown coefficient functions $\{m_l(\cdot)\}_{l=2}^p$ can be pre-estimated by linear splines and then substituted into (2.2). Suppose that U is distributed on a compact interval $[a, b]$, and without loss of generality, suppose $[a, b] = [0, 1]$. We preselect an integer $N \sim N_n = n^{1/4} \log n$ and define $H = (N+1)^{-1}$ (see Assumption (vii) in Appendix A). Let $0 = \kappa_0 < \kappa_1 < \cdots < \kappa_N < \kappa_{N+1} = 1$ denote a sequence of equally spaced points, called interior knots, and denote the degenerate knots $\kappa_{-1} = 0$, $\kappa_{N+2} = 1$. We define the linear B-spline basis as

$$b_J(u) = \left(1 - |u - \kappa_J|/H\right)_+ = \begin{cases} (N+1)u - J + 1, & \kappa_{J-1} \le u < \kappa_J, \\ J + 1 - (N+1)u, & \kappa_J \le u \le \kappa_{J+1}, \\ 0, & \text{otherwise.} \end{cases}$$

The spline estimator of $m_l(u)$, $l = 2, \ldots, p$, is

$$\hat{m}_{s,l}(u) = \hat{\lambda}_0 + \sum_{J=0}^{N+1} \hat{\lambda}_J b_J(u),$$

where the coefficients $\left(\hat{\lambda}_0, \hat{\lambda}_1, \ldots, \hat{\lambda}_N\right)$ are solutions of the least squares problem

$$\left\{\hat{\lambda}_0, \hat{\lambda}_1, \ldots, \hat{\lambda}_N\right\}^T = \underset{\Re^{N+1}}{\text{argmin}} \sum_{t=1}^{n} \left\{X_t - \left(\lambda_0 - \sum_{J=0}^{N+1} \lambda_J b_J(u)\right) x_{t-l}\right\}^2.$$

The estimated oracle pseudo-responses are defined as

$$\hat{Y}_{t,1} = X_t - \hat{c} - \sum_{l=1}^{p} \hat{m}_{s,l}(U_t) X_{t-l},$$

where $\hat{c} = 1/n \sum_{t=1}^n X_t$. Denote the vector of pseudo-responses as $\hat{\mathbf{Y}}_1 = \left(\hat{Y}_{1,1}, \cdots, \hat{Y}_{n,1}\right)^T$. The spline-backfitted kernel (SBK) estimator is defined as

$$\hat{m}_1(u) = (\mathbf{I}_{p\times p}, \mathbf{0}_{p\times p}) \left(\frac{1}{n}\mathbf{C}_1^T \mathbf{W}_1 \mathbf{C}_1\right)^{-1} \frac{1}{n}\mathbf{C}_1 \mathbf{W}_1 \hat{\mathbf{Y}}_1. \tag{2.4}$$

The SBK estimator mimics the oracle estimator but requires an extra estimation step. The marginal integration technique of Linton (1997) is similar to the SBK method but uses kernel-based integration to find the pre-estimates of the functions. Linton establishes that the error caused by this "cheating" is negligible. The idea behind SBK estimation is to under-smooth in the pre-estimates in order to reduce the bias. This under-smoothing leads to a larger variance which is reduced in the kernel estimation step.

2.4 Asymptotic Results

We present the asymptotic results for the oracle estimator $\tilde{m}_1(u)$ in Theorem 2.1. This theorem follows directly from Theorem 1 in Liu and Yang (2010) and the proof below is adapted from the proof of Theorem 1 found in the Appendix of that paper.

Let $\mu_2(K) = \int u^2 K(u)\,du$, $\|K\|_2^2 = \int K(u)^2\,du$, and $\mathbf{Q}(u) = \{q(u)\}_{l,l'=1}^p = E\left(\mathbf{X}\mathbf{X}^T|\mathbf{U}=\mathbf{u}\right)$. Define the bias coefficient as

$$b_{l,l'}(u) = \frac{1}{2}\mu_2(K)\, m_l''(u)\, f(u)\, q_{ll'}(u),$$

where $f(u)$ is defined in Assumption (i) in Appendix A, and the variance coefficient as

$$\{v_{l,l'}(u)\}_{l,l'=1}^p = \mathbf{Q}(u)^{-1}\,\Sigma(u)\,\mathbf{Q}(u)^{-1},$$

where

$$\Sigma(u) = \|K\|_2^2\, f(u)\, E\left(\mathbf{X}\mathbf{X}^T\sigma^2|\mathbf{U}=\mathbf{u}\right).$$

Theorem 2.1. Under Assumptions (i) - (vi) in Appendix A, for any $u \in [h, 1-h]$, as $n \to \infty$, the oracle local linear smoother $\tilde{m}_1(u)$ satisfies

$$\sqrt{nh}\left[\tilde{m}_1(u) - m_1(u) - \left\{\sum_{l=1}^p b_{l,l'}(u)\right\}_{l'=1}^p h^2\right] \to N\left(0, \{v_{l,l'}(u)\}_{l,l'=1}^p\right).$$

Proof. Define

$$\xi_{i,n,l'} = \xi_{i,n,l'}(u, \mathbf{u}, \mathbf{x}_i) = X_{il'}\sigma\varepsilon_i K_h(U_i - u)\left(\frac{U_i - u}{h}\right),$$

and let

$$V_{l'}(u) = \frac{1}{n}\sum_{i=1}^n \xi_{i,n,l'}.$$

For any $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)^T \in \Re^p$,

$$\boldsymbol{\lambda}^T \{V_{l'}(u)\}_{l'=1}^p = \sum_{l'=1}^p \lambda_{l'} V_{l'}(u) = \frac{1}{n} \sum_{i=1}^n \sum_{l'=1}^p \lambda_{l'} \xi_{i,n,l'}.$$

Define $\xi_{i,n} = \sum_{l'=1}^p \lambda_{l'} \xi_{i,n,l'}$ and $S_n = S_n(u) = \sum_{i=1}^n \xi_{i,n} = n\boldsymbol{\lambda}^T \{V_{l'}(u)\}_{l'=1}^p$, then we have $E[S_n] = 0$. Let

$$\gamma(k) = \gamma(k,u) = \operatorname{cov}(\xi_{i,n}, \xi_{i+k,n}) = \operatorname{cov}\left(\sum_{l'=1}^p \lambda_{l'} \xi_{i,n,l'}, \sum_{l'=1}^p \lambda_{l'} \xi_{i+k,n,l'}\right)$$

and

$$\begin{aligned}
\sigma_n^2 &= E\left[S_n^2\right] = \operatorname{var}(S_n) = \operatorname{var}\left(\sum_{i=1}^n \xi_{i,n}\right) = \sum_{i=1}^n \operatorname{var}(\xi_{i,n}) + \sum_{i \neq j}^n \operatorname{cov}(\xi_{i,n}, \xi_{j,n}) \\
&= n\operatorname{var}(\xi_{i,n}) + n \sum_{1 \leq |k| \leq qn-1} \left(1 - \frac{|k|}{n}\right) \gamma(k) = n\operatorname{var}(\xi_{i,n}) + nA_n.
\end{aligned}$$

In the above, $\operatorname{var}(\xi_{i,n}) = \operatorname{var}(\sum_{l'=1}^p \lambda_{l'} \xi_{i,n,l'}) = h^{-1} \boldsymbol{\lambda}^T \Sigma \boldsymbol{\lambda}$ where

$$\begin{aligned}
\Sigma &= h\{\operatorname{cov}(\xi_{i,n,l'}, \xi_{i,n,l''})\}_{l',l''=1}^p = hE\{\xi_{i,n,l'}, \xi_{i,n,l''}\}_{l',l''=1}^p \\
&= f(u) \|K\|_2^2 E\left\{\mathbf{X}^T \mathbf{X} \sigma | \mathbf{U} = \mathbf{u}\right\}
\end{aligned}$$

by Lemma A.15 in Liu and Yang (2010). While according to Lemma A.16 in Liu and Yang, we have

$$|\gamma(k)| \leq p^2 \max_{1 \leq l' \leq p} |\operatorname{cov}(\xi_{i,n,l'}, \xi_{i+k,n,l''})| \leq Ch^{-\frac{1+\eta}{2+\eta}} \alpha(k)^{\frac{\eta}{2+\eta}},$$

where η is defined in Assumption (v) in Appendix A. Hence,

$$\begin{aligned}
|A_n| &= \left| \sum_{1 \leq |l| \leq n-1} \gamma(k) \right| \leq \sum_{1 \leq |l| \leq n-1} \left(1 - \frac{|k|}{n}\right) h^{-\frac{1+\eta}{2+\delta}} \{K_0 \exp(-\lambda_0 k)\}^{\frac{\eta}{2+\eta}} \\
&\leq K_0 h^{-\frac{1+\eta}{2+\delta}} \sum_{1 \leq |l| \leq n-1} \exp\{-\lambda_0 k\eta/(2+\eta)\},
\end{aligned}$$

so there exists a constant C_1 such that $A_n \leq C_1 h^{-\frac{1+\eta}{2+\delta}}$. So $A_n/\operatorname{var}(\xi_{i,n}) \to 0$ as $n \to \infty$. Then $\sigma_n^2 \sim n\operatorname{var}(\xi_{i,n}) \geq c_0 n$ where n is large, so according to Lemma A.17 in Liu and Yang, there exist constants c_1 and c_2 such that for some $0 < \eta \leq 1$

$$\Delta_n = \sup_z \left|P\left\{\sigma_n^{-1} S_n < z\right\} - \Phi(z)\right| \leq c_1 \frac{d_\eta}{c_0 \sigma_n^\eta} \left\{\log\left(\frac{\sigma_n}{c_0^{1/2}}\right)/\lambda\right\}^{1+\eta},$$

for any λ with $\lambda_1 \leq \lambda \leq \lambda_2$, where

$$\lambda_1 = c_2 \left\{\log\left(\frac{\sigma_n}{c_0^{1/2}}\right)\right\}^b / n, b > \frac{2(1+\eta)}{\eta}; \lambda_2 = \frac{4(2+\eta)}{\eta} \log\left(\frac{\sigma_n}{c_0^{1/2}}\right).$$

For η, set $\lambda = 4(2+\eta)\eta^{-1}\log\left(\sigma_n/c_0^{1/2}\right)$, then we have

$$\begin{aligned} d_\eta &= \max_{1\le i\le n}\left\{E\left|\sum_{l'=1}^{p}\lambda_{l'}\xi_{i,n,l'}\right|^{2+\eta}\right\} \\ &= \max_{1\le i\le n}\left\{E\left|\sum_{l'=1}^{p}\lambda_{l'}X_{il'}\sigma\varepsilon_i K_h(U_i-u)\left(\frac{U_i-u}{h}\right)\right|^{2+\eta}\right\} \\ &\le CC_\delta C_\eta\left\{E\left|\sum_{l'=1}^{p}K_h(U_i-u)\left(\frac{U_i-u}{h}\right)\right|^{2+\eta}\right\} = O\left\{h^{-(1+\eta)}\right\}, \end{aligned}$$

i.e., $\Delta_n = O\left\{h^{-(1+\eta)}/\sigma_n^\eta\right\} = O\left\{n^{(1+\eta/2)/5-\eta/2}\right\} = O\left(n^{1/5-2\eta/5}\right) \to 0$ when $1/2 < \eta < 1$. So $S_n/\sigma_n \to N(0,1)$, then $\sqrt{nh}\boldsymbol{\lambda}^T\left\{V_{l'}(u)\right\}_{l'=1}^p \to N\left(0, \boldsymbol{\lambda}^T\Sigma\boldsymbol{\lambda}\right)$. By Cramér-Wold device, we have $\sqrt{n}\left\{V_{l'}(u)\right\}_{l'=1}^p \to N(0,\Sigma)$. Then according to Slutsky's theorem, we have

$$\sqrt{nh}E\left(\mathbf{X}^T\mathbf{X}|\mathbf{U}=\mathbf{u}\right)\left\{\tilde{m}_1(u) - m_1(u) - \sum_{l=1}^{p}b_{l,l'}(u)h^2\right\}_{l'=1}^{p} \to N(0,\Sigma)$$

i.e.,

$$\sqrt{nh}\left\{\tilde{m}_1(u) - m_1(u) - \sum_{l=1}^{p}b_{l,l'}(u)h^2\right\}_{l'=1}^{p} \to N\left(0, \mathbf{Q}(u)^{-1}\Sigma\mathbf{Q}(u)^{-1}\right).$$

□

The convergence result for $\tilde{m}_1(u)$ is provided in Theorem 2.2 and the convergence results for $\hat{m}_1(u)$ is provided in Theorems 2.3. These theorems follow directly from Theorems 2 and 3 in Liu and Yang (2010) and their proofs can be adapted from the proofs found in the Appendix of that paper similar to the proof of Theorem 2.1.

Theorem 2.2. Under Assumptions (i) - (vi) in Appendix A, as $n \to \infty$, the oracle local linear smoother $\tilde{m}_1(u)$ satisfies

$$\sup_{u\in[h,1-h]}\left|\tilde{m}_1(u) - m_1(u)\right| = O_p\left(\frac{\log n}{\sqrt{nh}}\right).$$

Theorem 2.3. Under Assumptions (i) - (vii) in Appendix A, as $n \to \infty$, the SBK estimator $\hat{m}_1(u)$ satisfies

$$\sup_{u\in[0,1]}\left|\hat{m}_1(u) - \tilde{m}_1(u)\right| = o_p\left(n^{-2/5}\right).$$

Note that in Theorem 2.3, the asymptotic uniform magnitude of the difference between $\hat{m}_1(u)$ and $\tilde{m}_1(u)$ is dominated by the asymptotic size of $\tilde{m}_1(u) - m_1(u)$. As a result, $\hat{m}_1(u)$ will have the same asymptotic distribution as $\tilde{m}_1(u)$. This result leads to the following corollary which follows Corollary 1 in Liu and Yang (2010).

Corollary 2.1. Under Assumptions (i) - (vii) in Appendix A, for any $u \in [h, 1-h]$, as $n \to \infty$, the SBK estimator $\hat{m}_1(u)$ satisfies

$$\sqrt{nh}\left[\hat{m}_1(u) - m_1(u) - \left\{\sum_{l=1}^{p} b_{l,l'}(u)\right\}_{l'=1}^{p} h^2\right] \to N\left(0, \{v_{l,l'}(u)\}_{l,l'=1}^{p}\right).$$

Note that the above theorems and corollary hold for $\hat{m}_\alpha(u)$ and $\tilde{m}_\alpha(u)$ similarly constructed for any $\alpha = 2, \ldots, p$.

2.5 Simulation Results

We ran simulations to test the performance of the SBK estimators and to compare them to the local linear (LL) method of Cai, Fan, and Yao (2000) (see Section 1.2.1). Realizations from four models were generated of lengths $n = 75, 150, 250, 500$. The realizations were from a exponential autoregressive (EXPAR) model and a self-exciting threshold autoregressive (SETAR) model each with order $p = 2, 4$. We ran $S = 500$ iterations for each model and computed the relative efficiency. For the jth sample, $j = 1, 2, \ldots, S$, the relative efficiency of $\hat{m}_\alpha$ with respect to $\tilde{m}_\alpha$ is defined as

$$\begin{aligned} \text{eff}_{\alpha,j} &= \frac{\frac{1}{n}\sum_{t=1}^{n}\left\{\hat{m}_\alpha(X_{t\alpha,j}) - m_\alpha(X_{t\alpha,j})\right\}^2}{\frac{1}{n}\sum_{t=1}^{n}\left\{\tilde{m}_\alpha(X_{t\alpha,j}) - m_\alpha(X_{t\alpha,j})\right\}^2} \\ \text{eff}_\alpha &= \frac{1}{S}\sum_{j=1}^{S}\text{eff}_{\alpha,j}, = 1, 2, \ldots, p. \end{aligned}$$

Theorems 2.1, 2.2, 2.3, and Corollary 2.1 indicate that the efficiency should be close to 1. For comparisons of total fit, we calculate the root mean square error (RMSE) for both the LL and the SBK methods. For the SBK method, the RMSE is defined as

$$\text{RMSE} = \left[\frac{1}{n}\sum_{t=1}^{n}\left(\hat{x}_{\text{SBK},t} - x_t\right)^2\right]^{1/2} \tag{2.5}$$

where

$$\hat{x}_{\text{SBK},t} = \hat{c} + \sum_{l=1}^{p}\hat{m}_l(u_t)\, x_{t-l}.$$

Similarly, we can calculate the RMSE of the LL method with $\hat{x}_{\text{LL},t}$ substituted into (2.5) for $\hat{x}_{\text{SBK},t}$ where

$$\hat{x}_{\text{LL},t} = \hat{c} + \sum_{l=1}^{p} \tilde{m}_l(u_t)\, x_{t-l}.$$

The EXPAR model was introduced by Haggan and Ozaki (1981) to model time series of non-linear random vibrations. The EXPAR model takes the form of (2.1) with $m_l(u) = a_l + (b_l + c_l u)\exp\{-\delta u^2\}$ where a_l, b_l, c_l, and δ are constants. The SETAR model (Tong and Lim, 1980) takes the form

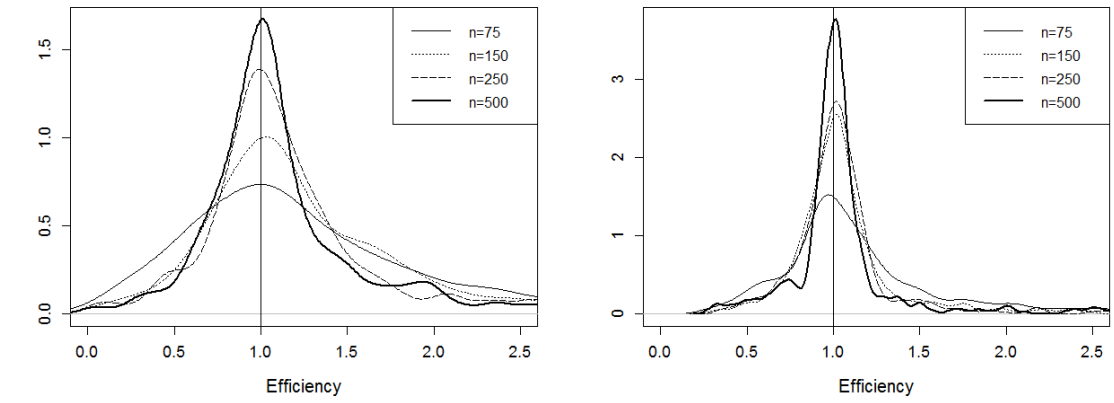
$$X_t = \phi_1^{(j)} X_{t-1} + \cdots + \phi_p^{(j)} X_{t-p} + \varepsilon_t \quad \text{if } X_{t-d} \in \Omega_j, \quad j = 1, \ldots, k,$$

where $\{\Omega_j\}$ form a partition of $\Re$. We chose to use the EXPAR model because the coefficient functions are continuous and the SETAR model to gauge the performance of the method when the coefficients were not continuous at all points.

