A New Approach to Modeling Multivariate Time Series on Multiple Temporal Scales

Tucker Zeleny
University of Nebraska-Lincoln, tzeleny1@huskers.unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/statisticsdiss
Part of the Applied Statistics Commons, and the Longitudinal Data Analysis and Time Series Commons

http://digitalcommons.unl.edu/statisticsdiss/16

This Article is brought to you for free and open access by the Statistics, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Dissertations and Theses in Statistics by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
A NEW APPROACH TO MODELING MULTIVARIATE TIME SERIES ON MULTIPLE TEMPORAL SCALES

by

Tucker Zeleny

A DISSERTATION

Presented to the Faculty of
The Graduate College at the University of Nebraska
In Partial Fulfillment of Requirements
For the Degree of Doctor of Philosophy

Major: Statistics

Under the Supervision of Professor David B. Marx

Lincoln, Nebraska

May, 2015
In certain situations, observations are collected on a multivariate time series at a certain temporal scale. However, there may also exist underlying time series behavior on a larger temporal scale that is of interest. Often times, identifying the behavior of the data over the course of the larger scale is the key objective. Because this large scale trend is not being directly observed, describing the trends of the data on this scale can be more difficult. To further complicate matters, the observed data on the smaller time scale may be unevenly spaced from one larger scale time point to the next. The existence of these multiple time scales means that it may be more appropriate to view the observations as coming from multiple, shorter multivariate time series occurring at each large scale time point as opposed to a single, long multivariate time series. Approaching the problem by examining the smaller scale time series separately, and then modeling the resulting estimates over the larger time scale, will provide an alternative to previous methods of dealing with similar situations while also producing additional information on the behavior of the data on the smaller observable time scale.
ACKNOWLEDGMENTS

A special thanks to Dr. David Marx for your help in completing this work. Also, thanks to the members of my committee for your contributions, as well as the University of Nebraska football team strength and conditioning coaches for your help in obtaining the data.
# TABLE OF CONTENTS

## Chapter 1

1.1 Introduction..............................................................................................................1  
1.2 Motivating Problem.................................................................................................3  
1.3 Other Approaches....................................................................................................4  
1.4 Vector Autoregressive Process..............................................................................6  
1.5 Deterministic Time Trends.....................................................................................8  
1.6 Estimation................................................................................................................9  

## Chapter 2

2.1 Modeling on the Small Scale..................................................................................14  
2.2 Multivariate Delta Method......................................................................................17  
2.3 Numerical Derivative Approximation.....................................................................22  
2.4 Simulation................................................................................................................23  
2.5 Simulation Results..................................................................................................26  

## Chapter 3

3.1 Testing Model on Weight Lifting Data.................................................................31  
3.2 Small Scale Linear Time Trends..........................................................................36  
3.3 Modeling with Small Scale Linear Time Trends...................................................38  
3.4 Model Application to Weight Lifting Data.........................................................40  
3.5 Weight Lifting Results..........................................................................................41  

## Chapter 4

4.1 Athletes with Unstable Data....................................................................................46  
4.2 Comparison of Simulation Results.........................................................................50
4.3 Identifying Trends ................................................................. 55

4.4 Lag Selection ................................................................. 57

4.5 Weight Lifting Results for Athletes with Unstable Data ........................................ 60

Chapter 5

5.1 Discussion ................................................................. 64

5.2 Future Research ............................................................. 65

5.3 Conclusion ................................................................. 66

References ................................................................. 68
LIST OF FIGURES AND TABLES

Table 2.5.1  Estimated Type I Error Rate with Modeling on Small Scale, $T = S = 100$.............27
Table 2.5.2  Estimated Power with Modeling on Small Scale, $T = S = 100$..............................27
Table 2.5.3  Estimated Type I Error Rate with Non-overlapping Averages, $T = S = 100$.......28
Table 2.5.4  Estimated Power with Non-overlapping Averages, $T = S = 100$.......................28
Table 2.5.5  Estimated Type I Error Rate with Modeling on Small Scale, $T = S = 20$...........29
Table 2.5.6  Estimated Power with Modeling on Small Scale, $T = S = 20$.............................29
Table 2.5.7  Estimated Type I Error Rate with Non-overlapping Averages, $T = S = 20$.......30
Table 2.5.8  Estimated Power with Non-overlapping Averages, $T = S = 20$.......................30
Table 3.1.1  Estimated Type I Error Rate with Modeling on Small Scale, $T = 10$, $S = 20$.......33
Table 3.1.2  Estimated Power with Modeling on Small Scale, $T = 10$, $S = 20$.....................33
Table 3.1.3  Estimated Type I Error Rate with Non-overlapping Averages, $T = 10$, $S = 20$...33
Table 3.1.4  Estimated Power with Non-overlapping Averages, $T = 10$, $S = 20$.............33
Figure 3.1.1  Plot of Athlete 11 Bench Press Data, Daily Estimates, and Trend Estimates........35
Table 3.1.5  Athlete 11 Bench Press Parameter Estimates and Standard Errors.....................35
Table 3.5.1  Bench Press Strength Trend Estimates and Standard Errors..................................42
Table 3.5.2  Bench Press Conditioning Trend Estimates and Standard Errors.......................43
Table 4.2.1  Number of Non-stationary Series Out of 1000 Simulations................................52
Table 4.2.2  Number of Non-stationary Large Scale Series Out of 1000 Simulations..............53
Table 4.2.3  Number of Non-stationary Small Scale Series Out of 1000 Simulations.............54
Table 4.3.1  Proportions of Simulations Identified Correctly As Containing Positive or Negative Large Scale Trends.................................................................56
Table 4.5.1  Bench Press Strength Trend Estimates and Standard Errors................................61
Table 4.5.2  Bench Press Conditioning Trend Estimates and Standard Errors.......................62
CHAPTER 1

1.1 Introduction

When working with data that follow the pattern of a time series, there may exist multiple temporal scales to consider. For example, observations are collected on a smaller time scale (e.g. minute-to-minute), and also exhibit time series behavior over a larger time scale (e.g. day-to-day). Variability of the data on the smaller scale can obscure the time series behavior of the data on the larger scale, making it more difficult to identify the larger scale trends. One common approach to better examine the data on a larger scale is to smooth the observed data on the smaller scale. Typically, the smoothing is done by utilizing some form of non-overlapping averages or non-overlapping sums, as will be discussed in Section 1.3. For example, if observations are made on a minute-to-minute scale, the data could be smoothed to examine the day-to-day behavior by taking daily averages or daily sums of the minute-to-minute observations.

However, there are times when the behavior of the data on the smaller scale is also of interest. Because of this interest, it would still be appropriate to model the data on the smaller scale, as well as smoothing it in some way to model the data on the larger scale. In addition, the spacing of the observations may not be constant throughout the entire data set. First consider a case where the spacing is constant: suppose that observations are made each minute of every hour over the course of a month. Here, the time spacing between every consecutive pair of observations is one minute. Because of the equal spacing throughout the entire sequence, the observations on the minute-to-minute scale may be modeled as a single time series in order to examine behavior on the fine scale. Daily averages could also be used to smooth the data in order to examine the behavior on the larger, day-to-day scale.

Now consider a case where observations are made each minute for one hour per day over the course of a month. In this situation, consecutive observations occurring on the
same day will be only one minute apart, but consecutive observations across days will be several hours apart. To examine the data on the day-to-day scale, daily averages could still be used to smooth the data to examine trends on a larger scale. However, when examining the behavior of the data on the finer, minute-to-minute scale, it may no longer be appropriate to treat the observations as coming from a single, long time series. The modeling of the data as a single time series is especially unreasonable if it is unexpected for the data to behave similarly from day to day.

Instead, under these circumstances, it is more appropriate to model the observations on the smaller scale as several shorter time series that can be modeled separately at each “large scale time point”. In the example with observations collected each minute for one hour each day, modeling the small scale time series separately would mean creating a separate model for the observations on the minute-to-minute scale for each day. Doing so allows for accurate estimation of the trends and a more detailed look at the data on the small scale at each large scale time point. Through this modeling, the stochastic autoregressive behavior of the data on the smaller scale can be separated from the non-autoregressive terms of the model, such as the intercept. By using the estimates of these non-autoregressive terms at each large scale time point, the data will be smoothed and the large scale trends of the data can then be modeled.

By using this approach of modeling separate time series on the small scale at each large scale time point to estimate the non-autoregressive terms, and then modeling these estimates over the large scale (as opposed to the use of sums and averages) not only can the larger scale trends be estimated, but the behavior of the data on the smaller scale can also be better examined. In addition, it allows for the changes in behavior on the fine scale from one large scale time point to the next to be investigated, which may provide additional insight into the analysis of the time series. While this applies to both the univariate and
multivariate cases, it is the multivariate case that will be dealt with from here, as explained in the following section.

### 1.2 Motivating Problem

The motivation for this research comes from data collected by the University of Nebraska-Lincoln football team on athletes during weight lifting sessions. Each weight lifting station is equipped with motion capture technology that measures the velocity of the bar as the athletes perform each lift. Prior to beginning a workout athletes sign into the device using a touchpad, which ensures the data collected from the lifts being performed are attributed to the correct athlete. The date of the workout, rep number, and exercise being performed are also recorded. The athletes also use the system to enter the weight on the bar prior to each set of lifts. Using the weight, along with the velocity measures, the Nebraska strength and conditioning coaches then calculate a multivariate power metric to evaluate the players’ performance.

For each completed rep, both the average power over the course of the rep, as well as the peak power over the course of that rep is recorded, resulting in a bivariate time series of data for each player. Because each player completes a program designed specifically for them, each player is evaluated independently of the others. Also, different exercises have different objectives so the data collected from different exercises are evaluated independently for each athlete.

The data in the given example occurs on two different temporal scales: a rep-to-rep scale within a single day and a day-to-day scale within a training session. For example, an athlete may complete between 15 to 30 reps for a given exercise in a given day. The winter training session lasts approximately six weeks with athletes working out two to three times per week, resulting in 12 to 18 days of data collection. While the observations are made on
the smaller rep-to-rep scale, the coaches are also interested in the trends of the athletes’ lifting performances on the larger day-to-day scale over the course of the training period in addition to the trends of the athletes within a single workout.

Within the data on the rep-to-rep scale over the course of the entire training period, there are wide variations in spacing. The time spacing between consecutive reps within the same day is typically only seconds apart, while the time spacing between consecutive reps across days is many hours apart. Because of the time separation, when considering the data on the rep-to-rep scale it is more appropriate to treat data collected on separate days as separate, shorter bivariate time series that may exhibit different behavior as opposed to a single, longer bivariate time series. By modeling the data from each day separately, the athletes’ daily performances are more specifically evaluated.

Once the rep-to-rep bivariate time series have been evaluated for each day, the resulting parameter estimates will provide a measure of an athlete’s daily performance for both variables. These measures of day-to-day performance for the two variables represent the bivariate time series on the larger day-to-day scale. The trends present in the large scale time series are then estimated by fitting a model to this larger scale time series. The focus of this research is comparing the modeling process outlined here to other approaches in general, as well as evaluating its ability to identify the small scale and large scale trends present in the real data collected during the weight lifting sessions.

1.3 Other Approaches

The remainder of this chapter focuses on the work from previous literature, beginning with previous methods of modeling time series data on various time scales and then also describing the modeling of vector autoregressive processes on a single time scale. Of the other methods that have been used in similar situations, most treat the observable
time series data on the smallest temporal scale as coming from a single, long time series. When attempting to analyze the larger temporal scale trends, these methods essentially reduce the resolution at which the data are examined. This is done in a variety of ways.

Ferreira et al. (2006) discuss modeling trends on a scale larger than the observed data by using non-overlapping and adjacent averages, and then implement this method on river flow data. Coughlin and Eto (2010) also used adjacent averages to model wind power data on a larger time scale. For example, if observations are made daily and the weekly trend is of greater interest, the daily observations will be averaged by week and these weekly averages will then be modeled using autoregressive methods.

Similarly for a time series of data collected daily, Kercheval and Liu (2010) discuss summing over an $n$-day period in order to examine the behavior of a time series at a larger time scale. They then apply this method of modeling $n$-day sums to financial data to make longer term forecasts.

Along with dividing the highest frequency observations into "grids" and then computing averages over these grids in order to estimate the behavior of a time series on a lower frequency scale, Zhang, Mykland and Ait-Sahalia (2005) also omit data in order to view the process at different scales. For example, if data is collected every second but a lower resolution time series would be of more use, only observations taken each minute could be utilized i.e. 59 out of every 60 observations are omitted.

Other methods involve filtering data in order to smooth noisy data. Through the smoothing, the smaller scale fluctuations can be eliminated and the underlying trends of the data can more easily be identified. This is done through the use of methods such as kernel functions and weighted averages as outlined in Vespier et al. (2012), Scott (2000), and Harvey and Koopman (2009).
Another method of smoothing is through the use of local linear regression, in which a linear regression model is fit at each observation point of the explanatory variable using weighted least squares. This method is described by Cleveland (1979) and its application is also utilized by Aneiros-Perez, Cao, and Vilar-Fernandez (2010) and Ruppert and Wand (1994).

However, as previously mentioned, these methods essentially are ways of viewing a single time series at different resolutions or frequencies. None of these methods consider the observations on the smallest temporal scale to be several separate time series, each occurring at a time point on the larger temporal scale where the behavior at each large scale time point also follows a time series pattern. Also, most of this previous work considers only the case of a univariate time series; only Ruppert and Wand (1994) describe the application for the multivariate case.

1.4 Vector Autoregressive Process

One of the most common, successful, and flexible models for analyzing multivariate time series is the vector autoregressive (VAR) model, attributed to Zivot and Wang (2006). VAR models have been found to be very useful in forecasting, particularly in the behavior of economic or financial time series as described by Luetkepohl (2011), Nason (2006), and Kercheval and Liu (2010). Consider first a multivariate time series occurring on only one time scale. In the context of the weight lifting example described previously, this would be the equivalent of having only one observation per day and thus only modeling on the larger day-to-day scale. Let \( \mathbf{x}_t = (x_{1,t}, x_{2,t}, \ldots, x_{n,t})' \) be an \((n \times 1)\) vector of time series variables at time \( t \). As shown by Luetkepohl (2011), then a general \( p \)-lag VAR, known as VAR\((p)\), for a purely stochastic autoregressive process is written as
\[ x_t = \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + \ldots + \Phi_p x_{t-p} + \epsilon_t \tag{1.4.1} \]

where \( \Phi_j (j = 1 \text{ to } p) \) are \((n \times n)\) matrices of coefficients and \( \epsilon_t (t = 1 \text{ to } T) \) are \((n \times 1)\) vectors of error terms such that \( E(\epsilon_t) = 0, E(\epsilon_t \epsilon_t') = \Sigma \) and \( E(\epsilon_t \epsilon_m') = 0 \) for all \( t \neq m \). Another common assumption, in addition to the mean and covariance matrix of the error terms, is that they also follow a multivariate normal distribution, i.e. \( \epsilon_t \sim N(0, \Sigma) \).

Using lag operator notation, the VAR\((p)\) can be rewritten as

\[ \Phi(L)x_t = \epsilon_t \tag{1.4.2} \]

where \( \Phi(L) = I_n - \Phi_1 L - \Phi_2 L^2 - \ldots - \Phi_p L^p \). The VAR\((p)\) process is stable if the roots of

\[ \det(I_n - \Phi_1 u - \Phi_2 u^2 - \ldots - \Phi_p u^p) = 0 \tag{1.4.3} \]

have modulus greater than one. Equivalently, the VAR\((p)\) is stable if the eigenvalues of

\[ F = \begin{bmatrix} \Phi_1 & \Phi_2 & \Phi_3 & \ldots & \Phi_p \\ I_n & 0 & 0 & \ldots & 0 \\ 0 & I_n & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \ldots & \vdots \\ 0 & 0 & \ldots & I_n & 0 \end{bmatrix} \tag{1.4.4} \]

have modulus less than one, where \( I_n \) is an \((n \times n)\) identity matrix, \( 0 \) is an \((n \times n)\) matrix with each element equal to zero, and thus \( F \) is \((np \times np)\) as shown by Zivot and Wang (2006).

Under the usual assumptions of a VAR\((p)\) process, a stable process has means, variances and covariance structure that are time invariant and thus the process is also a stationary
process, as stated by Luetkepohl (2011). Therefore, a VAR($p$) process that is shown to be stable can be estimated under the assumption of stationarity.

1.5 Deterministic Time Trends

The model outlined above is a purely stochastic autoregressive process with zero mean. In many cases, it is appropriate to also include a deterministic term that describes the time trends of the data. After adding the deterministic term the model is written as

$$y_t = \mu_t + x_t$$

(1.5.1)

where the deterministic term $\mu_t$ is a linear time trend i.e. $\mu_t = \mu_0 + \mu_1 t$, and $x_t$ is the stochastic autoregressive process described in the previous section so that (1.5.1) defines a VAR($p$) process that now includes a deterministic term, as described by Luetkepohl (2011). It is $y_t$ that is actually the vector of observed variables. Thus $\mu_1$, which is of greatest interest, is unobserved.