Example 2.1. We first consider an EXPAR(2) model with

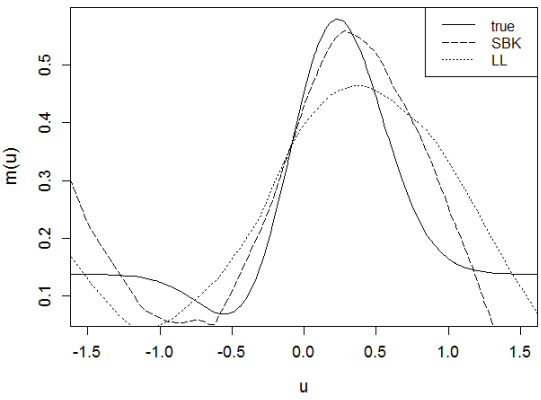
$$\begin{aligned} m_1(u) &= 0.138 + (0.316 + 0.982u)\exp\left\{-3.89u^2\right\}, \\ m_2(u) &= -0.437 - (0.659 + 1.260u)\exp\left\{-3.89u^2\right\}, \end{aligned}$$

$U_t = X_{t-2}$, and $\varepsilon_t \sim N(0,1)$. The densities of the empirical relative efficiencies were computed and can be found in Figure 2.1 (a) and (b). All of the series lengths had relative efficiencies close to 1 with the relative efficiencies for the larger series lengths centering closer to 1. We can also see that the variability of the densities is decreasing as the series lengths increase. These results are what we expect to see in terms of efficiency. For a comparison of the SBK method to the LL method, Figure 2.1 (c) shows the fits of both methods to an realization of this model for $l = 1$. We can see boxplots of the RMSE's for $n = 150$ for both methods in (d). These boxplots show that the SBK method has, on average, a lower RMSE but with slightly larger variability. The same results can be seen for the remaining series lengths (see Appendix B for remaining plots).

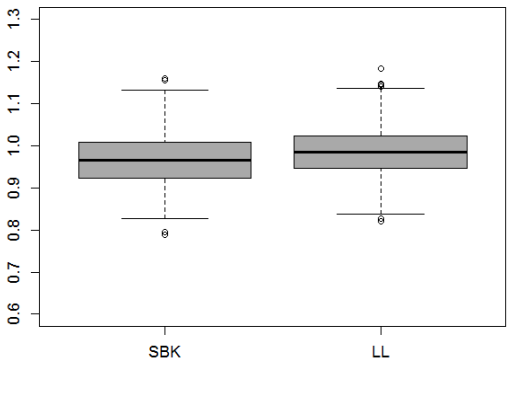


(a)

(b)



(c)



(d)

Figure 2.1: Empirical relative efficiencies for (a) $l = 1$ and (b) $l = 2$, the (c) fit of the coefficient function for $l = 1$, and (d) boxplots of the RMSE's for $n = 150$ for Example 2.1.

Example 2.2. We now consider an EXPAR(4) model with

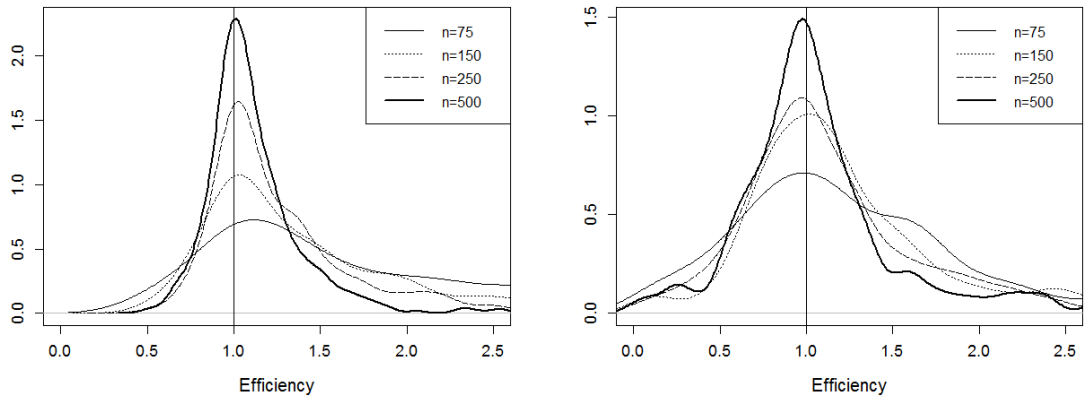
$$\begin{aligned} m_1(u) &= 0.138 + (0.316 + 0.982u)\exp\left\{-3.89u^2\right\}, \\ m_2(u) &= -0.437 - (0.346 + 1.260u)\exp\left\{-3.89u^2\right\}, \\ m_3(u) &= 0.138 + (0.659 + 0.982u)\exp\left\{-3.89u^2\right\}, \\ m_4(u) &= -0.437 - (0.659 + 1.260u)\exp\left\{-3.89u^2\right\}, \end{aligned}$$

$U_t = X_{t-2}$, and $\varepsilon_t \sim N(0,1)$. The empirical relative efficiencies for $l = 2$ and $l = 3$ can be seen in Figure 2.2 (a) and (b). The results for this example are similar to the results of Example 2.1 in that the efficiencies are close to one and the variability becomes smaller as the series length increases. With a degree of four, we would expect the SBK method to perform better in RMSE than the LL method for smaller series lengths since the SBK method does not suffer from the curse of dimensionality as much as the LL method. Figure 2.2 (c) and (d) show the boxplots of the RMSE's for $n = 75$ and $n = 150$. We see for $n = 75$ that the RMSE of the SBK method is, on average, lower than that of the LL method. For $n = 150$, the average RMSE's for the two methods become closer together. As the series length increases, the boxplots of the RMSE's for the two methods are approximately identical (see Appendix B for remaining plots).

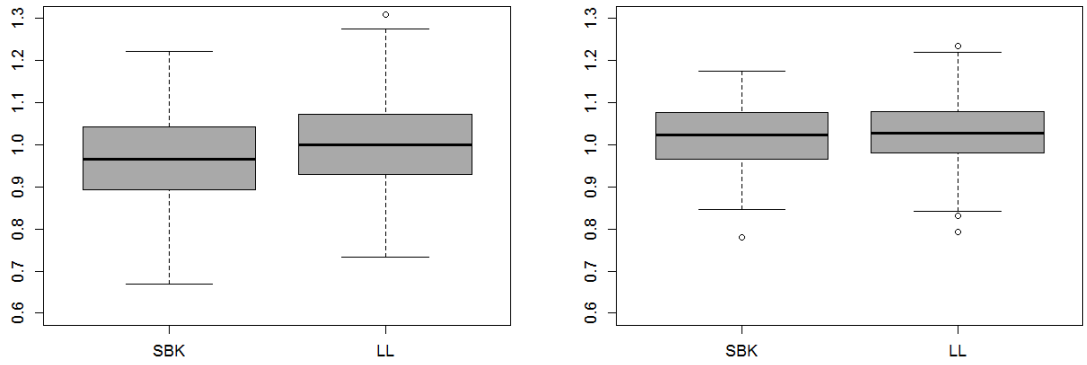
Example 2.3. To test the SBK method for when the coefficient function is not continuous at all points, we consider a SETAR(2) model with

$$\begin{aligned} m_1(u) &= 0.4I_{[u\leq 1]} - 0.8I_{[u>1]}, \\ m_2(u) &= -0.6I_{[u\leq 1]} + 0.2I_{[u>1]}, \end{aligned}$$

$U_t = X_{t-2}$, and $\varepsilon_t \sim N(0,1)$. The empirical relative efficiencies for $l = 1$ and $l = 2$ are shown in Figure 2.3 (a) and (b). The coefficient function and the fits of the SBK and LL methods can be seen in (c). We can see that both methods are biased around the point of discontinuity although, for the iteration shown, the SBK method is not as biased. The efficiencies seem to be effected by this bias since the efficiencies for $l = 1$ seem to be centered just above 1. For $l = 2$, we see that the efficiencies are severely off from what we expect. Although the efficiencies for all series lengths are



(a) (b)



(c) (d)

Figure 2.2: Empirical relative efficiencies for (a) $l = 2$ and (b) $l = 3$, the boxplots of the RMSE's for (c) $n = 75$ and (d) $n = 150$ for Example 2.2.

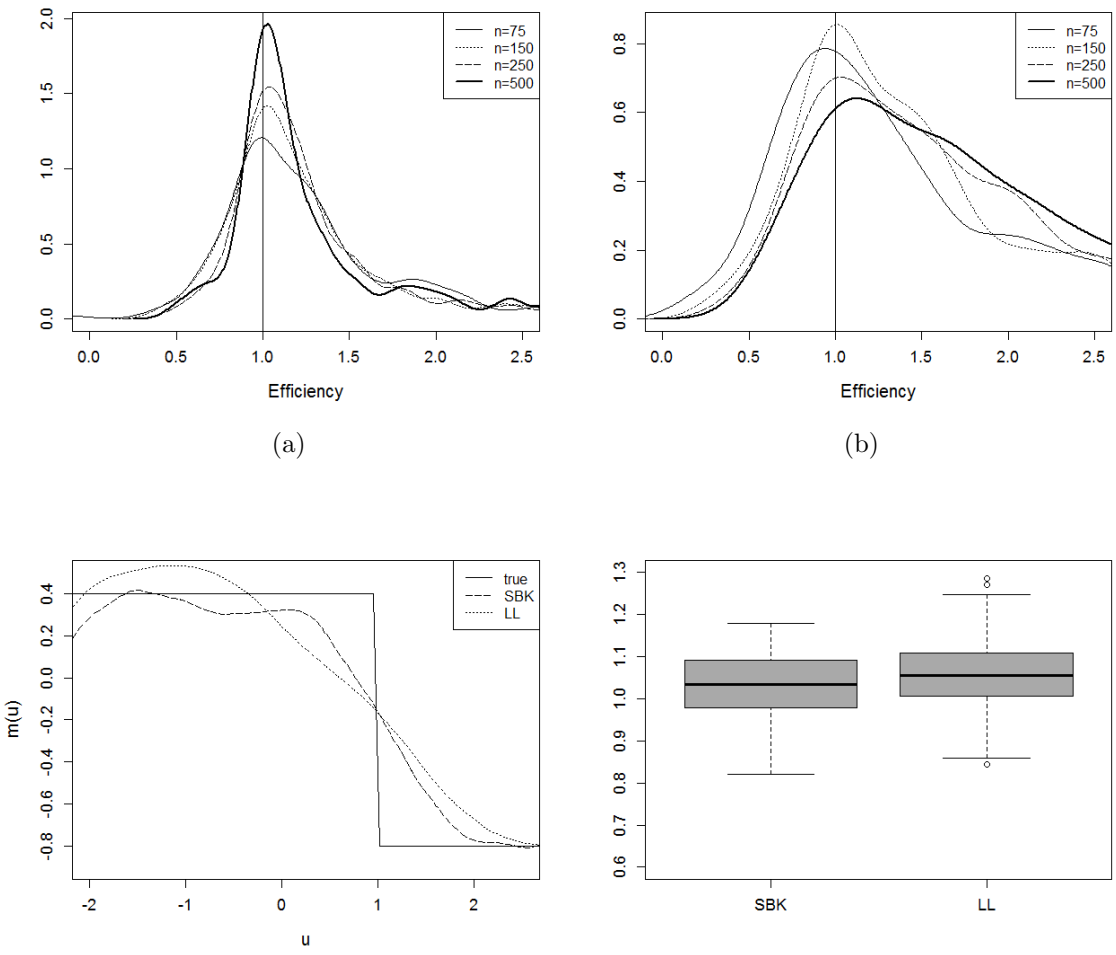


Figure 2.3: Empirical relative efficiencies for (a) $l = 1$ and (b) $l = 2$, the (c) fit of the coefficient function for $l = 1$, and (d) boxplots of the RMSE's for $n = 150$ for Example 2.3.

centered close to 1, the larger series lengths appear to be more likely to have very large efficiency values. In terms of RMSE, the SBK method performed better than the LL method for the smaller series lengths. The two methods seem to have more similar RMSE as the series lengths increase (see Appendix B for remaining plots). These RMSE results are similar to the results of Examples 2.1.

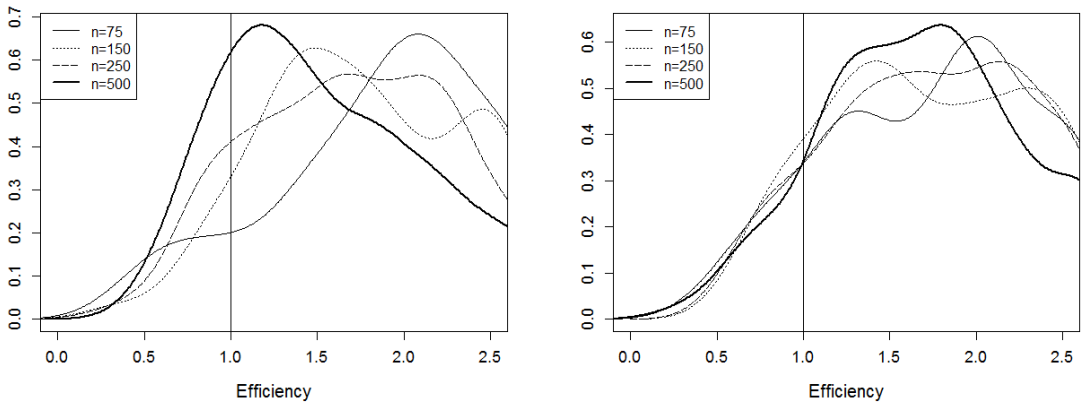
Example 2.4. Like Example 2.2, we would like to compare the SBK and LL methods for larger orders. To do this, we consider a SETAR(4) model with

$$\begin{aligned} m_1(u) &= 0.84I_{[u\leq 1.5]} + 1.44I_{[u>1.5]}, \\ m_2(u) &= 0.07I_{[u\leq 1.5]} - 0.84I_{[u>1.5]}, \\ m_3(u) &= -0.32I_{[u\leq 1.5]} + 0.06I_{[u>1.5]}, \\ m_4(u) &= 0.15I_{[u\leq 1.5]} + 0I_{[u>1.5]}, \end{aligned}$$

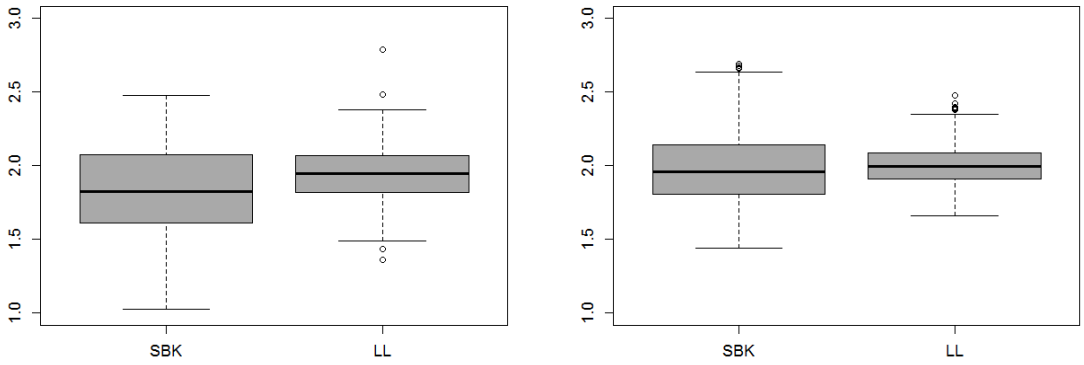
$U_t = X_{t-1}$, and $\varepsilon_t \sim N(0,4)$. The empirical relative efficiencies for $l = 1$ and $l = 2$ are shown in Figure 2.3 (a) and (b). Similar to Example 2.3, the SBK and LL methods both were biased when estimating the coefficient function at the point of discontinuity. We can see the effect of this bias in the relative efficiencies for $l = 1$ and $l = 2$ in Figure 2.4 (a) and (b). In terms of RMSE, the SBK method is on average smaller than the LL method. However, the variability of the SBK method appears much greater. The boxplots of the RMSE's for $n = 75$ and $n = 150$ can be seen in Figure 2.4 (c) and (d). We see similar results for $n = 250$ and $n = 500$ (see Appendix B for remaining plots).

2.6 Conclusion

We have introduced the SBK method as a better alternative to the LL method for estimating FCAR models, particularly for higher orders. We have shown that the SBK method is as asymptotically efficient as the oracle estimator through theoretical and simulation results. This method assumes that the coefficient functions are continuous at all points. For Examples 2.1 and 2.2 which had continuous coefficient functions, the empirical relative efficiencies were as we expected based on the theoretical results. Also, for these examples, the SBK method had lower RMSE than the LL method. For Examples 2.3 and 2.4, which did not have coefficient functions continuous at all points, the empirical relative efficiencies were not as we expected. When compared to the LL method, the SBK method had a smaller average RMSE for all series lengths. However, the variability of the SBK method was much greater than the LL method when the model was of order four. It is difficult to recommend



(a) (b)



(c) (d)

Figure 2.4: Empirical relative efficiencies for (a) $l = 1$ and (b) $l = 2$, the boxplots of the RMSE's for (c) $n = 75$ and (d) $n = 150$ for Example 2.4.

the SBK method over the LL method with this large variability for the case of coefficient functions not continuous at all points. In all other cases, we have shown the SBK method is a better method for estimating FCAR models.

CHAPTER THREE

Forecasting Using the Spline-Backfitted Kernel Method

3.1 Introduction

The spline-backfitted kernel estimation (SBK) method has been shown in the literature to be effective in estimating high-dimensional time series models (see, e.g., Wang and Yang, 2007). In Chapter Two, we showed the SBK method is effective in modeling functional-coefficient autoregressive (FCAR) models through theoretical and simulation results. The forecasting performance of the SBK method has not been addressed in the literature. In this chapter, we adapt three forecasting methods for FCAR models to be used with the SBK method.

This chapter is organized as follows. In Section 3.2, we discuss how the three forecasting methods can be used after the FCAR model is fitted by the SBK method. In Section 3.3, we discuss prediction intervals for these forecasting methods. We present simulation results in Section 3.4 and state a conclusion in Section 3.5.

3.2 Forecasting Methods

For convenience, we restate (1.1) here: the FCAR model is defined as

$$X_t = m_1(U_t) X_{t-1} + \cdots + m_p(U_t) X_{t-p} + \varepsilon_t, \quad t = 1, \ldots, n, \tag{3.1}$$

where p is a positive integer, $m_j(U_t)$ is a measurable function of the variable U_t for $j = 1, \ldots, p$, and $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with mean 0 and variance σ^2. As we discussed in Section 1.3, Harvill and Ray (2005) adapted two forecasting methods for FCAR models and compared them to the direct and iterative methods of Fan and Yao (2003). These methods are the bootstrap and the multi-stage forecasting methods. The authors estimated the FCAR model using local linear smoothing. Similar to Harvill and Ray (2005), we will use SBK estimation for the FCAR model and compare the three forecasting methods.

Assuming $m_l(\cdot)$ is known and U is exogenous in (3.1), we want to find an estimator of the conditional expectation

$$\begin{aligned} E[X_{n+M}|X_n,\ldots,X_{n-p}] &= E\left[\sum_{l=1}^{p} m_l(U_{n+M})\, X_{n+M-l}|X_n,\ldots,X_{n-p}\right] \\ &= \sum_{l=1}^{p} m_l(U_{n+M})\, E[X_{n+M-l}|X_n,\ldots,X_{n-p}] \\ &= \sum_{l=1}^{p} m_l(U_{n+M})\, \hat{X}_{n+M-l}. \end{aligned} \tag{3.2}$$

The expectation in (3.2) is no longer a simple linear operation when $U_t = X_{t-d}$ for some positive constant d. The three forecasting methods described below deal with this expectation in a different way.

3.2.1 Naive Predictor

The naive approach simply ignores the fact the expectation in (3.2) is not a linear function of X_{t+M-l} and substitutes $\hat{X}_{t+M-l}$ into the forecast equation. We estimate the coefficient function only using the within-sample series values. The naive predictor is defined as

$$\hat{X}_{n+M} = \sum_{l=1}^{p} \hat{m}_l\left(\hat{X}_{n+M-d}\right)\hat{X}_{n+M-l},$$

where $\hat{X}_t = X_t$, $t \le n$. For the SBK estimator, $\hat{X}_{n+M-d}$ is substituted for u in (2.2) and $\hat{m}_l(\cdot)$ is the value obtained by the general form of (2.4). Thus, the spline pre-estimate is not computed for each value of $\hat{X}_{n+M-d}$ but the local linear estimation is computed.

3.2.2 Bootstrap Predictor

The bootstrap predictor is like the naive predictor in that it estimates the functional coefficients using only the within-sample values. However, we bootstrap the within-sample residuals from the estimated model and find the predicted value as

$$\hat{X}_{n+M} = \sum_{l=1}^{p} \hat{m}_l\left(\hat{X}_{n+M-d}\right)\hat{X}_{n+M-l} + \epsilon^b,$$

where ϵ^b is the bootstrapped residual. We obtain bootstrapped forecasts for $b = 1,\ldots,B$, and use the average of these values as the M-step ahead forecast. For the

SBK method, we estimate $\hat{m}_l(\cdot)$ as with the naive predictor. An advantage of using the bootstrap values is that the set of all values allows us to estimate the predictive density of X_{t+M}. A disadvantage is that the estimated functional coefficients may become unreliable when $\hat{X}_{t+M-d}$ is outside or near the boundary of the range of the original X_{t-d}. This disadvantage was first noted by Huang and Yang (2004) and reiterated by Harvill and Ray (2005).

3.2.3 Multistage Predictor

Another way to handle the expectation in (3.2) is to incorporate the information from X_t encoded in the predicted response at time $n+j$, $j = 1, \ldots, M-1$. This is accomplished by updating the functional coefficients at each step and obtaining the forecasted value by

$$\hat{X}_{n+M} = \sum_{l=1}^{p} \hat{m}_l^M \left(\hat{X}_{n+M-d} \right) \hat{X}_{n+M-l},$$

where $\hat{X}_t = X_t$, $t \leq n$. The functional coefficient $\hat{m}_l^M(\cdot)$ is estimated by the SBK method at each step. That is, we include the predicted values $\hat{X}_t$, $t = n+1, \ldots, M-1$, with the original values X_t, $t = 1, \ldots, n$, and then re-estimate the functional coefficient with the SBK method using the new set of data.

3.3 Prediction Intervals

A number of approaches exist for computing prediction intervals. Chatfield (1993) states that, in practice, most $100(1-\alpha)\%$ prediction intervals for X_{n+M} follow the form

$$\hat{X}_{n+M} \pm z_{\alpha/2} \sqrt{\text{var}(e_{n+M})}, \tag{3.3}$$

where $z_{\alpha/2}$ denotes the upper $\alpha/2$ critical value for the standard normal distribution and $e_{n+M} = X_{n+M} - \hat{X}_{n+M}$ is the forecast error for $\hat{X}_{n+M}$. When we estimate $\text{var}(e_{n+M})$, then $z_{\alpha/2}$ should be replaced in (3.3) by a critical value from the t-distribution with an appropriate degrees of freedom, but this makes little difference except for short series.

Another approach for finding prediction intervals is to use the bootstrap predictor and find the $100(\alpha/2)\%$ and the $100(1-\alpha/2)\%$ sample quantiles of all of the bootstrapped values. Clearly, using this approach with the bootstrap predictor requires little additional computations. However, using the naive or multistage predictors would require us to bootstrap the residuals in addition to finding the forecasted values.

Chatfield (1993) discusses an "approximate" formula substituted into (3.3) that uses

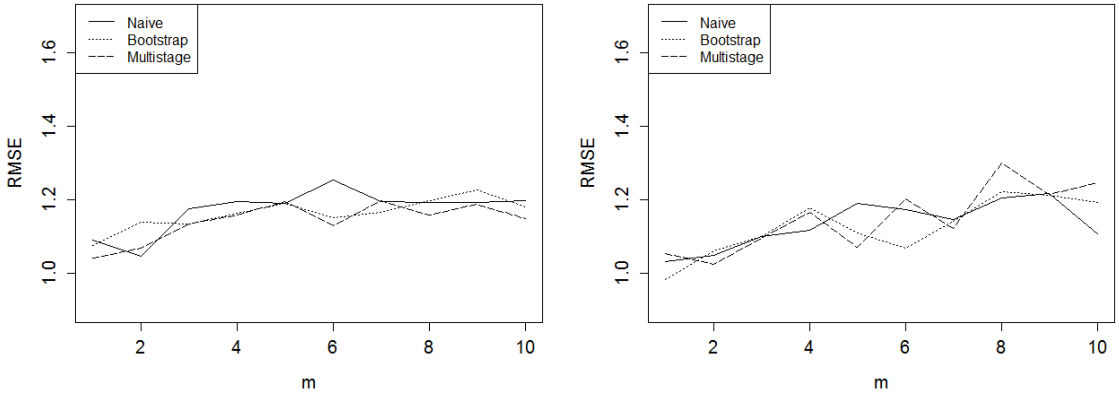
$$\text{var}\,[e_{n+M}] = M\sigma_e^2, \tag{3.4}$$

where $\sigma_e^2 = \text{var}\,[e_{n+1}]$ denotes the variance of the one-step-ahead forecast errors. Chatfield describes how this approximate formula is only valid for a random walk model but is frequently used for other models.

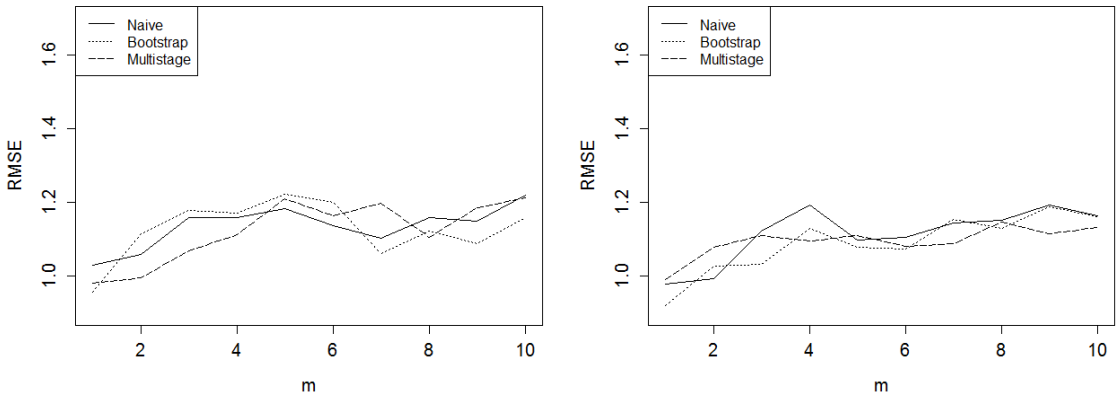
3.4 Simulation Results

In this section, we run simulations from the models presented in the examples in Chapter Two and compare the root mean squared errors (RMSE) for the three forecasting methods described in Section 3.2. We generate 500 realizations for each model with lengths $n = 75, 150, 250$, and 500. We also obtain empirical coverages of the 95% prediction intervals for the three forecasting methods. We obtain the prediction intervals by the z interval in (3.3), by the bootstrapped quantiles, and by the approximate formula in (3.4). We ran 500 iterations for each model, 400 replications for the bootstrap forecasting method, and found forecasts for $M = 1, \ldots, 10$.

Example 3.1. (Continuation of Example 2.1) For the EXPAR(2) model defined in Example 2.1, we show the RMSE for the four series lengths in Figure 3.1. For any of the series lengths, we don't see any method with significantly lower RMSE. For $n = 250$, we see the multistage method with lower RMSE for smaller values of M and the bootstrap method with the lower RMSE for larger values of M. However, these results do not hold for the other series lengths.

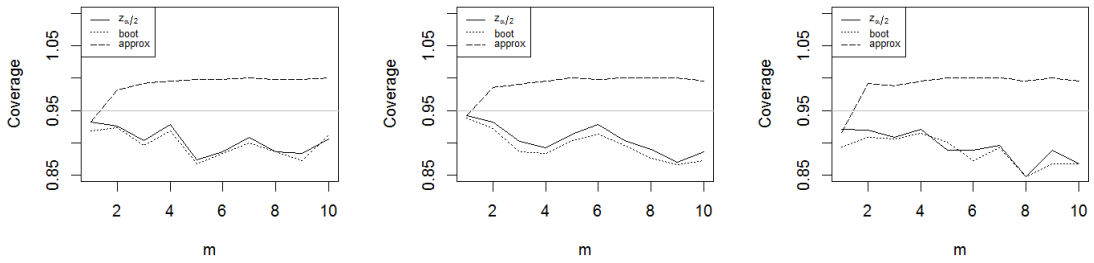


(a) (b)



(c) (d)

Figure 3.1: Plots of the RMSE's of the three forecasting methods for $M = 1, \ldots, 10$ for Example 3.1 with (a) $n = 75$, (b) $n = 150$, (c) $n = 250$, and (d) $n = 500$.



(a) (b) (c)

Figure 3.2: Plots of the empirical coverages of the 95% prediction intervals for the three approaches for $n = 150$ with predictions using (a) the naive approach, (b) the bootstrap approach, and (c) the multistage approach for Example 3.1.

The empirical coverages of the 95% prediction intervals for the three forecasting methods for $n = 150$ can be seen in Figure 3.2. As expected, we see that the approximate approach for prediction intervals provides invalid results. The plots show that the z-interval approach provides more accurate prediction intervals than the bootstrap approach. These results are more evident in the naive and the bootstrap forecasting methods (See Appendix D for additional plots).

Example 3.2. (Continuation of Example 2.2) For a higher order model, we used the same EXPAR(4) model from Example 2.2. We show the RMSE for the four series lengths in Figure 3.3. For this model, we see that the bootstrap forecasting method has slightly smaller RMSE for series lengths $n = 150, 250, 500$. For $n = 75$, the bootstrap method has the smallest RMSE for $M = 2, 3$, and the multistage method had the smallest for the larger steps.

Figure 3.4 shows the empirical coverages of the 95% prediction intervals for the three forecasting methods for $n = 150$. Similar to Example 3.1, the plots show that the z-interval approach provides more accurate prediction intervals than the bootstrap approach. For this example, these results are more evident in the bootstrap and the multistage forecasting methods. For the other series lengths, the coverages are approximately the same for the bootstrap and z approaches (See Appendix D for additional plots).

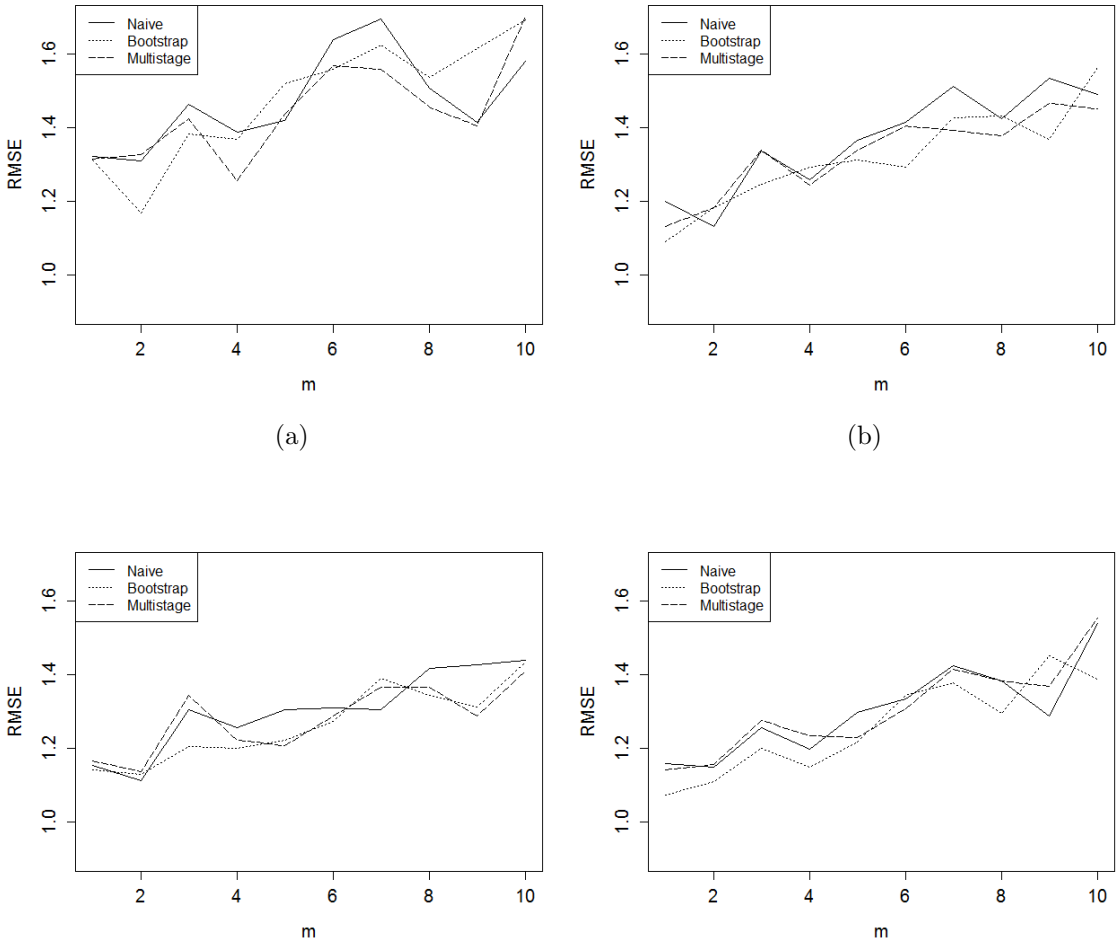


Figure 3.3: Plots of the RMSE's of the three forecasting methods for $M = 1, \ldots, 10$ for Example 3.2 with (a) $n = 75$, (b) $n = 150$, (c) $n = 250$, and (d) $n = 500$.

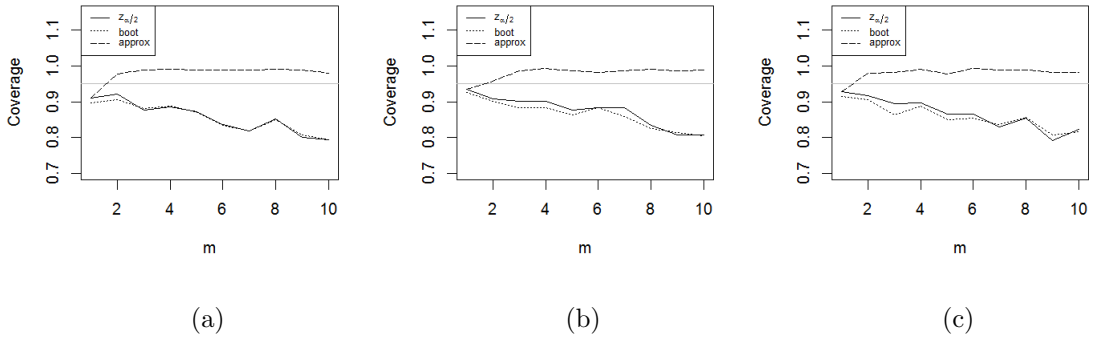


Figure 3.4: Plots of the empirical coverages of the 95% prediction intervals for the three approaches for $n = 150$ with predictions using (a) the naive forecast, (b) the bootstrap forecast, and (c) the multistage forecast for Example 3.2.

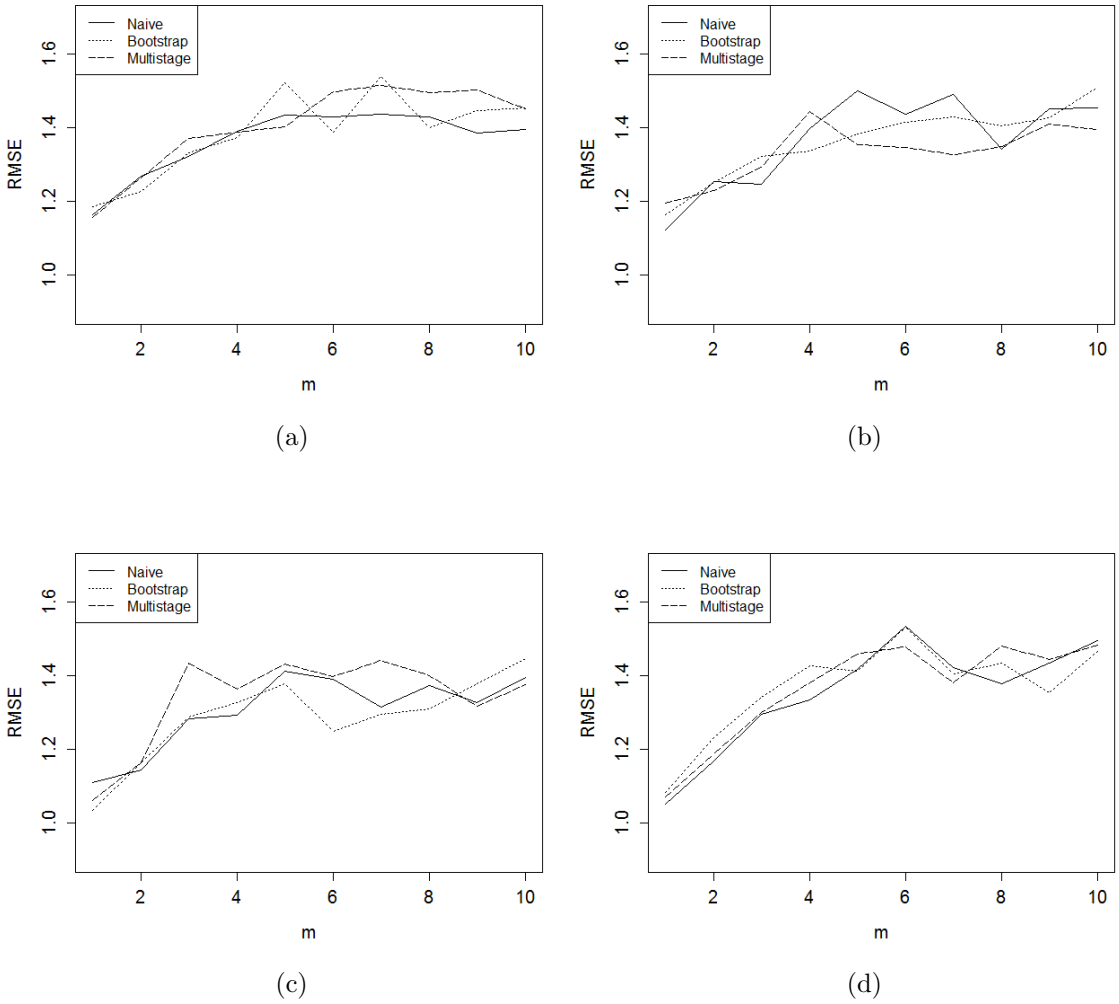
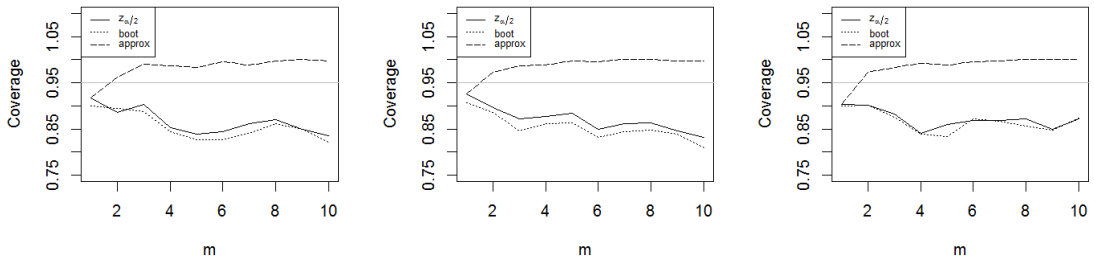


Figure 3.5: Plots of the RMSE's of the three forecasting methods for $M = 1, \ldots, 10$ for Example 3.3 with (a) $n = 75$, (b) $n = 150$, (c) $n = 250$, and (d) $n = 500$.