Luetkepohl (2011) goes on to show that pre-multiplying (1.5.1) by $\Phi(L)$ results in

$$\Phi(L)y_t = \Phi(L)\mu_t + \epsilon_t$$

(1.5.2)

which shows that $y_t$ has the same VAR($p$) structure as $x_t$. Equation 1.5.2 is then rewritten as

$$\Phi(L)y_t = \theta_t + \epsilon_t$$

or

(1.5.3)

$$y_t = \theta_0 + \theta_1 t + \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + \epsilon_t$$

(1.5.4)

where $\Phi(L)\mu_t = \theta_t = \theta_0 + \theta_1 t$ and thus
\[ \theta_0 = (I_n - \sum_{j=1}^{p} \Phi_j)\mu_0 + (\sum_{j=1}^{p} \Phi_j)\mu_1 \]  
(1.5.5)

\[ \theta_1 = (I_n - \sum_{j=1}^{p} \Phi_j)\mu_1. \]

The parameters as given by the model in Equation 1.5.4 are estimated. Once estimates have been found for \( \theta_t \) and each \( \Phi_j (j = 1 \text{ to } p) \), the estimates for \( \mu_t \), which represent the underlying trends of the data, are then found by solving the equations of (1.5.5) in terms of the parameters of \( \mu_t \), resulting in

\[ \mu_0 = (I_n - \sum_{j=1}^{p} \Phi_j)^{-1}\theta_0 - (I_n - \sum_{j=1}^{p} \Phi_j)^{-1}(\sum_{j=1}^{p} \Phi_j)(I_n - \sum_{j=1}^{p} \Phi_j)^{-1}\theta_1 \]  
(1.5.6)

\[ \mu_1 = (I_n - \sum_{j=1}^{p} \Phi_j)^{-1}\theta_1 \]

1.6 Estimation

Johnson and Wichern (2007) describe how the set of models for all \( t \) (\( t = 1 \text{ to } T \)) are written in matrix notation and estimated through linear regression in the multivariate case. Suppose a sample of size \( T \) is taken so that there are \( y_1, y_2, \ldots, y_T \), each an \( (n \times 1) \) vector of observations. Let \( Y = [y_1, y_2, \ldots, y_T]' \) so that

\[
Y = \begin{bmatrix}
y_{1,1} & y_{2,1} & y_{3,1} & \cdots & y_{n,1} \\
y_{1,2} & y_{2,2} & y_{3,2} & \cdots & y_{n,2} \\
y_{1,3} & y_{2,3} & y_{3,3} & \cdots & y_{n,3} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
y_{1,T} & y_{2,T} & y_{3,T} & \cdots & y_{n,T}
\end{bmatrix}
\]

where \( y_{i,t} \) is the observation of the \( i^{th} \) variable at time \( t \). Thus the \( t^{th} \) row of \( Y \) consists of the observations on the \( n \) variables at time \( t \).
Let $A = [a_1, a_2, ..., a_T]'$ where $a_t = [1, t, y_{t-1}', ..., y_{t-p}']'$ so that

$$A = \begin{bmatrix}
1 & 1 & y_{1,0} & ... & y_{n,0} & y_{1,-1} & ... & y_{n,-1} & ... & y_{1,-p} & ... & y_{n,-p} \\
1 & 2 & y_{1,1} & ... & y_{n,1} & y_{1,0} & ... & y_{n,0} & ... & y_{1,-2} & ... & y_{n,-2} \\
1 & 3 & y_{1,2} & ... & y_{n,2} & y_{1,1} & ... & y_{n,1} & ... & y_{1,-3} & ... & y_{n,-3} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & T & y_{1,T-1} & ... & y_{n,T-1} & y_{1,T-2} & ... & y_{n,T-2} & ... & y_{1,T-p} & ... & y_{n,T-p}
\end{bmatrix}$$

Finally, let $B = [\theta_0, \theta_1, \phi_1, ..., \phi_p]'$, and $E = [\varepsilon_1, \varepsilon_2, ..., \varepsilon_T]'$ where

$$\theta_0 = \begin{bmatrix}
\theta_{1,0} \\
\theta_{2,0} \\
\vdots \\
\theta_{n,0}
\end{bmatrix}, \quad \theta_1 = \begin{bmatrix}
\theta_{1,1} \\
\theta_{2,1} \\
\vdots \\
\theta_{n,1}
\end{bmatrix}, \quad \phi_j = \begin{bmatrix}
\phi_{1,1,j} & \phi_{1,2,j} & \cdots & \phi_{1,n,j} \\
\phi_{2,1,j} & \phi_{2,2,j} & \cdots & \phi_{2,n,j} \\
\vdots & \vdots & \ddots & \vdots \\
\phi_{n,1,j} & \phi_{n,2,j} & \cdots & \phi_{n,n,j}
\end{bmatrix}, \quad \varepsilon_t = \begin{bmatrix}
\varepsilon_{1,t} \\
\varepsilon_{2,t} \\
\vdots \\
\varepsilon_{n,t}
\end{bmatrix}$$

so that

$$B = \begin{bmatrix}
\theta_{1,0} & \theta_{2,0} & \cdots & \theta_{n,0} \\
\theta_{1,1} & \theta_{2,1} & \cdots & \theta_{n,1} \\
\phi_{1,1,1} & \phi_{2,1,1} & \cdots & \phi_{n,1,1} \\
\vdots & \vdots & \ddots & \vdots \\
\phi_{1,n,1} & \phi_{2,n,1} & \cdots & \phi_{n,n,1} \\
\phi_{1,1,2} & \phi_{2,1,2} & \cdots & \phi_{n,1,2} \\
\vdots & \vdots & \ddots & \vdots \\
\phi_{1,n,2} & \phi_{2,n,2} & \cdots & \phi_{n,n,2} \\
\phi_{1,1,p} & \phi_{2,1,p} & \cdots & \phi_{n,1,p} \\
\vdots & \vdots & \ddots & \vdots \\
\phi_{1,n,p} & \phi_{2,n,p} & \cdots & \phi_{n,n,p}
\end{bmatrix}, \quad E = \begin{bmatrix}
\varepsilon_{1,1} & \varepsilon_{2,1} & \varepsilon_{3,1} & \cdots & \varepsilon_{n,1} \\
\varepsilon_{1,2} & \varepsilon_{2,2} & \varepsilon_{3,2} & \cdots & \varepsilon_{n,2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\varepsilon_{1,T} & \varepsilon_{2,T} & \varepsilon_{3,T} & \cdots & \varepsilon_{n,T}
\end{bmatrix}$$

Thus, as shown by Johnson and Wichern (2007), the model given by (1.5.4) can be rewritten to include all times $t$ ($t = 1$ to $T$) as

$$Y = AB + E \quad (1.6.1)$$
The model is now in the form of a multivariate multiple regression model so that estimates of \( \theta_0, \theta_1, \Phi_1, \ldots, \Phi_p \) can be obtained by estimating \( \mathbf{B} \) through the use of ordinary least squares (OLS) as shown below:

\[
\hat{\mathbf{B}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Y} \tag{1.6.2}
\]

If no restrictions are imposed on the parameters then the OLS estimator of \( \mathbf{B} \) is also the generalized least squares (GLS) estimator. In addition, under the assumption that \( \varepsilon_t \) is distributed \( N(0, \Sigma) \), the OLS estimator of \( \mathbf{B} \) is also the maximum likelihood (ML) estimator, (Luetkepohl, 2011). The proof, as given by Hamilton (1994), is provided below:

Consider \( y_t \mid y_{t-1}, y_{t-2}, \ldots, y_{t-p+1} \sim N(\theta_0 + \theta_1 y_{t-1} + \Phi_1 y_{t-2} + \ldots + \Phi_p y_{t-p}, \Sigma) \). Then the log-likelihood is given by

\[
L(\mathbf{Y} ; \mathbf{B}) = -\frac{nT}{2} \log(2\pi) + \frac{T}{2} \log(\Sigma^{-1}) - \frac{1}{2} \sum_{t=1}^{T} (y_t - \mathbf{B}'\mathbf{a}_t)' \Sigma^{-1} (y_t - \mathbf{B}'\mathbf{a}_t) \tag{1.6.3}
\]