Example 3.3. (Continuation of Example 2.3) For the SETAR(2) model defined in Example 2.3, the RMSE for the four series lengths are shown in Figure 3.5. For this model, we do not see any of the three methods with significantly smaller RMSE. For $n = 150$, we see that the multistage method has slightly lower RMSE for $M = 5, \ldots, 10$. For $n = 250$, the bootstrap method has slightly lower RMSE for $M = 5, \ldots, 8$. For $n = 75$ and 500, we don't see any of the methods outperforming the other two consistently.

Figure 3.6 shows the empirical coverages of the 95% prediction intervals for the three forecasting methods for $n = 150$. For this model, it appears that the z-interval



(a) (b) (c)

Figure 3.6: Plots of the empirical coverages of the 95% prediction intervals for the three approaches for $n = 150$ with predictions using (a) the naive forecast, (b) the bootstrap forecast, and (c) the multistage forecast for Example 3.3.

approach has better coverage for all three forecasting methods. The same results can be seen in the plots for the other series lengths (See Appendix D for additional plots).

Example 3.4. (Continuation of Example 2.4) For the SETAR(4) model defined in Example 2.4, the RMSE for the four series lengths are shown in Figure 3.5. From these plots, we can see the multistage method having lower RMSE for smaller values of M for $n = 75$ and $n = 150$ and for larger values of M for $n = 500$. For $n = 250$, we see little difference in the three methods.

Plots of the empirical coverages of the 95% prediction intervals for the three forecasting methods for $n = 150$ can be seen in Figure 3.8. Overall, the empirical coverages of the prediction intervals are lower for this model than for the models in the previous examples. The z-interval approach still has higher coverage than the bootstrap approach. The plots for the other series lengths show the same results (See Appendix D for additional plots).

3.5 Conclusion

In this chapter, we have presented three forecasting methods to be used with the SBK estimator. Through simulation results, we showed that none of the methods were consistently better than the other methods in terms of RMSE. The bootstrap

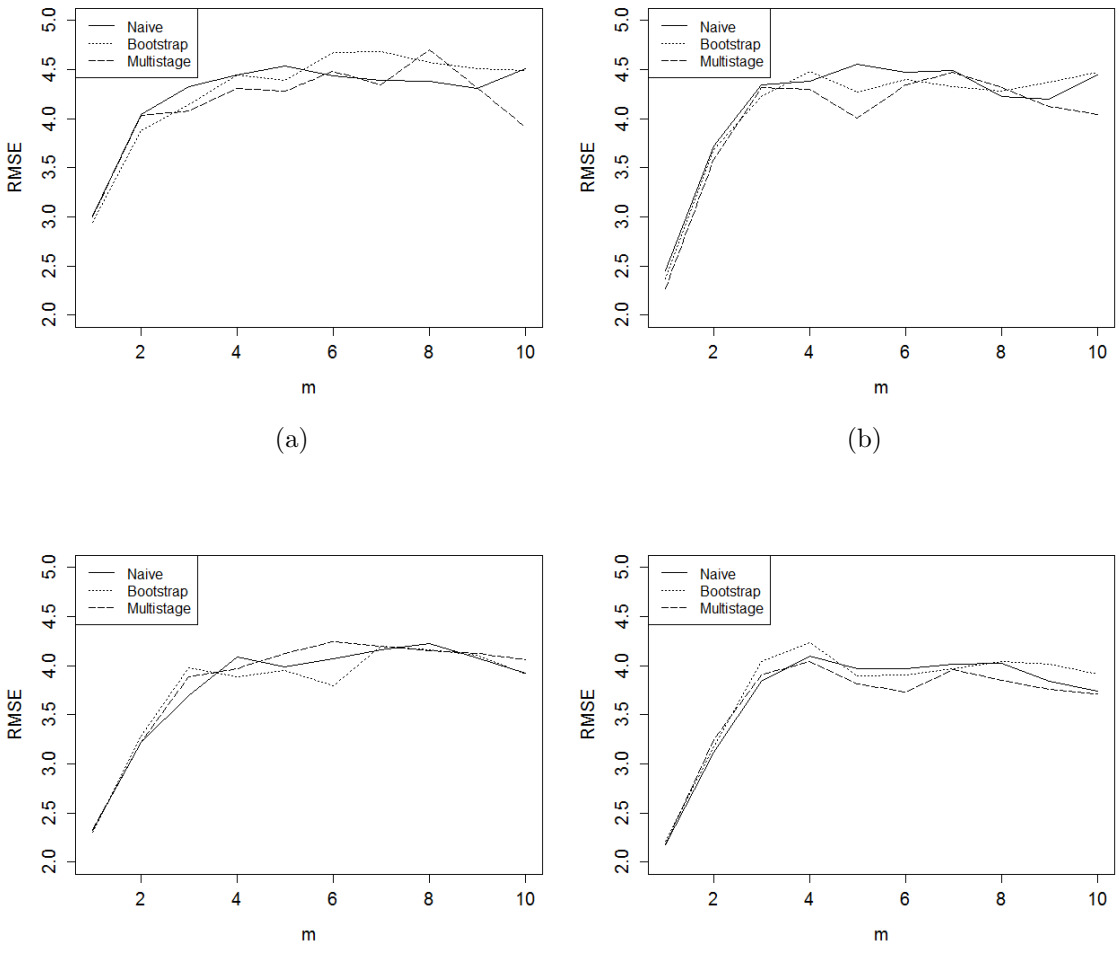


Figure 3.7: Plots of the RMSE's of the three forecasting methods for $M = 1, \ldots, 10$ for Example 3.4 with (a) $n = 75$, (b) $n = 150$, (c) $n = 250$, and (d) $n = 500$.

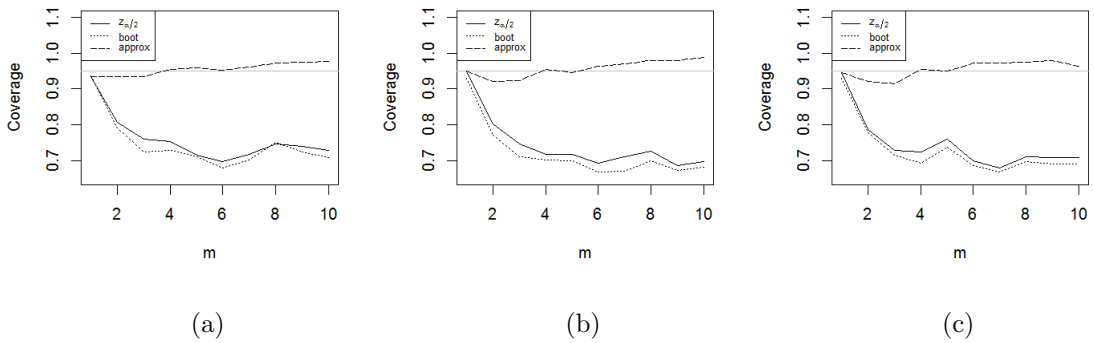


Figure 3.8: Plots of the empirical coverages of the 95% prediction intervals for the three approaches for $n = 150$ with predictions using (a) the naive forecast, (b) the bootstrap forecast, and (c) the multistage forecast for Example 3.4.

and multistage methods had smaller RMSE for some cases but the results did not show enough evidence to make one method the preferred choice.

We also explored three prediction interval approaches and found the approximate interval approach in (3.4) does not give accurate coverage. For the EXPAR(2) model, the bootstrap and z-interval approaches had empirical coverages around 0.95 for 95% prediction intervals for $M = 1$. The coverages began to decline to between 0.85 and 0.90 for larger values of M. For the EXPAR(4) model, the two approaches declined to between 0.80 and 0.90 for larger values of M. For SETAR models, the coverages declines to around 0.70. Overall, the z-interval approach had higher coverage than the bootstrap approach and is the preferred prediction interval for use with the SBK estimator.

CHAPTER FOUR

Application: Lanai Irradiance Data Set

4.1 Introduction

The Lanai irradiance data set contains irradiance measurements (in W/m^2) for 16 sensors recorded at one second intervals from January 1, 2010 to December 31, 2010 (Kuszamaul, Ellis, Stein, and Johnson, 2010). The 16 sensors are located within the 1.2 MW La Ola photovoltaic (PV) plant on the island of Lanai, Hawaii. The La Ola PV plant contains a grid of 12 solar panel tracking arrays arranged in three columns and four rows. Each tracking array is approximately 2,600 m^2 in size. Sandia National Laboratories and SunPower Corporation designed this system to study the effect of the movement of cloud shadows across the PV arrays on the power output of the plant. The power output is highly correlated with the measured irradiance so the goal is to model the irradiance measurements in this grid. In this chapter, we explore methods to achieve this goal through spatio-temporal models. We use the spline-backfitted kernel (SBK) method to explore the time series model.

The remainder of this chapter is organized as follows. In Section 4.2, we discuss how the diurnal time trend is removed from the data. We explore the data set through a preliminary analysis in Section 4.3. This preliminary analysis includes fitting the data with basic spatial models and time series models. In Section 4.4, we apply the SBK method to the data set to determine any patterns in the estimated coefficient functions. We explore spatio-temporal models in Section 4.5 and give a conclusion in Section 4.6.

4.2 Removing the Diurnal Trend

Before we analyze the data, we remove the diurnal trend. For irradiance data, several clear sky models can be found in the literature that can be helpful in removing this trend. These clear models are classified as very simple, simple, and complex. A review of some of these models can be found in Reno, Hansen, and

Stein (2012). Unfortunately, the clear sky models are for global horizontal irradiance (GHI) measurements. The GHI measurements are obtained by a sensor that is stationary. When the sensor moves and follows the sun across the sky, then the measurements are called plan-of-array (POA). The Lanai data set consists of POA measurements that were transformed to approximate GHI. This transformation is not exact and thus the clear sky models will not work for this data set. Instead, we use local polynomial kernel regression to model and remove the trend and then fit a time series model to the residuals.

For the observed irradiance Y_t at time t, and for kernel K with bandwidth h, the local polynomial kernel regression estimate is

$$\hat{\mathbf{Y}} = \left[\mathbf{X}(x)' \mathbf{K}(x) \mathbf{X}(x)\right]^{-1} \mathbf{X}(x)' \mathbf{K}(x) \mathbf{Y}$$

where $\mathbf{Y} = [y_1, \ldots, y_n]'$, $\mathbf{K}(x) = \text{diag}\{K_h(x_1 - x), \ldots, K_h(x_n - x)\}$, and

$$\mathbf{X}(x) = \begin{bmatrix} 1 & (x_1 - x) & (x_1 - x)^2 & (x_1 - x)^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (x_n - x) & (x_n - x)^2 & (x_n - x)^3 \end{bmatrix}.$$

Here, the predictor variable is the time index, $x_j = t_j$. This procedure is obtained using the `KernSmooth` package (Wand, 2012) in `R`. The local polynomial estimate for March 10 is found in Figure 4.1 along with a plot of the residuals. This method is able to remove the trend fairly well. However, we see from the plot of the residuals that the time series has non-constant variance. This heteroscedasticity will need to be addressed in a proper analysis of the data.

4.3 Preliminary Analysis

We first examine this data set through a preliminary analysis. This analysis includes some basic spatial and time series models fitted to the data. Using these basic models requires assumptions such as a separable covariance structure. That is, for a spatio-temporal field $Z(\mathbf{s}, t)$, where $\mathbf{s}$ represents space and t time, the covariance structure can be expressed as $\text{Cov}\{Z(\mathbf{s}, t), Z(\mathbf{s}', t')\} = C_1(\mathbf{s}, \mathbf{s}') C_2(t, t')$ for some spatial covariance C_1 and temporal covariance C_2. For the Lanai data, a separable covariance assumption may not hold. Intuitively, we expect the irradiance

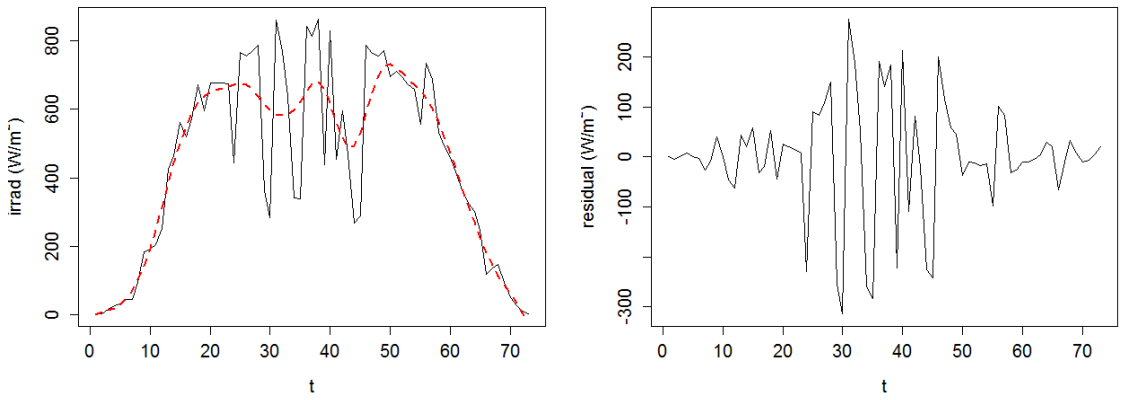


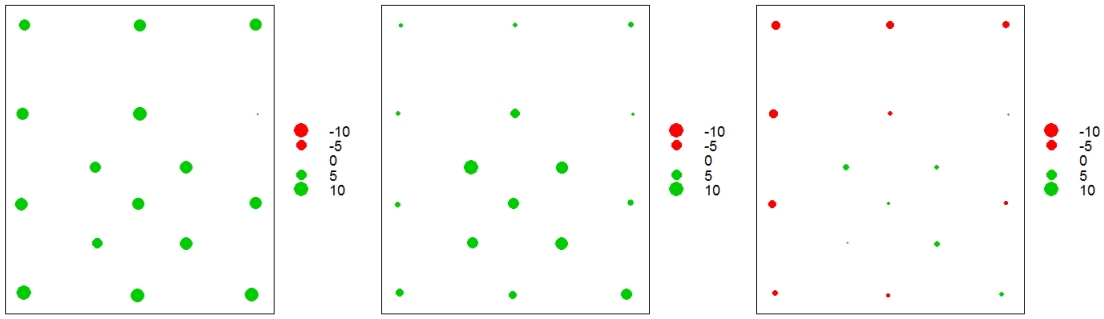
Figure 4.1: Plots of (a) irradiance measurements for March 10 and the local polynomial kernel estimate (dashed line) and (b) the residuals after removing the trend.

measurements to be fairly constant among all 16 sensors on clear and overcast days. However, on days that are partially cloudy we expect the clouds to move across the 16 sensors which will cause a spatial lag in the irradiance measurements. An example of this lag can be seen in Figure 4.2. At 8:00, the sensors are fairly uniform in the measurements except for one sensor on the right side which has a lower measurement. Ten minutes later, clouds move over the sensors at the top which cause measurements to fall. At 8:20, the measurements for the entire grid have fallen due to the clouds. Clearly, covariance separability will be an issue that will need to be addressed. For now, we assume a separable covariance structure.

4.3.1 Spatial Model

For the initial analysis of the Lanai data, for a fixed time t, the spatial structure is analyzed as lattice data. We fit the spatial structure using a conditional autoregressive (CAR) model. For a CAR model, we model the conditional mean and variance as

$$\begin{aligned} E\left[Y\left(\mathbf{s}_i\right)|\mathbf{Y}\left(\mathbf{s}\right)_{-i}\right] &= \mathbf{X}\left(\mathbf{s}_i\right)'\boldsymbol{\beta} + \sum_{j=1}^{n} c_{ij} e_j\left(\mathbf{s}_i\right) \\ \mathrm{Var}\left[Y\left(\mathbf{s}_i\right)|\mathbf{Y}\left(\mathbf{s}\right)_{-i}\right] &= \sigma_i^2, \end{aligned}$$



(a) (b) (c)

Figure 4.2: Spatial bubble plots for the 16 sensors on March 7, at (a) 8:00, (b) 8:10, and (c) 8:20.

where

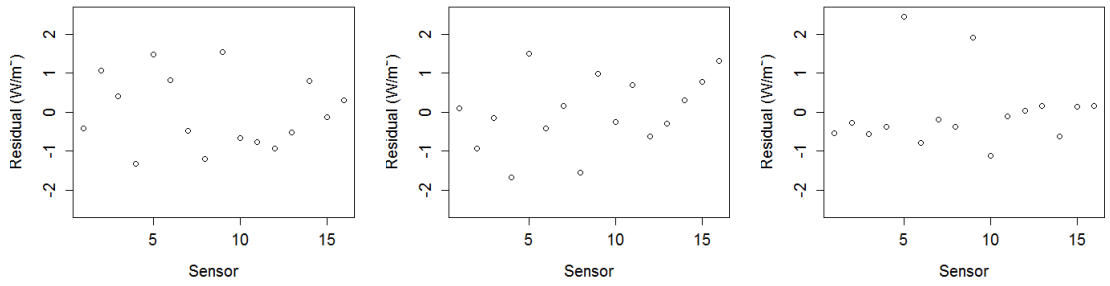
$$e_j(\mathbf{s}_i) = Y(\mathbf{s}_j) - \mathbf{X}(\mathbf{s}_i)\boldsymbol{\beta},$$

$\mathbf{Y}(\mathbf{s})_{-1}$ is the vector of observations with the ith observation removed, and c_{ij} are spatial dependence parameters with $c_{ii} = 0$. For the Lanai data set, there are no covariates so $\mathbf{X}(\mathbf{s}_i) = 1$ and a two nearest neighbors structure is used.