To find the estimate of \( \mathbf{B} \) that maximizes (1.6.3), \( \sum_{t=1}^{T} (y_t - \mathbf{B}'\mathbf{a}_t)' \Sigma^{-1} (y_t - \mathbf{B}'\mathbf{a}_t) \) must be minimized.

\[
\sum_{t=1}^{T} (y_t - \mathbf{B}'\mathbf{a}_t)' \Sigma^{-1} (y_t - \mathbf{B}'\mathbf{a}_t) \\
= \sum_{t=1}^{T} (y_t - \hat{\mathbf{B}}'\mathbf{a}_t + \hat{\mathbf{B}}'\mathbf{a}_t - \mathbf{B}'\mathbf{a}_t)' \Sigma^{-1} (y_t - \hat{\mathbf{B}}'\mathbf{a}_t + \hat{\mathbf{B}}'\mathbf{a}_t - \mathbf{B}'\mathbf{a}_t) \\
= \sum_{t=1}^{T} (\hat{\varepsilon}_t + (\hat{\mathbf{B}} - \mathbf{B})'\mathbf{a}_t)' \Sigma^{-1} (\hat{\varepsilon}_t + (\hat{\mathbf{B}} - \mathbf{B})'\mathbf{a}_t) \\
= \sum_{t=1}^{T} (\epsilon_t + \Sigma^{-1}\epsilon_t + 2\sum_{t=1}^{T} \epsilon_t' \Sigma^{-1}(\hat{\mathbf{B}} - \mathbf{B})'\mathbf{a}_t + \sum_{t=1}^{T} \mathbf{a}_t' (\hat{\mathbf{B}} - \mathbf{B}) \Sigma^{-1}(\hat{\mathbf{B}} - \mathbf{B})'\mathbf{a}_t \\
= \sum_{t=1}^{T} \hat{\varepsilon}_t' \Sigma^{-1}\hat{\varepsilon}_t + 2tr(\sum_{t=1}^{T} \epsilon_t' \Sigma^{-1}(\hat{\mathbf{B}} - \mathbf{B})'\mathbf{a}_t) + \sum_{t=1}^{T} \mathbf{a}_t' (\hat{\mathbf{B}} - \mathbf{B}) \Sigma^{-1}(\hat{\mathbf{B}} - \mathbf{B})'\mathbf{a}_t \\
= \sum_{t=1}^{T} \hat{\varepsilon}_t' \Sigma^{-1}\hat{\varepsilon}_t + 2tr(\Sigma_{t=1}^{T} \epsilon_t' \Sigma^{-1}(\hat{\mathbf{B}} - \mathbf{B})'\mathbf{a}_t) + \sum_{t=1}^{T} \mathbf{a}_t' (\hat{\mathbf{B}} - \mathbf{B}) \Sigma^{-1}(\hat{\mathbf{B}} - \mathbf{B})'\mathbf{a}_t
\]
Thus, the estimate of $B$ that maximizes the likelihood function is the estimate of $B$ that minimizes $\sum_{t=1}^{T} a_t (\hat{B} - B) \Sigma^{-1} (\hat{B} - B)^T a_t$. Because $\Sigma$ is a positive definite matrix, $\Sigma^{-1}$ is also a positive definite matrix, which means that the smallest value that this term can occur when $\hat{B} - B = 0$. Thus the maximum likelihood estimate of $B$ is given by $\hat{B}$ which is the ordinary least squares estimate.

Similarly, as stated by Hamilton (1994), it follows that the maximum likelihood estimate of $\Sigma$ is

$$\hat{\Sigma} = \frac{1}{T} \hat{E}_{\hat{E}} = \frac{1}{T} (Y - A\hat{B})' (Y - A\hat{B}) \quad (1.6.4)$$

The proof is given below:

$$L(Y; \hat{B}, \Sigma) = -\frac{TN}{2} \log(2\pi) + \frac{T}{2} \log(\Sigma^{-1}) - \frac{1}{2} \sum_{t=1}^{T} (y_t - \hat{B}'a_t)' \Sigma^{-1} (y_t - \hat{B}'a_t)$$

$$\frac{\partial}{\partial \Sigma^{-1}} L(Y; \hat{B}, \Sigma) = \frac{T}{2} \Sigma - \frac{1}{2} \sum_{t=1}^{T} (y_t - \hat{B}'a_t)(y_t - \hat{B}'a_t)' = 0$$

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{B}'a_t)(y_t - \hat{B}'a_t)'$$

The ML estimate, however, results in biased estimates for the elements of $\Sigma$ so the bias is corrected by accounting for degrees of freedom for error and dividing by $(T - (2 + np))$ rather than by $T$ as $(2 + np)$ gives the total number of parameters being fit in each model.

If $B$ is considered as $B = [\beta_1, \beta_2, ..., \beta_n]$ then $\hat{B}$ will also follow a multivariate normal distribution with $E(\hat{B}) = B$ and $\text{Cov}(\hat{\beta}_i, \hat{\beta}_k) = \sigma_{ik}(A'A)^{-1}$ where $i, k = 1, ..., n$ and $\sigma_{ik}$ is the $i^{th}$,
Using this information, the standard errors for the estimates of $\theta_0, \theta_1, \Phi_1, ..., \Phi_p$ are also calculated. In addition, $\mu_0$ and $\mu_1$ can be estimated by entering the estimates of $\theta_0, \theta_1, \Phi_1, ..., \Phi_p$ into (1.5.6). Finding the standard errors of the estimates of $\mu_0$ and $\mu_1$ will be found using the multivariate delta method, which will be addressed in the next chapter.
CHAPTER 2

2.1 Modeling on the Small Scale

A model is proposed here that accounts for a multivariate time series on multiple temporal scales; its development, application, and evaluation are the focus of the remaining chapters. For simplicity’s sake, first consider only a bivariate VAR(1) model with two temporal scales. These temporal scales will be referred to as the small scale (for example, the rep-to-rep scale within a given day on which observations are made in the previous weight lifting example) and the large scale (the day-to-day scale over the course of a training period).

Let \( T \) be the number of large scale time points for which observations are made on the large scale, e.g. weight lifting data is collected over \( T \) days. In Chapter 1 a vector autoregressive model on a single time scale was considered, which corresponds to only one rep recorded per day in the context of the weight lifting data. However, because several reps are recorded each day, there is also the small scale within day series. Thus, for each large scale time \( t (t = 1, \ldots, T) \) there is a series of small scale observations. Again, for simplicity’s sake, consider the number of small scale observations to be constant for each \( t \), i.e. for each \( t \) \((t = 1, \ldots, T)\), there are \( S \) observations on each small scale series.

Similar to hierarchical linear modeling, as described by Woltman et al. (2012), the small scale models at each large scale time point are considered first. The model for the small scale bivariate time series at a given large scale time \( t \) is given as follows: Let \( \mathbf{z}_{s,t} = (z_{1,s,t}, z_{2,s,t})' \) be the vector of observations at small scale time \( s \) \((s = 1, \ldots, S)\) and large scale time \( t \) \((t = 1, \ldots, T)\). Then

\[
\mathbf{z}_{s,t} = \mathbf{y}_t + \mathbf{w}_{s,t} \tag{2.1.1}
\]
where at large scale time \( t \), \( y_t \) is strictly an intercept for simplicity (a linear time trend term may be included so that \( y_t = y_{0,t} + y_{1,t}^s \)) and \( w_{s,t} \) is a stochastic autoregressive term so that

\[
    w_{s,t} = \Gamma_{1,t} w_{s-1,t} + \omega_{s,t} 
\]

where \( \Gamma_{1,t} \) is a \((2 \times 2)\) coefficient matrix and \( \omega_{s,t} \sim N(0, \Sigma_t) \) and \( E(\omega_{s,t} \omega_{l,t}') = 0 \) for all \( s \neq l \).

Thus (2.1.1) can be rewritten as

\[
    z_{s,t} = y_t + \Gamma_{1,t} w_{s-1,t} + \omega_{s,t} \quad \text{or} \\
    z_{s,t} = \alpha_t + \Gamma_{1,t} z_{s-1,t} + \omega_{s,t} 
\]

where \( \alpha_t = (I_2 - \Gamma_{1,t}) y_t \), which can be done because of the properties, assumptions, and steps outlined in section 1.5, specifically in Equations 1.5.1 through 1.5.5.

Equations 2.1.1 through 2.1.4 deal with a case of a multivariate time series as described in section 1.5 where \( n=2 \) and \( p=1 \). Section 1.5 illustrated the case where there was only one observation at each large scale time point \( t \), and thus there was only one time scale. Now, because there are multiple small scale observations at each large scale time point \( t \), there are two time scales. Notation has been adjusted in order to fit a model to those observations on the small scale at each large scale time point \( t \).

Using the model given by (2.1.4) and the methods described in Section 1.6, the OLS estimates for \( \alpha_t \) and \( \Gamma_{1,t} \) are obtained. Under the assumptions of stability and \( z_{s,t} \) being a normal process, these OLS estimates are also the ML estimates. Using these estimates, estimates for \( y_t \) are calculated by utilizing

\[
    \hat{y}_t = (I_2 - \hat{\Gamma}_{1,t})^{-1} \hat{\alpha}_t 
\]
Each $y_t (t = 1, \ldots, T)$ indicates the behavior of the large scale bivariate time series at large scale time $t$ because it represents the intercept portion of the model once it has been separated from the stochastic autoregressive portion of the model, $w_{x,t}$, as shown in (2.1.1). Thus, each $y_t$ is a single multivariate observation at each large scale time $t$, which is the case explained in Section 1.5. Now, not only can the time series be examined on the small scale through the model given by (2.1.4) at each large scale time $t$, but by fitting a model to the $y_t$’s, the behavior of the time series on the large scale is also examined. The model is given as

$$y_t = \theta_0 + \theta_1 t + \Phi_1 y_{t-1} + \varepsilon_t$$

(2.1.6)

which is the model outlined in (1.5.4) with the number of lags, $p$, equal to one.

Because the $y_t$’s cannot be directly observed, they must be estimated from the observable $z_{x,t}$’s. As is done in hierarchical linear modeling, these estimates are obtained through the use of OLS (Woltman et al., 2012). The estimation is described previously and thus the $\tilde{y}_t$’s that result from (2.1.5) are entered into the model given in (2.1.6) as $y_t$’s. As described in Section 1.6, it is then possible to obtain estimates for $\theta_0$, $\theta_1$, and $\Phi_1$. These estimates are then used to estimate $\mu_0$ and $\mu_1$ as illustrated in (1.5.6), as $\mu_0$ and $\mu_1$ are the deterministic large scale time trend parameters that are of the greatest interest in this case.

At this point $\tilde{\mu}_0$ and $\tilde{\mu}_1$, which are functions of $\tilde{\theta}_0$, $\tilde{\theta}_1$, and $\tilde{\Phi}_1$, will have been obtained. Consider $\boldsymbol{\psi} = \text{vec}(\theta_0, \theta_1, \Phi_1)$. As explained in Section 1.6, under the usual assumptions, $\tilde{\boldsymbol{\psi}}$ follows an approximately multivariate normal distribution, with mean vector $\boldsymbol{\psi}$. The elements of the covariance matrix of $\tilde{\boldsymbol{\psi}}$ are found as described in Section 1.6. Denote this covariance matrix of $\tilde{\boldsymbol{\psi}}$ as $\Omega$, so that $\tilde{\boldsymbol{\psi}} \sim N(\boldsymbol{\psi}, \Omega)$. 

2.2 Multivariate Delta Method

Because $\psi$ follows a multivariate normal distribution, and $\mu_0$ and $\mu_1$ are functions of $\psi$, the multivariate delta method, as described by Agresti (2002), can be utilized which shows that $\hat{\mu}_0$ and $\hat{\mu}_1$ are also approximately normally distributed with mean vectors $\mu_0$ and $\mu_1$, respectively, and that the covariance matrices can be found through the use of partial derivatives. These estimates allow for statistical inference on the $\mu_0$ and $\mu_1$ parameters of interest.

As described in the previous section, let $\psi = \text{vec}(\theta_0, \theta_1, \Phi_1)$. Also, consider the function $g: \mathbb{R}^8 \rightarrow \mathbb{R}^4$ so that $g(\psi) = \text{vec}(\mu_0, \mu_1)$. Then using the notation from Section 1.6,

$$g(\psi) = \mu = \begin{bmatrix} \mu_{1,0} \\ \mu_{2,0} \\ \mu_{1,1} \\ \mu_{2,1} \end{bmatrix}$$

First, suppose that

$$f_i = \theta_{1,1} \phi_{1,1,1} + \theta_{2,1} \phi_{1,2,1} - 2\theta_{1,1} \phi_{1,1,1} \phi_{2,2,1} + 2\theta_{1,1} \phi_{1,2,1} \phi_{2,1,1} + \theta_{1,1} \phi_{1,1,1} \phi_{2,2,1}^2$$

$$+ \theta_{2,1} \phi_{2,1,1} \phi_{1,2,1}^2 - \theta_{2,1} \phi_{1,2,1} \phi_{2,1,1} \phi_{2,2,1} - \theta_{2,1} \phi_{1,1,1} \phi_{1,2,1} \phi_{2,2,1}$$

and
\[ f_2 = \theta_{2,1} \phi_{2,2,1} + \theta_{1,1} \phi_{2,1,1} - 2 \theta_{2,1} \phi_{2,2,1} \phi_{1,1,1} + 2 \theta_{2,1} \phi_{2,1,1} \phi_{1,2,1} + \theta_{2,1} \phi_{2,2,1} \phi_{1,1,1}^2 \]

\[ + \theta_{1,1} \phi_{2,2,1} \phi_{2,1,1}^2 - \theta_{2,1} \phi_{2,1,1} \phi_{1,2,1} \phi_{1,1,1} - \theta_{1,1} \phi_{1,1,1} \phi_{2,2,1} \phi_{2,1,1} \]

Then using (1.5.6) it is seen that

\[ \mu_{1,0} = \frac{\theta_{1,0} - \theta_{1,0} \phi_{2,2,1} + \theta_{2,0} \phi_{1,2,1}}{1 - \phi_{1,1,1} - \phi_{2,2,1} + \phi_{1,1,1} \phi_{2,2,1} - \phi_{1,2,1} \phi_{2,1,1}} - \frac{f_1}{(1 - \phi_{1,1,1} - \phi_{2,2,1} + \phi_{1,1,1} \phi_{2,2,1} - \phi_{1,2,1} \phi_{2,1,1})^2} \tag{2.2.1} \]

\[ \mu_{2,0} = \frac{\theta_{2,0} - \theta_{2,0} \phi_{1,1,1} + \theta_{1,0} \phi_{2,2,1}}{1 - \phi_{1,1,1} - \phi_{2,2,1} + \phi_{1,1,1} \phi_{2,2,1} - \phi_{1,2,1} \phi_{2,1,1}} - \frac{f_2}{(1 - \phi_{1,1,1} - \phi_{2,2,1} + \phi_{1,1,1} \phi_{2,2,1} - \phi_{1,2,1} \phi_{2,1,1})^2} \tag{2.2.2} \]

\[ \mu_{1,1} = \frac{\theta_{1,1} - \theta_{1,1} \phi_{2,2,1} + \theta_{2,1} \phi_{1,2,1}}{1 - \phi_{1,1,1} - \phi_{2,2,1} + \phi_{1,1,1} \phi_{2,2,1} - \phi_{1,2,1} \phi_{2,1,1}} \tag{2.2.3} \]

\[ \mu_{2,1} = \frac{\theta_{2,1} - \theta_{2,1} \phi_{1,1,1} + \theta_{1,1} \phi_{2,2,1}}{1 - \phi_{1,1,1} - \phi_{2,2,1} + \phi_{1,1,1} \phi_{2,2,1} - \phi_{1,2,1} \phi_{2,1,1}} \tag{2.2.4} \]