The CAR model is fitted to data on March 10 at three different times using the `spautolm` function in the `R` package `spdep` (Bivand et al., 2013). To test the hypothesis of no spatial autocorrelation, the likelihood ratio test gives p-values of 0.0098, 0.0175, and 0.0205 for times 8:00, 12:00, and 14:00 respectively. All three tests give a significant p-value indicating the presence of spatial autocorrelation. The studentized residuals of the fit can be seen in Figure 4.3. For 8:00 and 12:00, there are no obvious patterns in the studentized residuals to indicate non-normality. However, it is difficult to conclude non-normality with only 16 sensors. For 14:00, two outliers are evident which indicates that a more complex model may be needed for this data. The CAR model was fit to other time periods with similar results.

4.3.2 Time Series Model

As seen in Figure 4.1 (b), the residuals after the trend is removed indicate non-constant variance in the time series. One solution to this problem is to remove from the measurements a set number of time after sunrise and before sunset. This



(a) (b) (c)

Figure 4.3: Studentized residuals of irradiance data for March 10, at (a) 8:00, (b) 12:00, and (c) 14:00 for all 16 sensors after fit of a CAR model.

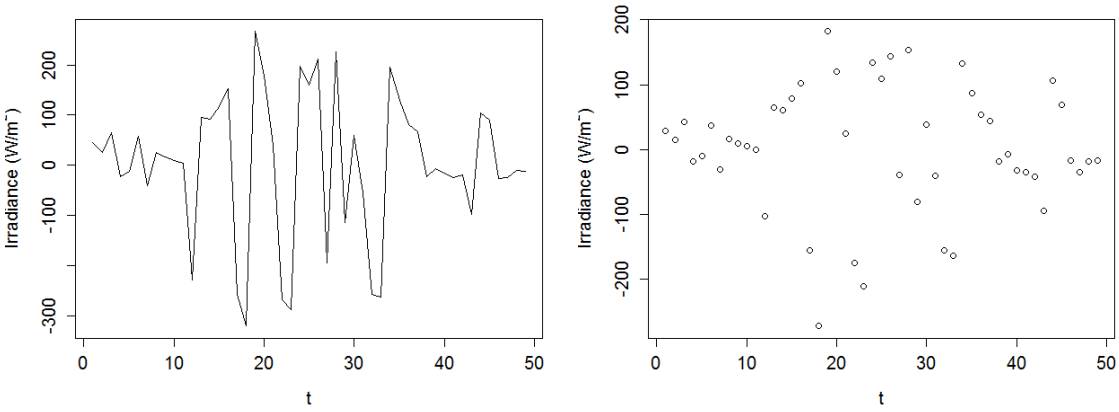
solution may work for some days, but not for days with more variable weather. One type of model that may work for this type of data is the self-exciting threshold autoregressive (SETAR) model. For a univariate time series y_t with a $k \times 1$ vector $\mathbf{Y}_{t-1} = (1, y_{t-1}, y_{t-2}, \cdots, y_{t-p})^T$, the SETAR model takes the form

$$Y_i = \phi_1^{(j)} X_{i-1} + \cdots + \phi_p^{(j)} X_{i-p} + \varepsilon_i \quad \text{if } X_{i-d} \in \Omega_j, \quad j = 1, \ldots, k,$$

where $\{\Omega_i\}$ form a non-overlapping partition of $\Re$. The SETAR model is flexible for modeling nonlinear time series and it will be the first model used in the preliminary analysis of the Lanai data.

First, we examined data on March 10 for sensor one but for observations only between 8:40 and 16:40. These times represent two hours after sunrise and before sunset, respectively. The time series plot of the data can be seen in Figure 4.4. The parameters of the SETAR model can be found using the `tsDyn` package (Antonio et al., 2009) in `R` and the model fit using the `setar` function. For this data, the number of regimes found was $m = 2$ and the delay parameter was $d = 2$. Just from examining the time series plot, we can see the problem of heteroscedasticity still exists which can also be seen in the plot of the residuals.

A second model that we can fit is the generalized autoregressive conditional heteroscedasticity (GARCH) model. A GARCH(q, p) model is useful when the variance is believed to be dependent on time. For these models, it is assumed that the error variance follows an autoregressive moving average (ARMA) model. This



(a) (b)

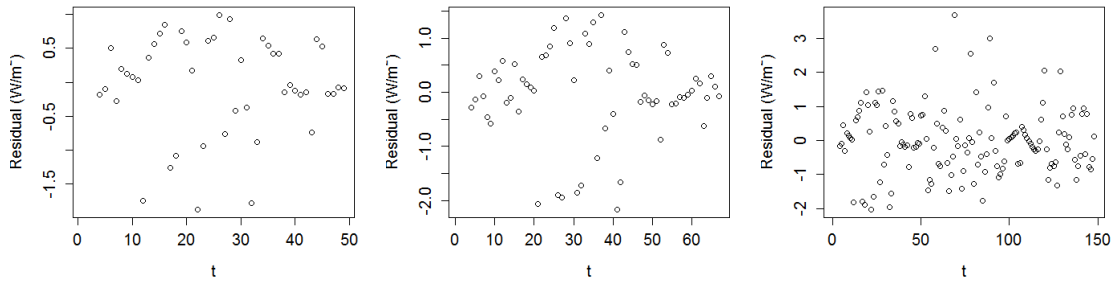
Figure 4.4: (a) Time series plot of March 10 form 8:40 to 4:40 and (b) the plot of the residuals after fit of SETAR(2,2) model.

model is expressed as

$$\sigma_t^2 \quad = \quad \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-1}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2,$$

where α_i is the autoregressive coefficient, ϵ_t is the error term, and β_i is the moving average coefficient.

We can fit the GARCH model in `R` using the `garch` function in the `tseries` package (Trapletti and Hornik, 2012). This function uses maximum likelihood to estimate the parameters. To determine the values of q and p, we run the function for a number of combinations of q and p and chose the values that correspond to the lowest root mean squared error (RMSE). Using the same data as in the SETAR model above, we find the lowest RMSE was with $q = 3$ and $p = 1$. The plot of the residuals can be found in Figure 4.5. The residuals appear to have non-constant variance with the residuals on the ends being less variant. We also fit the GARCH model to all the data on March 10 from sunrise to sunset. The residual plot for this data shows non-constant variance more clearly. To explore the data on multiple days, we fit the model to three days of data (March 10 - 12). The residual plot for this fit shows a definite pattern that needs to be addressed. It appears the GARCH model is a better fit than SETAR model but more improvement is needed for this data.



(a) (b) (c)

Figure 4.5: Plots of the residuals after fitting (a) a GARCH$(3, 1)$ model to March 10, 8:40 to 16:40, (b) a GARCH$(1, 4)$ model to March 10, sunrise to sunset, and (c) a GARCH$(1, 1)$ model to March 10 - 12.

4.3.3 Preliminary Conclusion

The assumption of separable covariance structure will need to be addressed for this data set. The use of CAR spatial models seems to work well for some time periods but not for all. For the time series aspect, the problem of heteroscedasticity will be a major obstacle for fitting and forecasting for each sensor. This problem may be compounded once the spatial covariance is also taken into account. The SETAR model was not able to fit the data well due to heteroscedasticity. The GARCH model did a better job at fitting the data but still had issues. We believe a better model to use for this data will be a functional-coefficient autoregressive model fitted using spline-backfitted kernel estimation. These models are flexible for modeling non-linear time series and the estimators perform well even with non-constant variance.

4.4 Applying the SBK Method

We introduced the SBK method for estimating FCAR models in Section 2.4. We fit a FCAR model to the Lanai data set using the SBK method and examine the estimated coefficient functions. Before we fit the model, we examine the data and see that the weather conditions play a major role in the behavior of the irradiance. For instance, January 30 was cloudy for most of the day while March 8 was clear. In Figure 4.6, we see the two days have much different irradiance measurements. We group days based on weather conditions, fit an FCAR model using the SBK

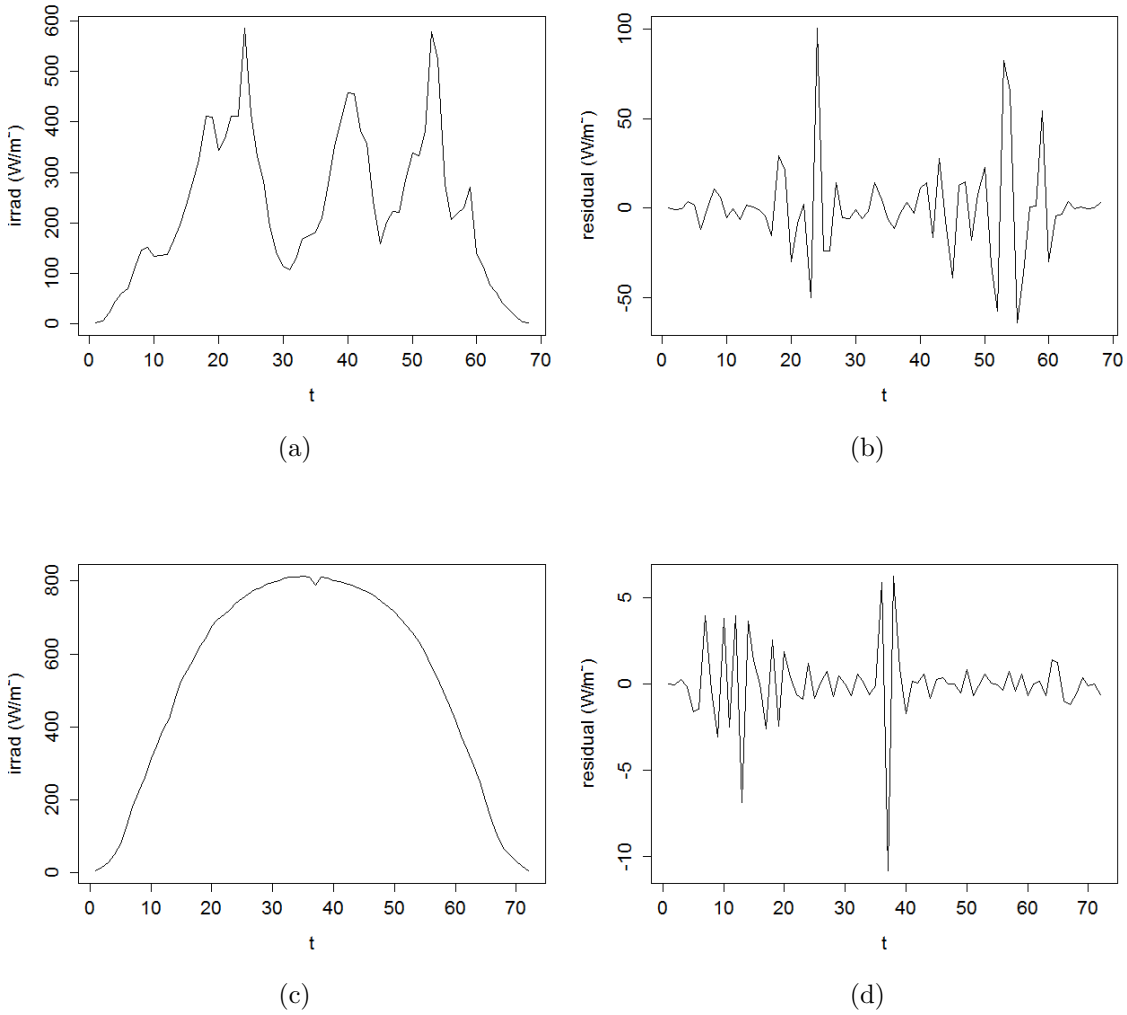


Figure 4.6: Irradiance measurements for January 30 (a) original data and (b) with trend removed. Irradiance measurements for March 8 (a) original data and (b) with trend removed.

method, and then examine the plots of the estimated coefficient functions. If these coefficient functions show a pattern among days with the same weather conditions, then we may be able to use a parametric model based on the shape of the estimated coefficient functions.

We begin by examining clear days. For each day, we fit a FCAR model with dimension p and delay d to the data using the SBK method. For this section, we will only be using the measurements from sensor one. Different values for p and d were fitted with the values resulting in the smallest root mean squared error (RMSE) being chosen. Of the eight clear days examined, five days had $p = 5$ and three days

had $p = 3$. The value for d ranged from two to five for these clear days. Although, the values of p and q differ from day to day, we compare the coefficient functions for $l = 1$ among six of the clear days. We see the estimated coefficient functions in Figure 4.7. From these plots, we see that the coefficient functions have a 'v' shape close to zero. It appears that the 'v' moves further to the right for days later in the year. We also examine the coefficient functions for $l = 2$ which can be seen in Figure 4.8. The same 'v' shape pattern can be seen in these estimated coefficient functions although it is not as evident for March 18 and December 16. The pattern we see in these coefficient functions may lead us to a FCAR model with a parametric form for the coefficient functions. For the 'v' shape seen in Figures 4.7 and 4.8, a threshold autoregressive model may be appropriate.

For non-clear days, the weather conditions may vary between partly cloudy, mostly cloudy, and overcast. We examine days which are classified as mostly cloudy, of which we select six days. Similar to the clear days, we fit a FCAR model to these days with a p and d value selected by RMSE. We see the estimated functions for $l = 1$ in Figure 4.9. From these plots, we do not see any patterns in the functions, thus it would be difficult to use a parametric form for the functions. We do not see any patterns in the estimated coefficient functions for $l = 2$ either (see Appendix D for additional plots).

We also examine days where the weather is classified as rainy, of which we select six days. We see the fit of the estimated coefficient functions $l = 1$ in Figure 4.10. Similar to the cloudy days, we see no discernable pattern in the coefficient functions. We also see no patterns in $l = 2$ (see Appendix D for additional plots).

The overall fit of the FCAR model also varies based on the weather conditions. We show the RMSE and the corresponding values of p and d for all three weather conditions in Table 4.1. We see that the RMSE for the clear days are much smaller than for the cloudy and rainy days. This difference is due to the variability of the irradiance measurements being much greater when rain and clouds are present. A model with a covariate for weather condition may be useful for this data set. One

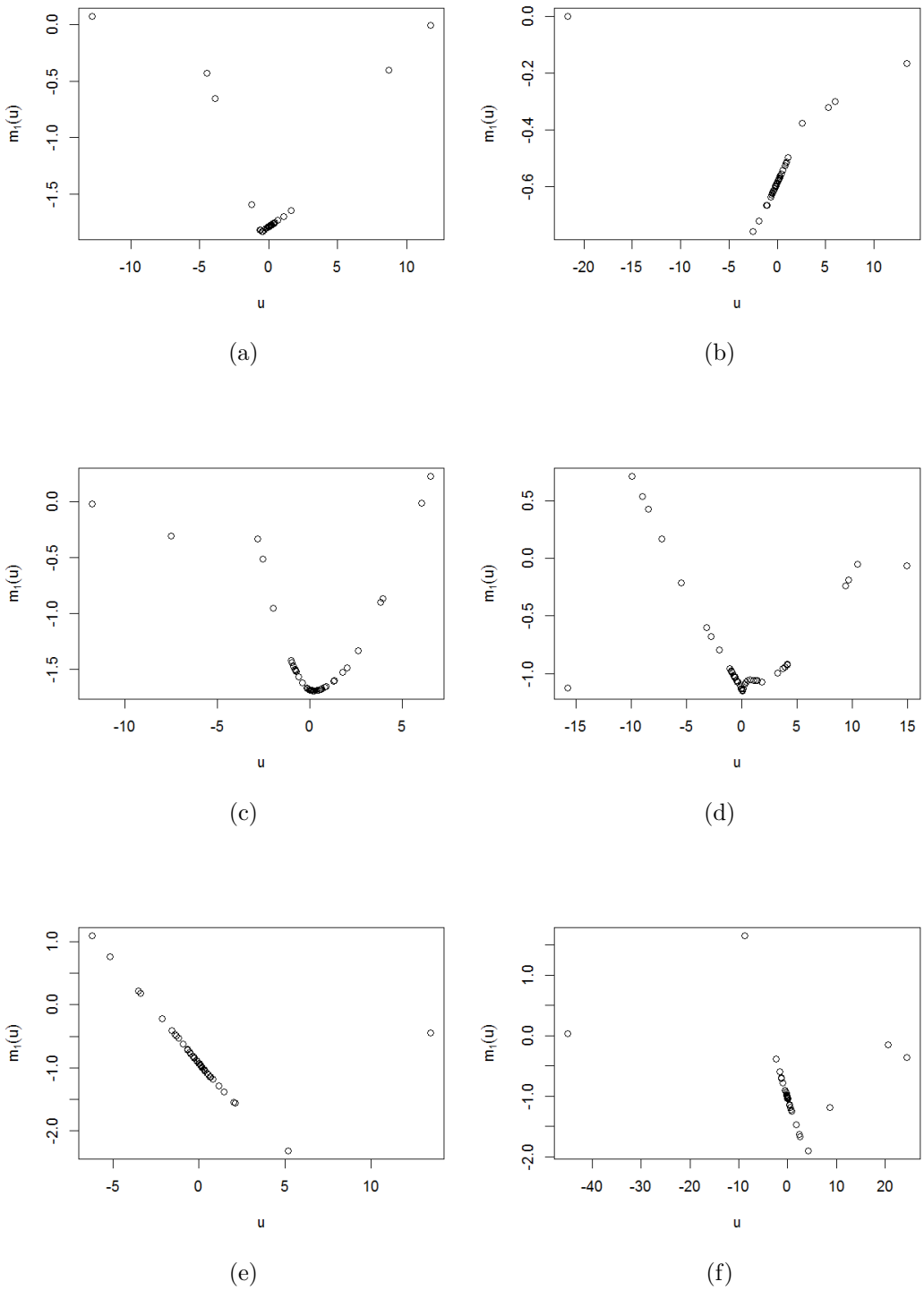
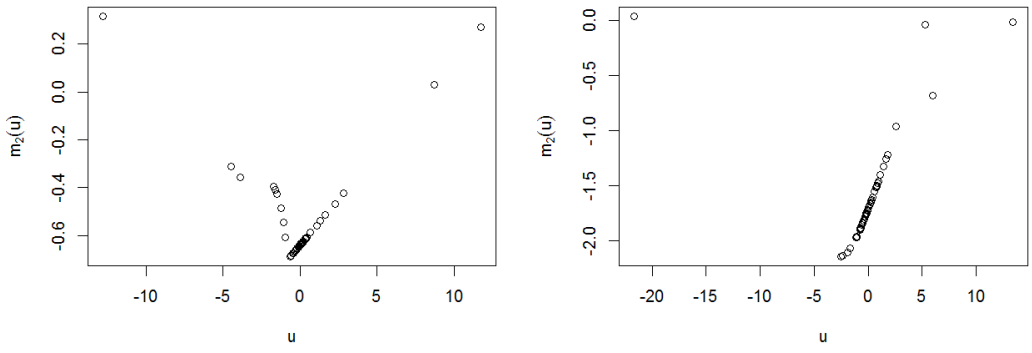
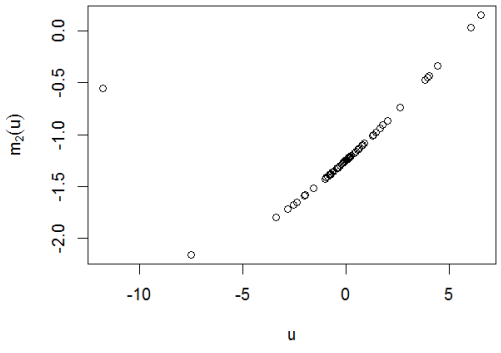


Figure 4.7: Estimated coefficient functions for $l = 1$ for clear days: (a) February 3, (b) February 16, (c) March 18, (d) March 19, (e) October 21, and (f) December 16.