By the multivariate delta method, since \( \hat{\Psi} \sim N(\Psi, \Omega) \), it is also known that \( \hat{\mu} \sim N(\mu, D \Omega D') \)

where

\[ D = \frac{\partial g(\Psi)}{\partial \Psi} = \begin{bmatrix} \frac{\partial \mu_{1,0}}{\partial \mu_{1,0}} & \frac{\partial \mu_{1,0}}{\partial \phi_{2,1,1}} & \frac{\partial \mu_{1,0}}{\partial \phi_{2,2,1}} & \frac{\partial \mu_{1,0}}{\partial \phi_{2,1,1,1}} & \frac{\partial \mu_{1,0}}{\partial \phi_{2,2,1,1}} \\ \frac{\partial \mu_{1,0}}{\partial \phi_{2,1,1}} & \frac{\partial \mu_{1,0}}{\partial \phi_{2,2,1}} & \frac{\partial \mu_{1,0}}{\partial \phi_{1,1,1}} & \frac{\partial \mu_{1,0}}{\partial \phi_{1,1,1,1}} & \frac{\partial \mu_{1,0}}{\partial \phi_{1,2,1}} \\ \frac{\partial \mu_{1,0}}{\partial \phi_{2,2,1}} & \frac{\partial \mu_{1,0}}{\partial \phi_{1,1,1}} & \frac{\partial \mu_{1,0}}{\partial \phi_{2,2,1,1}} & \frac{\partial \mu_{1,0}}{\partial \phi_{2,2,1,1,1}} & \frac{\partial \mu_{1,0}}{\partial \phi_{2,2,2,1}} \\ \frac{\partial \mu_{1,0}}{\partial \phi_{2,1,1,1}} & \frac{\partial \mu_{1,0}}{\partial \phi_{1,1,1,1}} & \frac{\partial \mu_{1,0}}{\partial \phi_{2,2,1,1,1}} & \frac{\partial \mu_{1,0}}{\partial \phi_{2,2,2,1,1}} & \frac{\partial \mu_{1,0}}{\partial \phi_{2,2,2,2,1}} \\ \frac{\partial \mu_{1,0}}{\partial \phi_{2,2,1,1}} & \frac{\partial \mu_{1,0}}{\partial \phi_{2,2,2,1}} & \frac{\partial \mu_{1,0}}{\partial \phi_{2,2,2,1,1}} & \frac{\partial \mu_{1,0}}{\partial \phi_{2,2,2,2,1}} & \frac{\partial \mu_{1,0}}{\partial \phi_{2,2,2,2,2,1}} \end{bmatrix} \tag{2.2.5} \]

These partial derivatives are as follows:

\[ \frac{\partial \mu_{1,0}}{\partial \theta_{1,0}} = \frac{1 - \phi_{2,2,1}}{1 - \phi_{1,1,1} - \phi_{2,2,1} + \phi_{1,1,1} \phi_{2,2,1} - \phi_{1,2,1} \phi_{2,1,1}} \]
\[
\frac{\partial \mu_{1,0}}{\partial \theta_{2,0}} = \frac{\phi_{1,2,1}}{1 - \phi_{1,1,1} - \phi_{2,2,1} + \phi_{1,1,1} \phi_{2,2,1} - \phi_{1,2,1} \phi_{2,2,1}}
\]

\[
\frac{\partial \mu_{1,0}}{\partial \theta_{1,1}} = -\frac{(\phi_{1,1,1}^2 - 2\phi_{1,1,1} \phi_{2,2,1} + 2\phi_{1,2,1} \phi_{2,2,1} + \phi_{1,1,1} \phi_{2,2,1}^2 - \phi_{1,2,1} \phi_{2,2,1}^2)}{(1 - \phi_{1,1,1} - \phi_{2,2,1} + \phi_{1,1,1} \phi_{2,2,1} - \phi_{1,2,1} \phi_{2,2,1})^2}
\]

\[
\frac{\partial \mu_{1,0}}{\partial \theta_{2,1}} = -\frac{(\phi_{1,2,1} + \phi_{1,1,1} \phi_{2,2,1}^2 - \phi_{1,1,1} \phi_{2,2,1} \phi_{2,2,1})}{(1 - \phi_{1,1,1} - \phi_{2,2,1} + \phi_{1,1,1} \phi_{2,2,1} - \phi_{1,2,1} \phi_{2,2,1})^2}
\]

\[
\frac{\partial \mu_{1,0}}{\partial \phi_{1,1,1}} = \frac{2f_1(\phi_{2,2,1} - 1)}{(1 - \phi_{1,1,1} - \phi_{2,2,1} + \phi_{1,1,1} \phi_{2,2,1} - \phi_{1,2,1} \phi_{2,2,1})^3}
\]

\[
\frac{\partial \mu_{1,0}}{\partial \phi_{2,2,1}} = \frac{2f_1(\phi_{1,1,1} - 1)}{(1 - \phi_{1,1,1} - \phi_{2,2,1} + \phi_{1,1,1} \phi_{2,2,1} - \phi_{1,2,1} \phi_{2,2,1})^3}
\]

\[
\frac{\partial \mu_{1,0}}{\partial \phi_{1,2,1}} = \frac{2f_1(\phi_{1,1,1} - 1)}{(1 - \phi_{1,1,1} - \phi_{2,2,1} + \phi_{1,1,1} \phi_{2,2,1} - \phi_{1,2,1} \phi_{2,2,1})^3}
\]
\[
\frac{\partial u_{2,0}}{\partial \theta_{1,0}} = \frac{\Phi_{2,1,1}}{1 - \phi_{1,1,1} - \phi_{2,2,1} + \phi_{1,1,1}\phi_{2,2,1} - \phi_{1,2,1}\phi_{2,2,1}}
\]

\[
\frac{\partial u_{2,0}}{\partial \theta_{2,0}} = \frac{1 - \phi_{1,1,1}}{1 - \phi_{1,1,1} - \phi_{2,2,1} + \phi_{1,1,1}\phi_{2,2,1} - \phi_{1,2,1}\phi_{2,2,1}}
\]

\[
\frac{\partial u_{2,0}}{\partial \theta_{1,1}} = \frac{-(\phi_{2,1,1} + \phi_{1,2,1}\phi_{2,1,1} - \phi_{1,1,1}\phi_{2,1,1})}{(1 - \phi_{1,1,1} - \phi_{2,2,1} + \phi_{1,1,1}\phi_{2,2,1} - \phi_{1,2,1}\phi_{2,2,1})^2}
\]

\[
\frac{\partial u_{2,0}}{\partial \phi_{1,1,1}} = \frac{2f_2(\phi_{2,2,1} - 1)}{(1 - \phi_{1,1,1} - \phi_{2,2,1} + \phi_{1,1,1}\phi_{2,2,1} - \phi_{1,2,1}\phi_{2,2,1})^3}
\]

\[
+ \frac{(-\theta_{2,0} + \theta_{2,0}\phi_{1,1,1} + \theta_{2,0}\phi_{2,2,1} - \theta_{2,0}\phi_{1,1,1}\phi_{2,2,1} + \theta_{2,0}\phi_{1,2,1}\phi_{2,2,1})^2 + (-\theta_{2,0} - \theta_{2,0}\phi_{1,1,1} + \theta_{1,0}\phi_{2,2,1})(1 - \phi_{2,2,1})}{(1 - \phi_{1,1,1} - \phi_{2,2,1} + \phi_{1,1,1}\phi_{2,2,1} - \phi_{1,2,1}\phi_{2,2,1})^2}
\]

\[
\frac{\partial u_{2,0}}{\partial \phi_{2,1,1}} = \frac{-2f_2\phi_{1,2,1}}{(1 - \phi_{1,1,1} - \phi_{2,2,1} + \phi_{1,1,1}\phi_{2,2,1} - \phi_{1,2,1}\phi_{2,2,1})^3}
\]

\[
+ \frac{(\theta_{1,0} - \theta_{1,0}\phi_{1,1,1} - \theta_{1,0}\phi_{2,2,1} + \theta_{1,0}\phi_{1,1,1}\phi_{2,2,1} + \theta_{2,0}\phi_{1,2,1}\phi_{2,2,1})^2}{(1 - \phi_{1,1,1} - \phi_{2,2,1} + \phi_{1,1,1}\phi_{2,2,1} - \phi_{1,2,1}\phi_{2,2,1})^2}
\]

\[
\frac{\partial u_{2,0}}{\partial \phi_{1,1,1}} = \frac{-2f_2\phi_{1,2,1}}{(1 - \phi_{1,1,1} - \phi_{2,2,1} + \phi_{1,1,1}\phi_{2,2,1} - \phi_{1,2,1}\phi_{2,2,1})^3}
\]

\[
+ \frac{(\theta_{1,0} - \theta_{1,0}\phi_{1,1,1} - \theta_{1,0}\phi_{2,2,1} + \theta_{1,0}\phi_{1,1,1}\phi_{2,2,1} + \theta_{2,0}\phi_{1,2,1}\phi_{2,2,1})^2}{(1 - \phi_{1,1,1} - \phi_{2,2,1} + \phi_{1,1,1}\phi_{2,2,1} - \phi_{1,2,1}\phi_{2,2,1})^2}
\]
\[
\frac{\partial \mu_{2,0}}{\partial \phi_{2,1}} = \frac{2f_2(\phi_{1,1,1}-1)}{(1-\phi_{1,1,1}+\phi_{1,1,1}\phi_{2,2,1}-\phi_{1,2,1})^2} + \frac{(\theta_{2,0}-\phi_{1,1,1}+\theta_{1,0}\phi_{2,2,1})(1-\phi_{1,1,1})-(\theta_{2,1}-2\theta_{2,1}\phi_{1,1,1}+\theta_{2,2,1}\phi_{1,1,1}-\phi_{2,2,1})}{(1-\phi_{1,1,1}-\phi_{2,2,1}+\phi_{1,1,1}\phi_{2,2,1}-\phi_{1,2,1})^2}
\]

\[
\frac{\partial \mu_{1,1}}{\partial \theta_{1,0}} = 0
\]

\[
\frac{\partial \mu_{1,1}}{\partial \theta_{2,0}} = 0
\]

\[
\frac{\partial \mu_{1,1}}{\partial \theta_{1,1}} = \frac{1-\phi_{2,2,1}}{1-\phi_{1,1,1}-\phi_{2,2,1}+\phi_{1,1,1}\phi_{2,2,1}-\phi_{1,2,1}\phi_{2,1,1}}
\]

\[
\frac{\partial \mu_{1,1}}{\partial \theta_{2,1}} = \frac{\phi_{1,2,1}}{1-\phi_{1,1,1}-\phi_{2,2,1}+\phi_{1,1,1}\phi_{2,2,1}-\phi_{1,2,1}\phi_{2,1,1}}
\]

\[
\frac{\partial \mu_{1,1}}{\partial \phi_{1,1,1}} = \frac{(\theta_{1,1}+\phi_{1,2,1}\phi_{2,2,1}+\theta_{2,1}\phi_{1,1,1}\phi_{2,2,1}+\theta_{1,2,1}\phi_{2,2,1})}{(1-\phi_{1,1,1}-\phi_{2,2,1}+\phi_{1,1,1}\phi_{2,2,1}-\phi_{1,2,1}\phi_{2,1,1})^2}
\]

\[
\frac{\partial \mu_{1,1}}{\partial \phi_{2,2,1}} = \frac{(\theta_{2,1}-\theta_{2,1}\phi_{1,1,1}-\theta_{2,2,1}\phi_{2,2,1})}{(1-\phi_{1,1,1}-\phi_{2,2,1}+\phi_{1,1,1}\phi_{2,2,1}-\phi_{1,2,1}\phi_{2,1,1})^2}
\]

\[
\frac{\partial \mu_{1,1}}{\partial \phi_{2,2,1}} = \frac{(-\phi_{1,1,1}+\theta_{1,1}\phi_{1,1,1}+\theta_{1,1}\phi_{2,2,1}-\theta_{1,1}\phi_{1,1,1}\phi_{2,2,1}+\theta_{1,1}\phi_{2,2,1})}{(1-\phi_{1,1,1}-\phi_{2,2,1}+\phi_{1,1,1}\phi_{2,2,1}-\phi_{1,2,1}\phi_{2,1,1})^2}
\]

\[
\frac{\partial \mu_{2,1}}{\partial \theta_{1,0}} = 0
\]
After estimating the parameters of $\psi$, these parameter estimates are entered into the equations and partial derivatives above in order to obtain the estimated means and standard errors of the estimates for the parameters of $\mu$, which are the large scale deterministic time trend parameters.

### 2.3 Numerical Derivative Approximation

As has been seen in the previous section, the calculation of the exact partial derivatives necessary to utilize the multivariate delta method can be quite intensive, even for the relatively simple bivariate ($n = 2$) VAR(1) ($p = 1$) case. For a model with $n > 2$ or $p >$...
1, or both, the process will become even more complicated. Because of this, it is appropriate
to consider numerical derivative approximations.

By using centered differencing, the estimated values of the derivatives can be
approximated as shown by Levy (2010). Consider (2.2.1) which gives $\mu_{1,0}$, the first element
of $\mu$, as a function of the elements of $\psi$. As an example, the first element of $\psi$ is $\theta_{1,0}$. By
centered differencing, as $h$ approaches zero for $h > 0$,

$$
\frac{\partial \mu_{1,0}(\theta_{1,0})}{\partial \theta_{1,0}} \approx \frac{\mu_{1,0}(\theta_{1,0} + h) - \mu_{1,0}(\theta_{1,0} - h)}{2h}
$$

(2.3.1)

where all other elements of $\psi$ are held constant. After obtaining $\hat{\psi}$, the parameter estimates
are entered into (2.3.1) to estimate $\frac{\partial \mu_{1,0}}{\partial \theta_{1,0}}$. This is done similarly to estimate all necessary
partial derivatives.

Estimating the numerical values of the partial derivatives using the elements of $\hat{\psi}$ in
such a way is much simpler and more time efficient than calculating the exact partial
derivatives, especially for more complicated models with $n > 2$ or $p > 1$. In addition, the
numerical approximation also further simplifies the process because it allows for the $\mu$
and $\psi$ parameters to be considered in vector and matrix notation for all $n$ and all $p$ as in (1.5.6)
as opposed to solving for the parameters of $\mu$ individually as in Equations 2.2.1 through
2.2.4.

2.4 Simulation

In order to test the ability of the model that has been described to identify the $\mu$
parameters, data that follow the pattern of a multivariate time series with multiple
temporal scales, assuming normally distributed error terms, were simulated. The model
outlined previously was fit to the data in order to examine its effectiveness. Following the
calculation of $\hat{\mu}_0$ and $\hat{\mu}_1$, the multivariate delta method was utilized to approximate the
distribution of the estimates.

The exact partial derivatives were first calculated in order to estimate the
covariance matrices of the estimates. However, as discussed in the previous section,
because the calculation of these derivatives may be quite complicated even for the simplest
form of the model, numerical methods were also used to approximate the derivatives. The
estimated covariance matrices resulting from using the exact derivatives and from the
numerical approximations were compared to ensure that simplifying the calculations
through the use of numerical approximation is a valid substitution.

After verifying the validity of using numerical approximations for the partial
derivatives in order to estimate the distribution of the large scale time trend parameters,
the simulation was then repeated numerous times. By simulating the data 1000 times for
varying $\mu$ parameter values, the type I error rate and power for hypothesis testing on these
large scale deterministic time trend parameters using the suggested model and estimation
methods were approximated. The results were then compared to the type I error rate and
power estimates resulting from hypothesis testing on these same parameters by instead
using non-overlapping averages, which is the other most commonly used method of viewing
time series on multiple temporal scales as discussed in the previous chapter.

In keeping with simplicity, as described in Section 2.1, consider the bivariate ($n = 2$)
VAR(1) ($p = 1$) case for both the small and large scale. Thus the notation, the model, the
assumptions, and the parameters used in the simulation will be identical to those described
in the previous sections of this chapter. In order to first compare the estimated covariance
matrices resulting using the exact partial derivatives and the numerical approximation, the
following parameter values were used:
\( T = 100 \)

\[
\begin{bmatrix}
\mu_0 \\
\mu_1
\end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix}
\Phi_1 \\
\Sigma_1
\end{bmatrix} = \begin{bmatrix} 0.7 & -0.2 \\ 0.2 & 0.7 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
\]

\( S = 100 \)

\[
\begin{bmatrix}
\Gamma_{1,t} \\
\Sigma_t
\end{bmatrix} = \begin{bmatrix} 0.5 & -0.2 \\ 0.2 & 0.5 \end{bmatrix}, \quad \Sigma_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ for all } t \text{ (} t = 1 \text{ to } 100 \text{)}
\]

In addition to the error term assumptions, it is also seen that these selections of \( \Phi_1 \) and \( \Gamma_{1,t} \) do indeed produce stable vector autoregressive processes as the modulus of the eigenvalues of \( \Phi_1 \) are both equal to 0.728011 and the modulus of the eigenvalues of \( \Gamma_{1,t} \) are both equal to 0.538517. Because each of these eigenvalues have modulus less than one, the requirement of stability is met as described in Section 1.4 which will allow for valid parameter estimation under the assumption of stationarity.