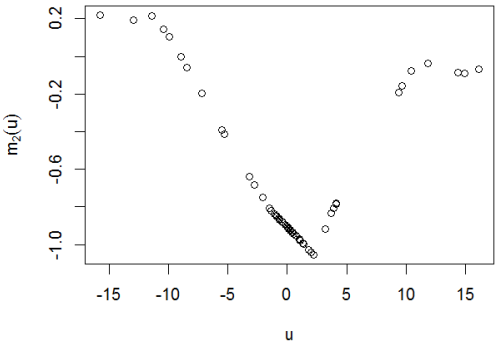


(a)

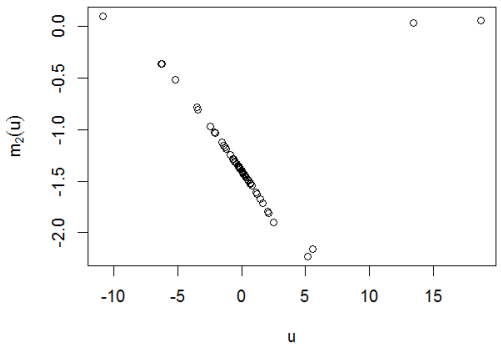
(b)



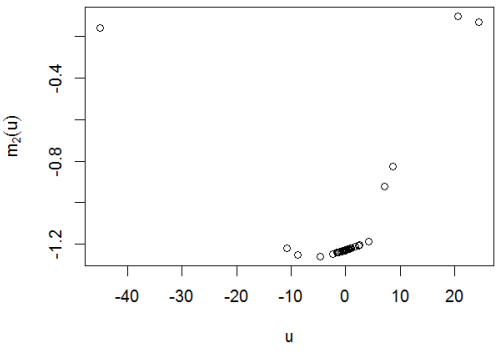
(c)



(d)

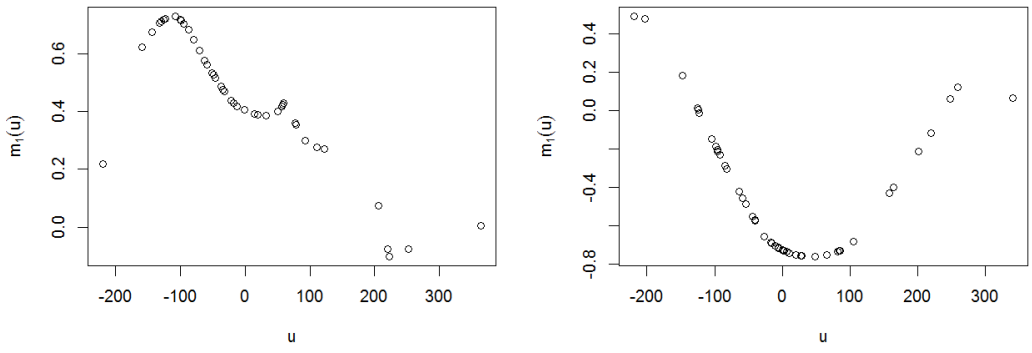


(e)



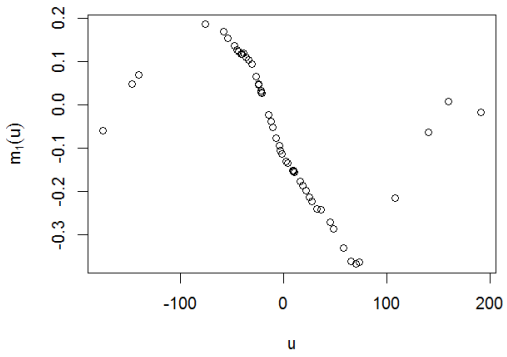
(f)

Figure 4.8: Estimated coefficient functions for $l = 2$ for clear days: (a) February 3, (b) February 16, (c) March 18, (d) March 19, (e) October 21, and (f) December 16.

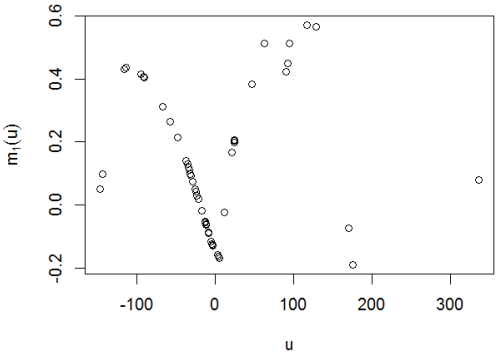


(a)

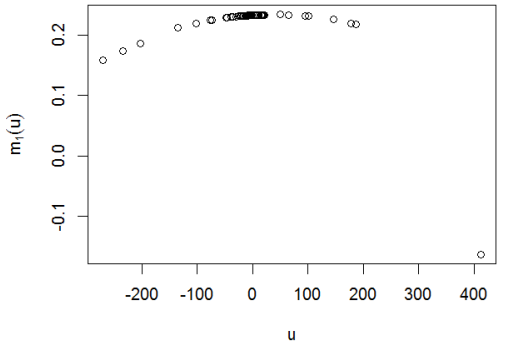
(b)



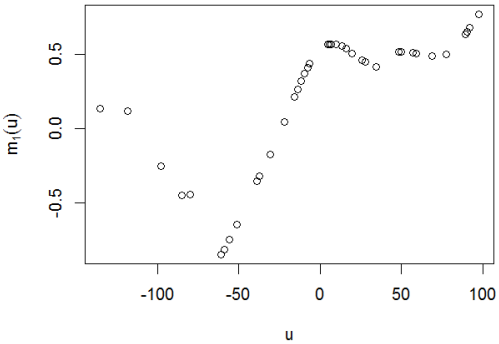
(c)



(d)

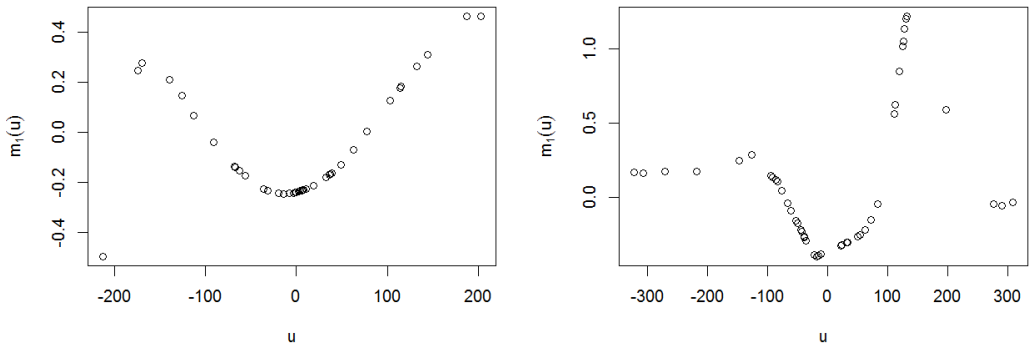


(e)



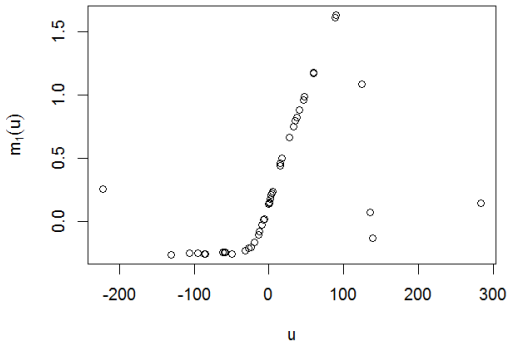
(f)

Figure 4.9: Estimated coefficient functions for $l = 1$ for cloudy days: (a) March 7, (b) April 1, (c) May 10, (d) June 4, (e) June 28, and (f) November 15.

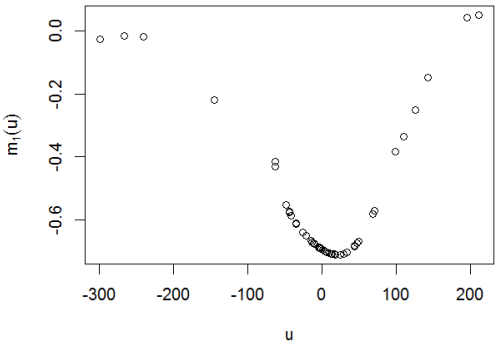


(a)

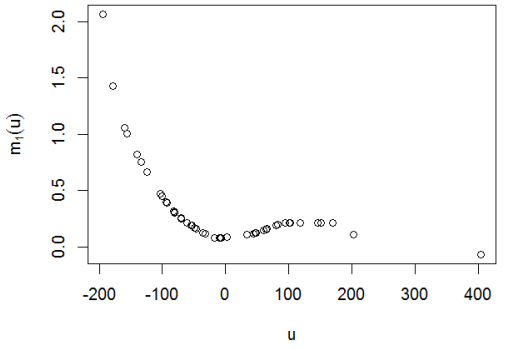
(b)



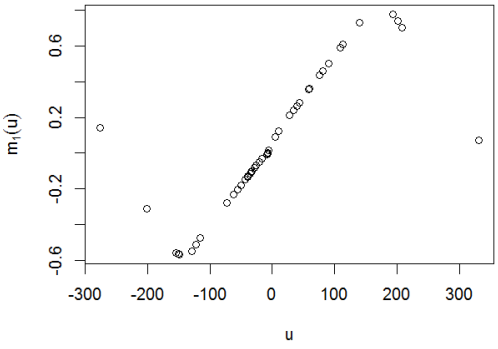
(c)



(d)



(e)



(f)

Figure 4.10: Estimated coefficient functions for $l = 1$ for rainy days: (a) February 1, (b) March 15, (c)April 6, (d) May 31, (e) August 3, and (f) October 27.

Table 4.1: The RMSE of the days with clear, cloudy, and rainy conditions and the corresponding values of p and d.

Condition	Date	p	d	RMSE
clear	Feb. 3	3	3	1.1763
	Feb. 16	5	2	0.6486
	Mar. 18	3	4	1.0094
	Mar. 19	5	3	1.3852
	Oct. 21	3	4	1.3214
	Dec. 16	5	2	2.7418
cloudy	Mar. 7	5	2	73.9713
	Apr. 1	4	5	80.7983
	May 10	5	1	46.4456
	Jun. 4	5	2	51.7008
	Jun. 28	4	2	57.9384
	Nov. 15	5	3	33.6671
rain	Feb. 1	5	5	62.9553
	Mar. 15	4	2	106.4262
	Apr. 6	3	5	55.5763
	May 31	5	3	54.6207
	Aug. 3	5	5	87.2082
	Oct. 27	3	3	74.5524

drawback to using a covariate for weather condition is the reliability of weather forecasts one day in advance and beyond. When we forecast the irradiance measurements, we will need the forecasted weather conditions which may be unavailable or inaccurate.

4.5 Spatio-Temporal Models

We now attempt to model the data in space and time through a spatio-temporal model. We do this in two different ways. One way is to first fit a spatial CAR model to the 16 sensors for each time t. This fit will give us a residual for each t at each sensor. We then fit the times series model to the residuals for each sensor. The other way to fit a spatio-temporal model is to first fit the time series for each sensor and then fit the CAR model to each time t of the residuals of the time series fit. For the first method, we model the spatial correlation first and then model the time series correlation. For the second method, we model the time series correlation and then the spatial correlation. We can compare the overall fit of

the spatio-temporal model by the total root mean squared error (TRMSE). For the RMSE_i of the residuals at sensor i, the TRMSE is defined as

$$\text{TRMSE} = \frac{1}{16}\sum_{i=1}^{16}\text{RMSE}_i.$$

The TRMSE's for the 18 days analyzed in Section 4.4 are found in Table 4.2. We see that for each day, the TRMSE of the model fit to space then to time is lower than with the model fit to time then to space. This difference in TRMSE is another argument for the need of a spatio-temporal model with a nonseparable covariance structure.

4.6 Conclusion

For the Lanai data set, we have shown that a CAR model performs fairly well in fitting the spatial structure for some time periods but not for all. For modeling the time structure, we have seen that the SETAR model was inadequate for modeling the heteroscedasticity of the data. The GARCH model performed better than the SETAR model although it did not completely model the heteroscedasticity. After fitting a FCAR model to the data by the SBK method, we see that a parametric form of the coefficient functions may be used for modeling clear days. For cloudy and rainy days, we do not see patterns in the coefficient functions. We believe a data driven approach to estimating the functions, such as the SBK method, is the best estimation method to use. We have also seen evidence for a nonseparable covariance structure for a spatio-temporal model through spatial plots and TRMSE. For the Lanai data set, future research will be conducted to model the spatial and temporal structures using a nonseparable covariance structure. The SBK method will be adapted to this spatio-temporal model for modeling the time structure. Future goals of this project is to forecast irradiance measurements both in time and space. The research conducted in this dissertation will be usefully in achieving these goals.

Table 4.2: The TRMSE for the space-time model and the time-space model of the days with clear, cloudy, and rainy conditions.

Condition	Date	Space-Time	Time-Space
clear	Feb. 3	0.28	0.67
	Feb. 16	2.97	3.30
	Mar. 18	0.30	0.60
	Mar. 19	0.14	3.57
	Oct. 21	0.63	0.92
	Dec. 16	0.80	1.84
cloudy	Mar. 7	11.09	36.33
	Apr. 1	8.43	38.38
	May 10	4.99	19.77
	Jun. 4	14.30	34.46
	Jun. 28	5.30	22.34
	Nov. 15	8.37	28.55
rain	Feb. 1	9.65	32.60
	Mar. 15	12.31	46.94
	Apr. 6	11.24	22.01
	May 31	10.14	25.98
	Aug. 3	7.78	42.65
	Oct. 27	4.82	34.41

APPENDICES

APPENDIX A

Assumptions for Theorems in Chapter Two

The necessary assumptions for Theorems 2.1, 2.2, and 2.3 which are taken from the Appendix of Liu and Yang (2010):

(i) The delay variable $\mathbf{U}$ has a continuous probability density function $f(u)$ that satisfies $0 < c_f \leq \min_{u\in\mathcal{X}} f(u) \leq \max_{u\in\mathcal{X}} f(u) \leq C_f < \infty$ for some constants c_f and C_f and $f(u) = 0$, $u \notin \mathcal{U} = [0,1]$.

(ii) There exist constants $0 < c_{\mathbf{Q}} \leq C_{\mathbf{Q}} < \infty$ and $0 < c_\delta \leq C_\delta < \infty$ and some $\delta > 1/2$ such that $c_{\mathbf{Q}}\mathbf{I}_{p\times p} \leq \mathbf{Q}(\mathbf{u}) = \{q(\mathbf{u})\}_{l,l'}^{p} = E(\mathbf{X}\mathbf{X}'|\mathbf{U}=\mathbf{u}) \leq C_{\mathbf{Q}}\mathbf{I}_{p\times p}$ and $c_\delta \leq E\left\{(X_l X_{l'})^{2+\delta}|\mathbf{U}=\mathbf{u}\right\} \leq C_\delta$ for all $\mathbf{u}\in\mathcal{U}$ and $l,l' = 1,\ldots,p$.

(iii) The vector process $\{\boldsymbol{\varsigma}_t\}_{t=-\infty}^{\infty} = \{(U_t, \mathbf{X}_t)\}_{t=-\infty}^{\infty}$ is strictly stationary and geometrically strongly mixing.

(iv) The coefficient functions, $m_l \in C^1[0,1]$, $m_l' \in \mathrm{Lip}([0,1], C_\infty)$, $\forall 1 \leq l \leq p$.

(v) The conditional variance function $\sigma^2(\mathbf{u},\mathbf{x})$ is measurable and bounded. The errors $\{\varepsilon_i\}_{i=1}^{n}$ satisfy $E(\varepsilon_i|\mathcal{F}_i) = 0$, $E(\varepsilon_i^2|\mathcal{F}_i) = 1$, $E\left(|\varepsilon_i|^{2+\eta}|\mathcal{F}_i\right) \leq C_\eta$ for some $\eta \in (1/2, 1]$ and the sequence of σ-fields $F_i = \sigma\{(U_j, \mathbf{X}_j), j \leq i; \varepsilon_j, j \leq i-1\}$ for $i = 1,\ldots,n$.

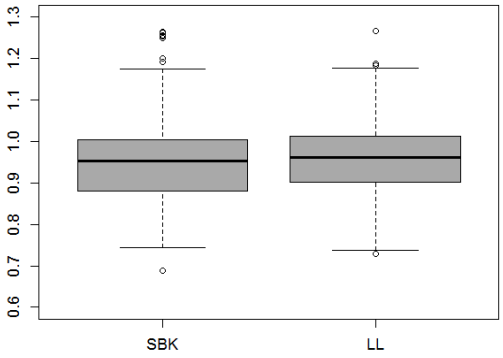
(vi) The function K is a symmetric probability density function supported on $[-1,1]$, and $K \in \mathrm{Lip}([-1,1], C_K)$ for some $C_K > 0$, while the bandwidth $h > 0$, $h \sim n^{-1/5}$.

(vii) The number of interior knots $N = N_n \sim n^{1/4}\log n$ and hence $H = (N+1)^{-1} \sim n^{-1/4}(\log n)^{-1}$.

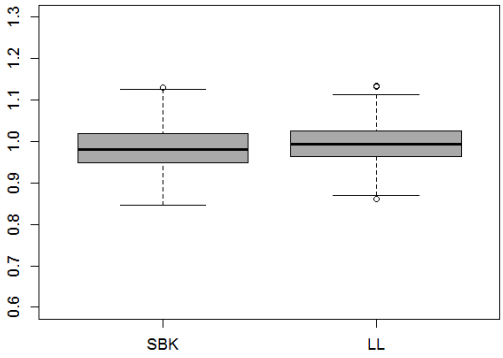
APPENDIX B

Chapter Two Simulation Results

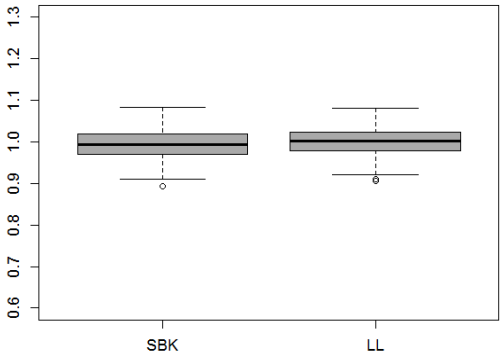
The following plots are additional boxplots referenced in Section 2.5.



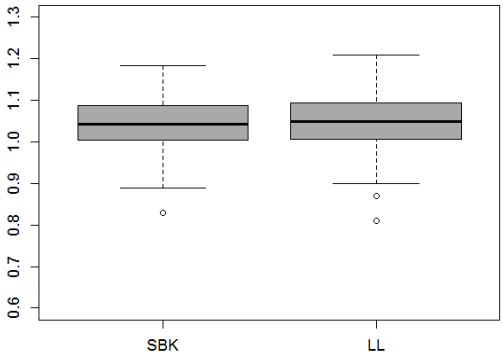
Example 2.1: $n = 75$



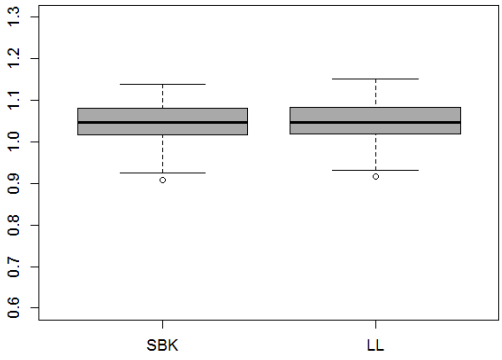
Example 2.1: $n = 250$



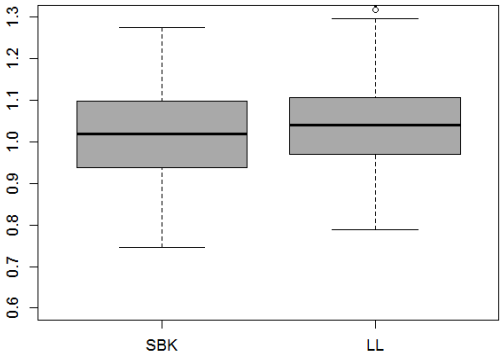
Example 2.1: $n = 500$



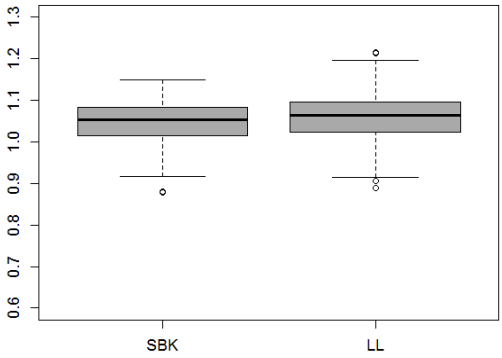
Example 2.2: $n = 250$



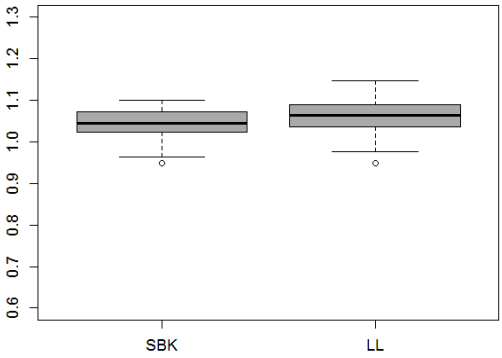
Example 2.2: $n = 500$



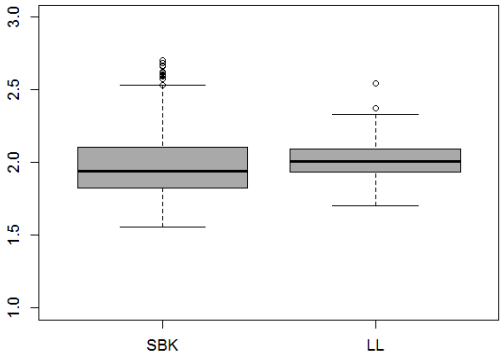
Example 2.3: $n = 75$



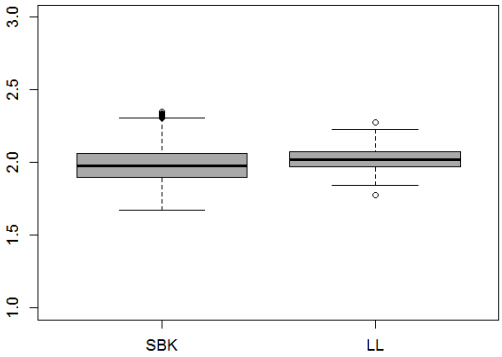
Example 2.3: $n = 250$



Example 2.3: $n = 500$



Example 2.4: $n = 250$

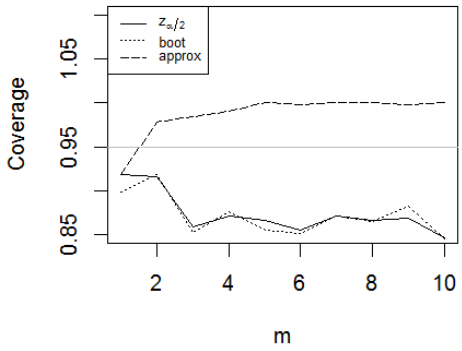


Example 2.4: $n = 500$

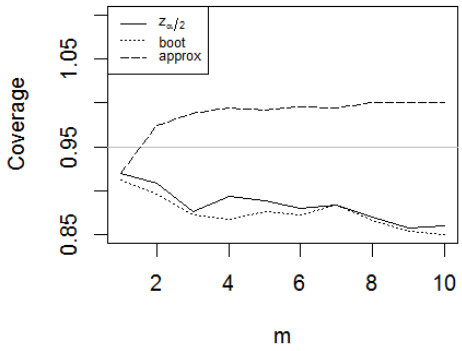
APPENDIX C

Chapter Three Simulation Results

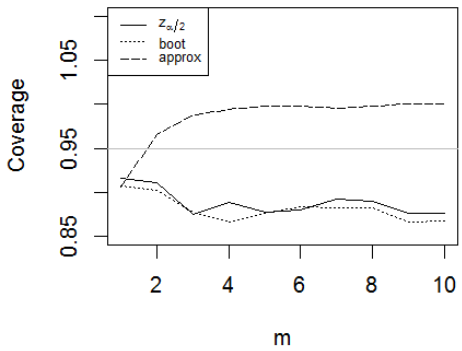
The following plots are additional results referenced in Section 3.4.



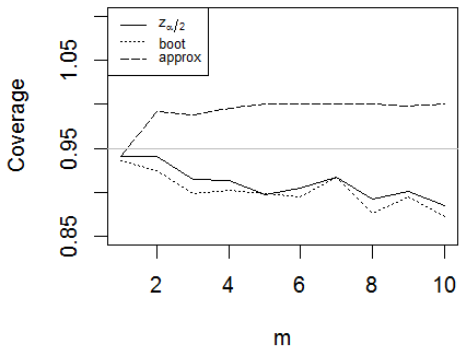
Example 3.1: $n = 75$, naive method



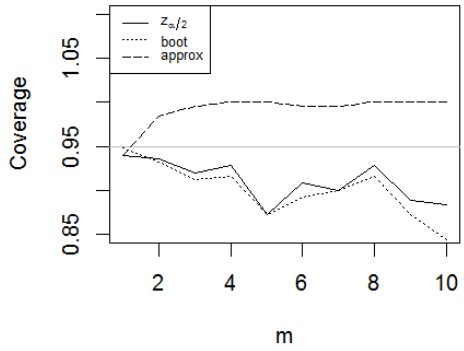
Example 3.1: $n = 75$, bootstrap method



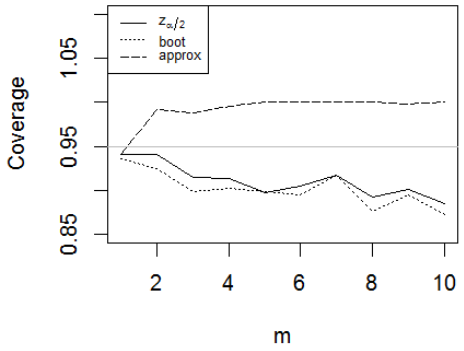
Example 3.1: $n = 75$, multistage method



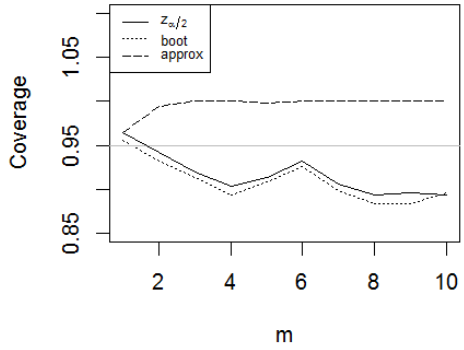
Example 3.1: $n = 250$, naive method



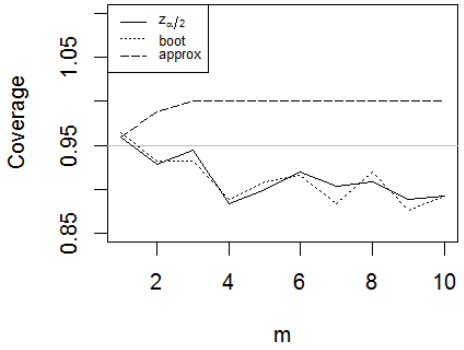
Example 3.1: $n = 250$, bootstrap method



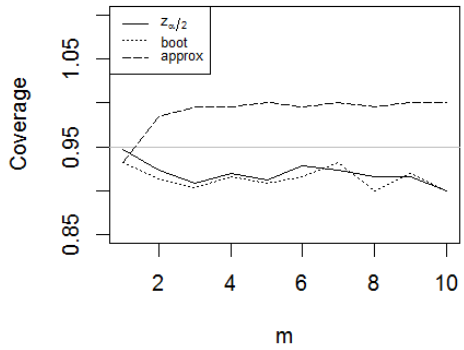
Example 3.1: $n = 250$, multistage method



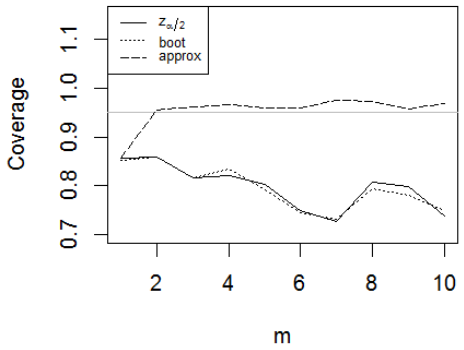
Example 3.1: $n = 500$, naive method



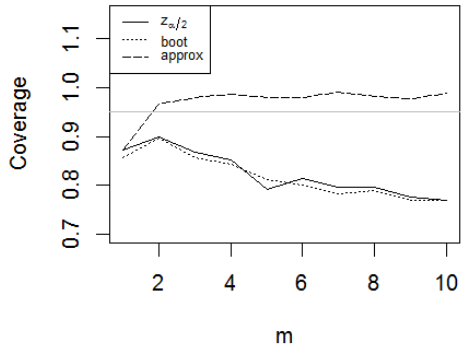
Example 3.1: $n = 500$, bootstrap method



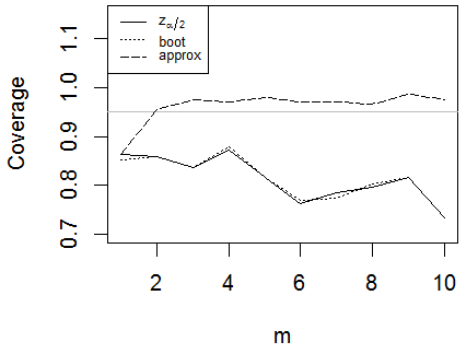
Example 3.1: $n = 500$, multistage method



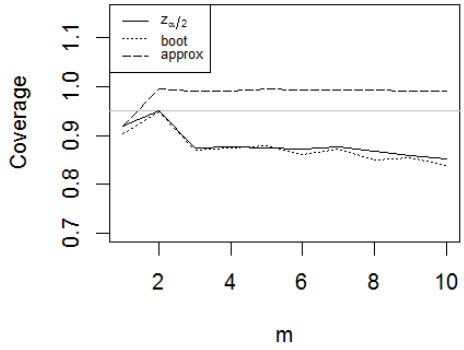
Example 3.2: $n = 75$, naive method



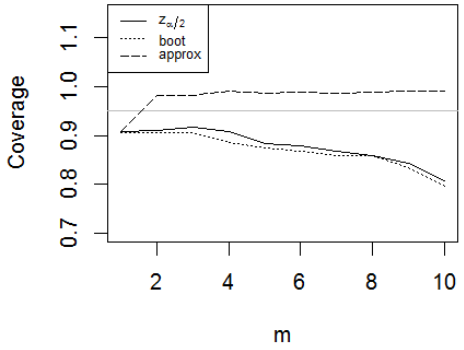
Example 3.2: $n = 75$, bootstrap method



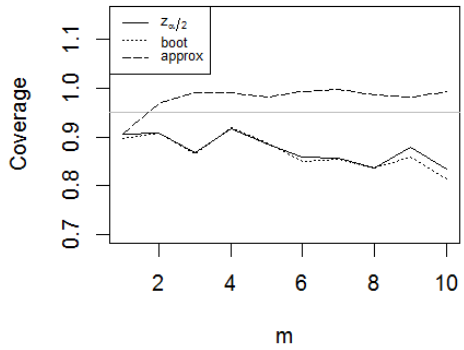
Example 3.2: $n = 75$, multistage method



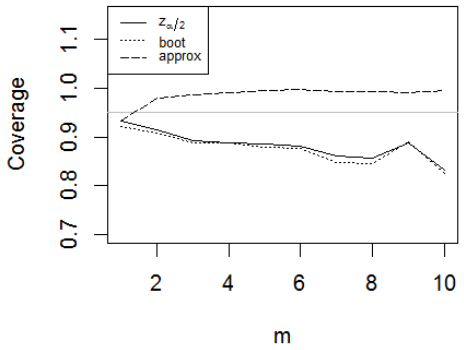
Example 3.2: $n = 250$, naive method



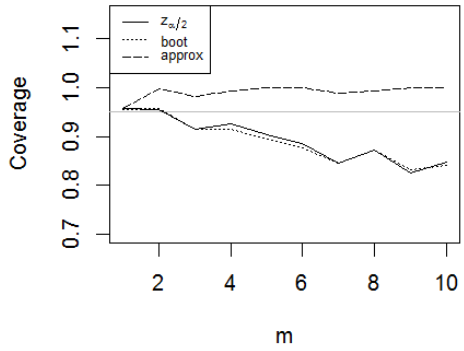
Example 3.2: $n = 250$, bootstrap method



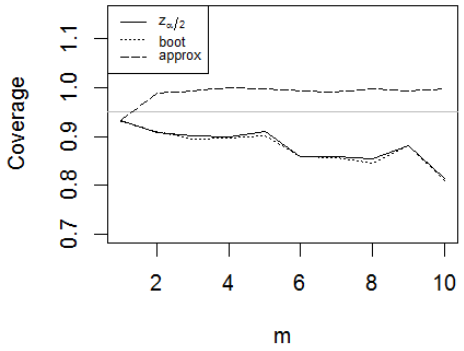
Example 3.2: $n = 250$, multistage method



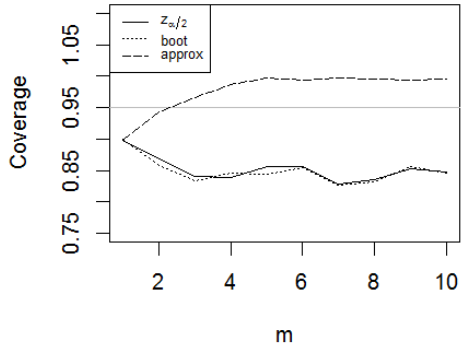
Example 3.2: $n = 500$, naive method



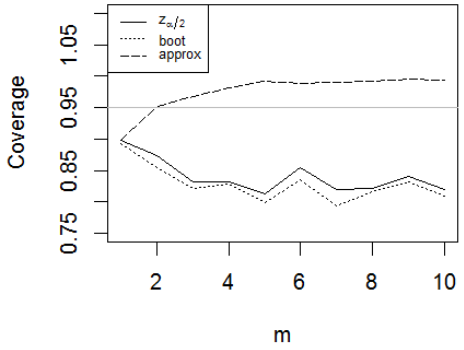
Example 3.2: $n = 500$, bootstrap method



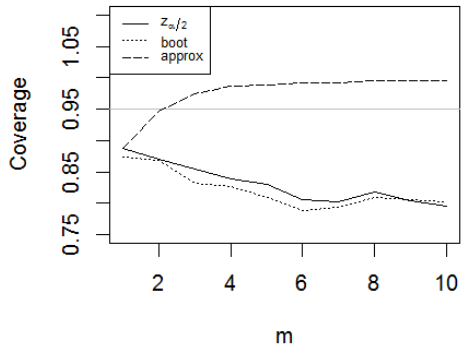
Example 3.2: $n = 500$, multistage method



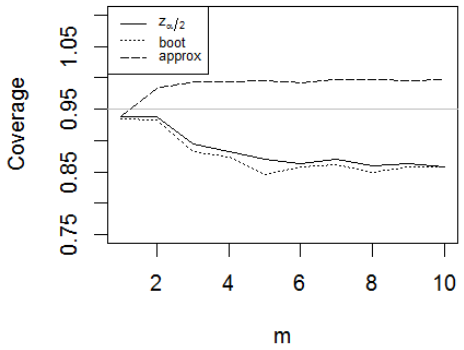
Example 3.3: $n = 75$, naive method



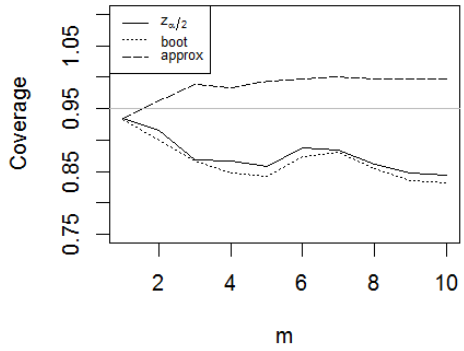
Example 3.3: $n = 75$, bootstrap method



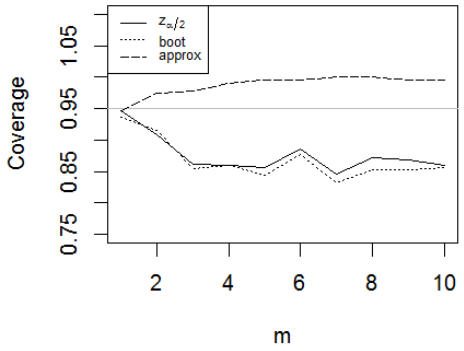
Example 3.3: $n = 75$, multistage method



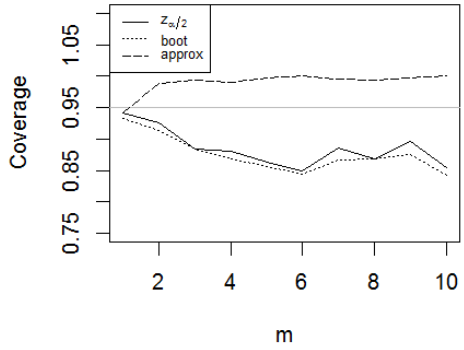
Example 3.3: $n = 250$, naive method



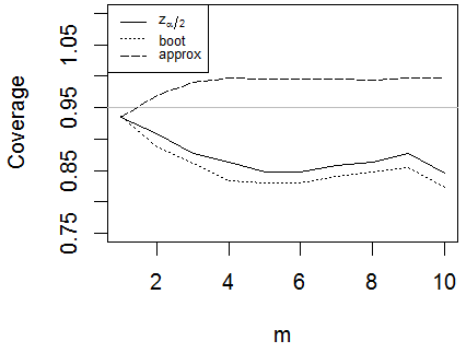
Example 3.3: $n = 250$, bootstrap method



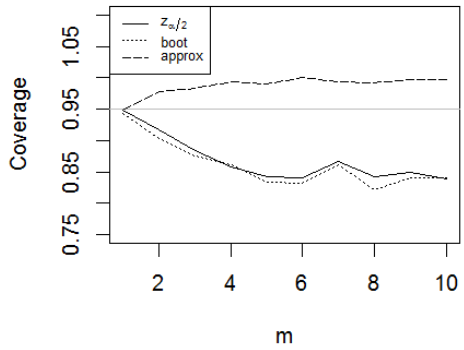
Example 3.3: $n = 250$, multistage method



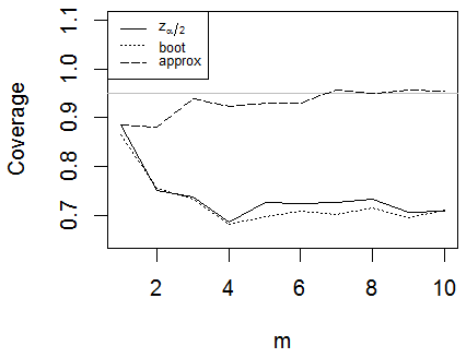
Example 3.3: $n = 500$, naive method



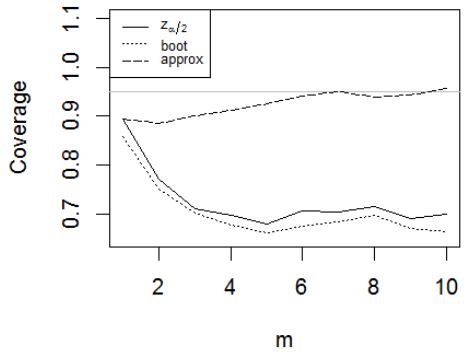
Example 3.3: $n = 500$, bootstrap method



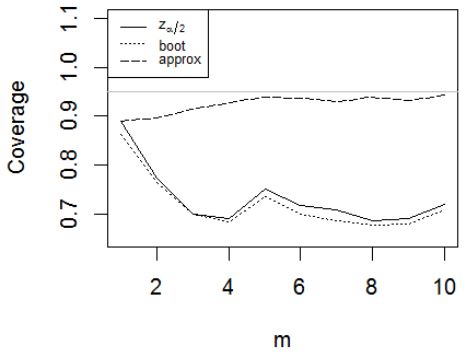
Example 3.3: $n = 500$, multistage method



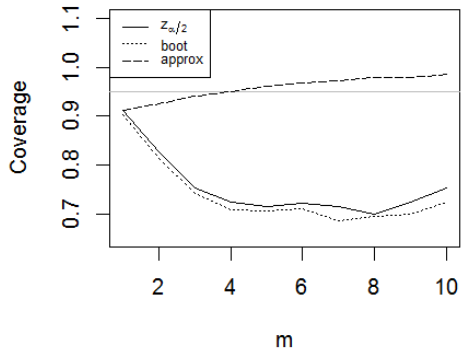
Example 3.4: $n = 75$, naive method



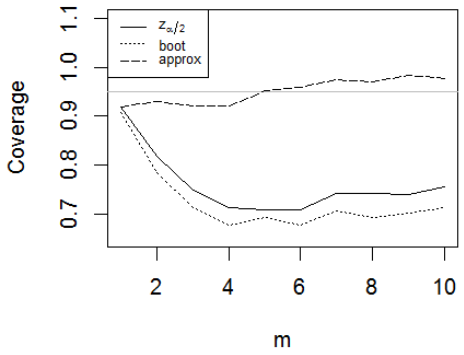
Example 3.4: $n = 75$, bootstrap method



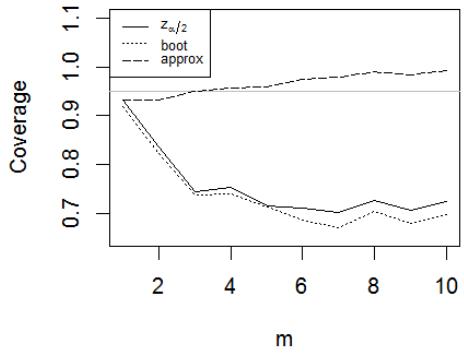
Example 3.4: $n = 75$, multistage method



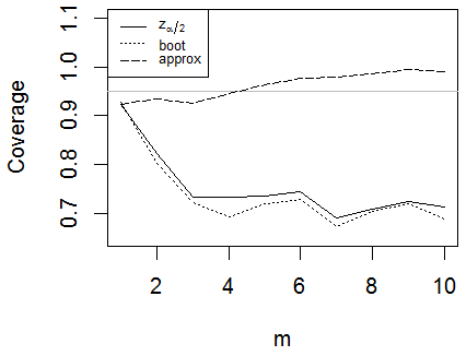
Example 3.4: $n = 250$, naive method



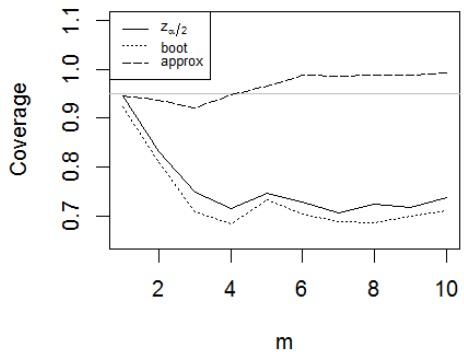
Example 3.4: $n = 250$, bootstrap method



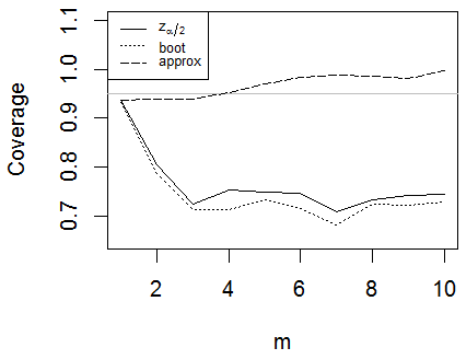
Example 3.4: $n = 250$, multistage method



Example 3.4: $n = 500$, naive method



Example 3.4: $n = 500$, bootstrap method

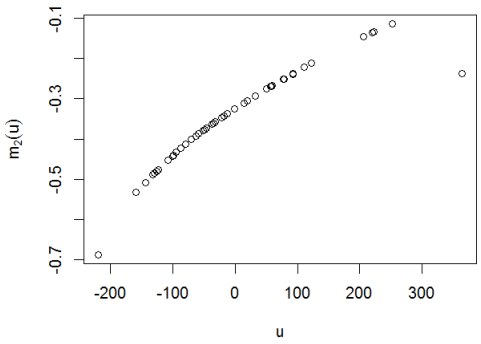


Example 3.4: $n = 500$, multistage method

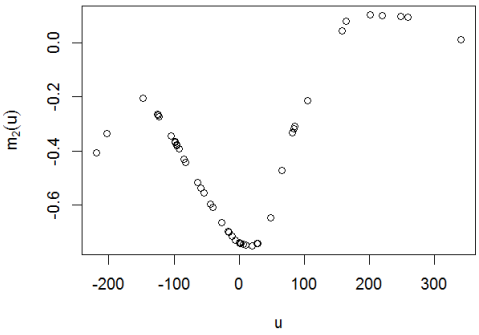
APPENDIX D

Chapter Four Additional Figures

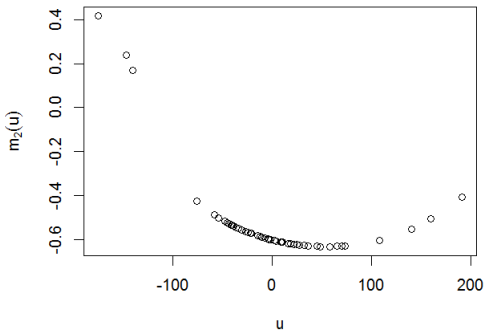
The following figures are additional plots referenced in Section 4.4.



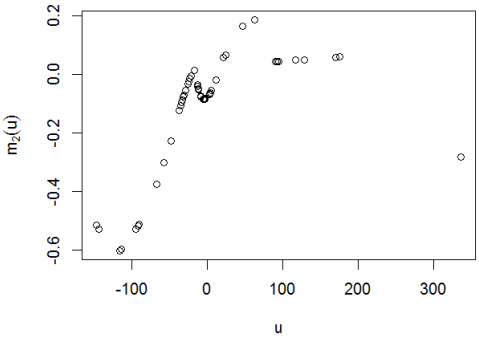
March 7: $l = 2$



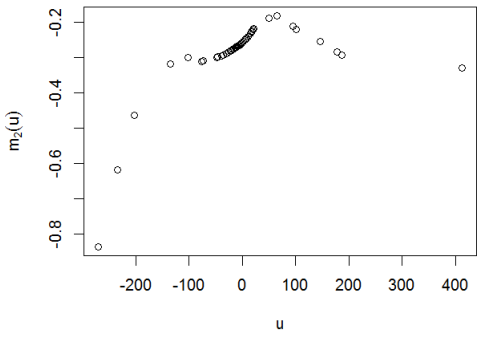
April 1: $l = 2$



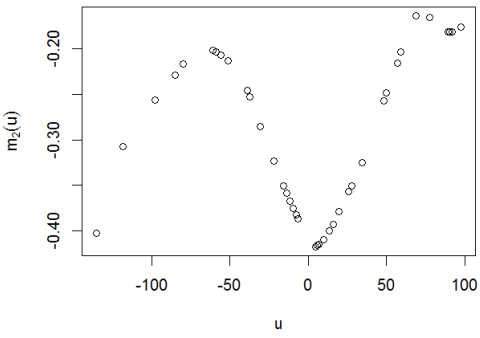
May 10: $l = 2$



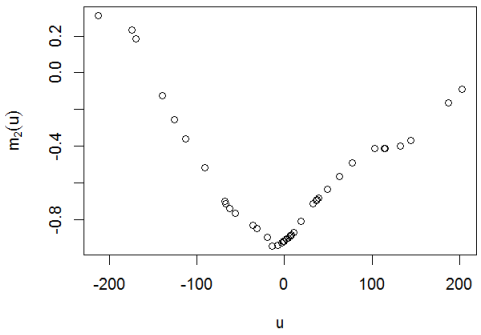
June 9: $l = 2$



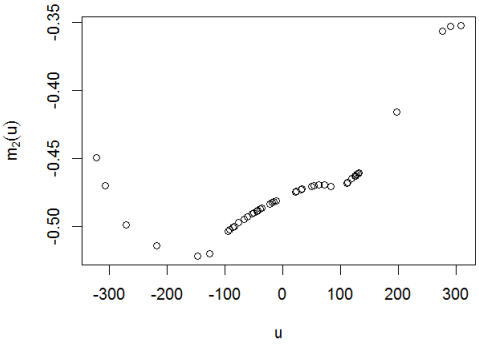
June 28: $l = 2$



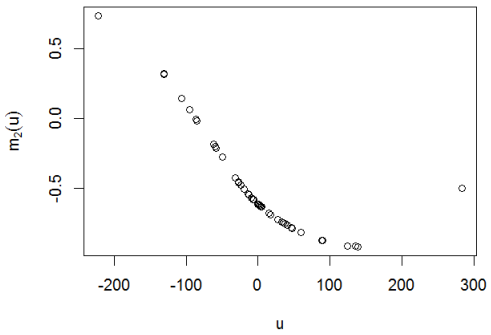
November 15: $l = 2$



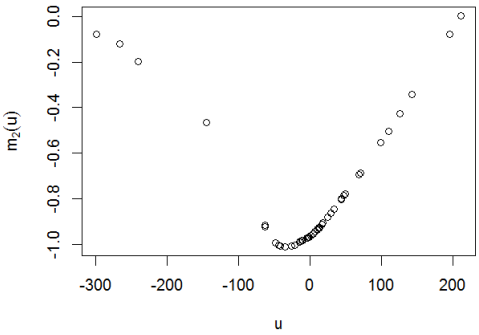
February 1: $l = 2$



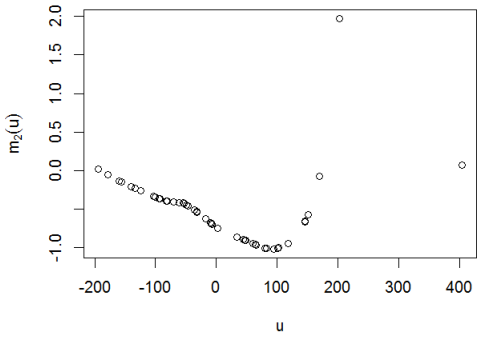
March 15: $l = 2$



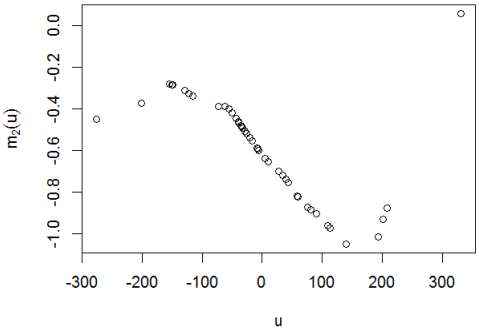
April 6: $l = 2$



May 31: $l = 2$



August 3: $l = 2$



October 27: $l = 2$

BIBLIOGRAPHY

Antonio, Di Narzo, F., Jose Aznarte, J. L., and Stigler, M. (2009), *tsDyn: Time series analysis based on dynammical systems theory*, r package version 0.5-56.

Bivand, R., with contributions by Micah Altman, Anselin, L., Assuno, R., Berke, O., Bernat, A., Blanchet, G., Blankmeyer, E., Carvalho, M., Christensen, B., Chun, Y., Dormann, C., Dray, S., Halbersma, R., Krainski, E., Legendre, P., Lewin-Koh, N., Li, H., Ma, J., Millo, G., Mueller, W., Ono, H., Peres-Neto, P., Piras, G., Reder, M., Tiefelsdorf, M., and Yu., D. (2013), *spdep: Spatial dependence: weighting schemes, statistics and models*, r package version 0.5-56.

Bosq, D. (1998), *Nonparametric statistics for stochastic processes: estimation and prediction, 2nd Edition*, Berlin: Springer-Verlag.

Cai, Z., Fan, J., and Yao, Q. (2000), "Functional-coefficient regression models for nonlinear time series," *Journal of the American Statistical Association*, 95, 941–956.

Chatfield, C. (1993), "Calculating interval forecasts," *Journal of Business & Economic Statistics*, 11, 121–135.

Chen, R. (1996), "A nonparametric multi-step prediction estimator in Markovian structures," *Statistica Sinica*, 6, 603–616.

Chen, R. and Liu, L. (2001), "Functional coefficient autoregressive models: estimation and tests of hypotheses," *Journal of Time Series Analysis*, 22, 151–173.

Chen, R. and Tsay, R. (1993), "Functional-coefficient autoregressive models," *Journal of the American Statistical Association*, 298–308.

Clements, M. and Smith, J. (1997), "The performance of alternative forecasting methods for SETAR models," *International Journal of Forecasting*, 13, 463–475.

Fan, J. and Yao, Q. (2003), *Nonlinear time series: nonparametric and parametric methods*, Springer Verlag.

Haggan, V. and Ozaki, T. (1981), "Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model," *Biometrika*, 68, 189–196.

Härdle, W., Lütkepohl, H., and Chen, R. (1997), "A Review of Nonparametric Time Series Analysis," *International Statistical Review/Revue Internationale de Statistique*, 49–72.

Harvill, J. and Ray, B. (2005), "A note on multi-step forecasting with functional coefficient autoregressive models," *International Journal of Forecasting*, 21, 717–727.

Hastie, T. J. and Tibshirani, R. J. (1990), *Generalized additive models*, vol. 43, Chapman & Hall/CRC.

Huang, J. (2001), "Concave extended linear modeling: a theoretical synthesis," *Statistica Sinica*, 11, 173–198.

Huang, J. and Shen, H. (2004), "Functional Coefficient Regression Models for Nonlinear Time Series: A Polynomial Spline Approach," *Scandinavian Journal of Statistics*, 31, 515–534.

Huang, J. and Yang, L. (2004), "Identification of nonlinear additive autoregressive models," *Journal of the Royal Statistical Society Series B*, 66, 463–477.

Kuszamaul, S., Ellis, A., Stein, J., and Johnson, L. (2010), "Lanai High-Density irradiance sensor network for characterizing solar resource variability of MW-scale PV system," in *35th Photovoltaic Specialists Conference, Honolulu, HI.*

Linton, O. (1997), "Efficient estimation of additive nonparametric regression models," *Biometricka*, 84, 469–473.

Liu, R. and Yang, L. (2010), "Spline-backfitted kernel smoothing of additive coefficient model," *Econometric Theory*, 12, 29.

Liu, R., Yang, L., and Härdle, W. K. (2011), "Oracally Efficient Two-Step Estimation of Generalized Additive Model," Tech. rep., Humboldt University, Collaborative Research Center 649.

Ma, S. and Yang, L. (2011), "Spline-backfitted kernel smoothing of partially linear additive model," *Journal of Statistical Planning and Inference*, 141, 204–219.

Reno, M., Hansen, C., and Stein, J. (2012), "Global Horizontal Irradiance Clear Sky Models: Implementation and Analysis," Tech. Rep. SAND2012-2389, Sandia National Laboratories, Albuquerque, NM.

Song, Q. and Yang, L. (2010), "Oracally efficient spline smoothing of nonlinear additive autoregression models with simultaneous confidence band," *Journal of Multivariate Analysis*, 101, 2008–2025.

Stone, C. (1994), "The use of polynomial splines and their tensor products in multivariate function estimation," *The Annals of Statistics*, 118–171.

Tong, H. (1990), *Nonlinear time series: a dynamical system approach*, Oxford: Oxford University Press.

Tong, H. and Lim, K. S. (1980), "Threshold autoregression, limit cycles and cyclical data," *Journal of the Royal Statistical Society. Series B (Methodological)*, 245–292.

Trapletti, A. and Hornik, K. (2012), *tseries: Time Series Analysis and Computational Finance*, r package version 0.10-30.

Wand, M. (2012), *KernSmooth: Functions for kernel smoothing for Wand Jones (1995)*, r package version 2.23-8.

Wang, J. and Yang, L. (2009), "Efficient and fast spline-backfitted kernel smoothing of additive models," *Annals of the Institute of Statistical Mathematics*, 61, 663–690.

Wang, L. and Yang, L. (2007), "Spline-backfitted kernel smoothing of nonlinear additive autoregression model," *Annals of Statistics*, 35, 2474–2503.