In this situation, the parameters of \( \mu_1 \), which describe the large scale linear time trend, are of greatest interest and so it is the type I error rate and power for hypothesis testing on these parameters that is investigated. To do so, the values of \( \mu_1 \) were varied while all other parameter values were held constant at the values given above and 1000 simulations were run for each unique value of \( \mu_1 \). For the given set of hypotheses, the proportion of rejections was used to estimate the type I error rate when the true value of \( \mu_1 \) used in the simulation matched that of the null hypothesis. The proportion of rejections was also used to estimate the power for various differences when the true value of \( \mu_1 \) used in the simulation did not match that of the null hypothesis. This was done using both the method proposed here as well as the method of utilizing non-overlapping averages in order to make comparisons. The values of \( \mu_1 \) used in these simulations and the results are discussed in the next section.
2.5 Simulation Results

The results from the initial simulation show that using numerical approximations to the partial derivatives is in fact a valid substitution. The matrix of partial derivatives, \( \mathbf{D} \), as described in (2.2.5) was estimated using the elements of \( \hat{\mathbf{\psi}} \). The estimation was done first by using the calculations of the exact partial derivatives, given in Section 2.2. Refer to this first estimate of \( \mathbf{D} \) as \( \hat{\mathbf{D}}_{\text{exact}} \). \( \mathbf{D} \) was then also estimated by the numerical approximation method with centered differencing, as shown in (2.3.1), again using the elements of \( \hat{\mathbf{\psi}} \). Refer to this second estimate of \( \mathbf{D} \) as \( \hat{\mathbf{D}}_{\text{approx}} \). For the numerical approximation with centered differencing, as seen in (2.3.1), it was decided to set \( h = 1 \times 10^{-5} \). The resulting estimates are given below.

\[
\hat{\mathbf{D}}_{\text{exact}} = \begin{bmatrix}
2.285430 & -1.715525 & -1.289059 & 5.900402 & 1.855064 & -1.409440 & 1.608399 & -1.170721 \\
0.961048 & 2.153985 & -3.305444 & -0.836965 & 0.789577 & 1.747072 & 0.655846 & 1.518697 \\
0 & 0 & 2.285430 & -1.715525 & -0.007859 & 0.005899 & 0.016957 & -0.012729 \\
0 & 0 & 0.961048 & 2.153985 & -0.003305 & -0.007407 & 0.007131 & 0.015982 \\
\end{bmatrix}
\]

\[
\hat{\mathbf{D}}_{\text{approx}} = \begin{bmatrix}
2.285430 & -1.715525 & -1.289059 & 5.900402 & 1.855064 & -1.409440 & 1.608399 & -1.170721 \\
0.961048 & 2.153985 & -3.305444 & -0.836965 & 0.789577 & 1.747072 & 0.655846 & 1.518697 \\
0 & 0 & 2.285430 & -1.715525 & -0.007859 & 0.005899 & 0.016957 & -0.012729 \\
0 & 0 & 0.961048 & 2.153985 & -0.003305 & -0.007407 & 0.007131 & 0.015982 \\
\end{bmatrix}
\]

As is seen from these estimates, with \( h \) set at \( 1 \times 10^{-5} \), using the numerical approximations of the partial derivatives with centered differencing and the elements of \( \hat{\mathbf{\psi}} \) to estimate \( \mathbf{D} \) yields results that are identical to estimating \( \mathbf{D} \) with \( \hat{\mathbf{\psi}} \) and the exact partial derivatives to the \( 1 \times 10^{-6} \) decimal place in this situation. Because of the lack of difference between the estimates of the two methods, and because the numerical approximations are much easier to implement, the numerical approximation method will be used to utilize the multivariate delta method going forward.
As described in the previous section, the values of $\mu_1$ were then varied and 1000 simulations were run for each unique $\mu_1$ with all other parameter values held constant in order to estimate the type I error rate and power of hypothesis testing on the $\mu_1$ parameters. Because $\mu_1$ consists of two parameters, two null hypotheses were tested. The pair of hypotheses being tested is as follows:

$$H_{01}: \mu_{1,1} = 0 \text{ vs. } H_{a1}: \mu_{1,1} \neq 0$$

$$H_{02}: \mu_{2,1} = 0 \text{ vs. } H_{a2}: \mu_{2,1} \neq 0$$

Each hypothesis test was conducted using a t-test due to the fact that $\bar{\mu} \sim N(\mu, D\Omega D^T)$, as explained in Section 2.2, at a level of $\alpha = 0.05$. Because the two null hypotheses are being tested simultaneously, the Bonferroni adjustment was applied. If either $H_{01}$ or $H_{02}$ was rejected, the pair of null hypotheses was rejected.

After repeating the process through each of the 1000 simulations where $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, the proportion of times that the pair of null hypotheses was rejected was used to estimate the type I error rate for hypothesis testing on $\mu_1$. The same was then done for the varying values of $\mu_1$ to estimate the power. This was first done by using the modeling process described in Section 2.1. The results are found in the tables below.

<table>
<thead>
<tr>
<th>$\mu_{1,1}$</th>
<th>$\mu_{1,2}$</th>
<th>Proportion Rejected</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.00</td>
<td>0.042</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\mu_{1,1}$</th>
<th>$\mu_{1,2}$</th>
<th>Proportion Rejected</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.01</td>
<td>0.277</td>
</tr>
<tr>
<td>0.02</td>
<td>0.02</td>
<td>0.720</td>
</tr>
</tbody>
</table>
In order to compare the proposed method of modeling observations on the small
time scale at each large scale time point and modeling the resulting estimates on the larger
time scale to the method of using non-overlapping averages, the same process of simulating
data was carried out. Again, the proportion of times that the pair of hypotheses was rejected
was used to estimate type I error rate and the power. All parameter values used in the
simulations remained the same, with the values of \( \mu_1 \) again varying for power estimation
but with all other parameter values held constant. The methods of parameter and
distribution estimation also remain the same, with the only exception being that each \( \hat{y}_t \) is
now obtained by taking the arithmetic mean of the \( z_{st} \)'s rather than through the use of the
model as shown in (2.1.5). These results are found in the tables below.

| Table 2.5.3 Estimated Type I Error Rate with Non-overlapping Averages, \( T = S = 100 \) |
|---------------------------------|---------------------------------|-----------------|
| \( \mu_{1,1} \)                | \( \mu_{1,2} \)                | Proportion Rejected |
| 0.00                           | 0.00                           | 0.043            |

| Table 2.5.4 Estimated Power with Non-overlapping Averages, \( T = S = 100 \) |
|---------------------------------|---------------------------------|-----------------|
| \( \mu_{1,1} \)                | \( \mu_{1,2} \)                | Proportion Rejected |
| 0.01                           | 0.01                           | 0.277            |
| 0.02                           | 0.02                           | 0.721            |
| 0.03                           | 0.03                           | 0.960            |
| 0.04                           | 0.04                           | 1.000            |
| 0.05                           | 0.05                           | 1.000            |

From these results it is seen that the proposed method of modeling observations on
the small observable time scale at each large scale time point to obtain estimates of the
larger scale behavior of the time series performs comparably to using non-overlapping
averages to examine the behavior of the time series on a larger time scale in terms of identifying the large scale time trends of the data. Thus, in addition to allowing for a closer analysis of the small scale behavior of the data from large scale time point to large scale time point by providing more information than a simple average, modeling the data on both scales is also a valid alternative to using non-overlapping averages to test the large scale time trend parameters.

The simulations were then re-run with smaller sample sizes to better approximate the sample sizes likely to be seen in the context of the weight lifting example. This time, both $T$ and $S$ were set to 20 while all other parameters were held constant. The parameters of $\mu_1$ were again varied appropriately in order to estimate power. The hypotheses and alpha level also remained the same. The resulting type I error rate and power estimates from the simulations using the method of modeling observations on the small scale are given in the first two tables below, followed by the results of using non-overlapping averages.

<table>
<thead>
<tr>
<th>$\mu_{1,1}$</th>
<th>$\mu_{1,2}$</th>
<th>Proportion Rejected</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.00</td>
<td>0.091</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\mu_{1,1}$</th>
<th>$\mu_{1,2}$</th>
<th>Proportion Rejected</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.05</td>
<td>0.209</td>
</tr>
<tr>
<td>0.10</td>
<td>0.10</td>
<td>0.374</td>
</tr>
<tr>
<td>0.15</td>
<td>0.15</td>
<td>0.563</td>
</tr>
<tr>
<td>0.20</td>
<td>0.20</td>
<td>0.716</td>
</tr>
<tr>
<td>0.25</td>
<td>0.25</td>
<td>0.820</td>
</tr>
<tr>
<td>0.30</td>
<td>0.30</td>
<td>0.885</td>
</tr>
<tr>
<td>0.35</td>
<td>0.35</td>
<td>0.929</td>
</tr>
<tr>
<td>0.40</td>
<td>0.40</td>
<td>0.954</td>
</tr>
<tr>
<td>0.45</td>
<td>0.45</td>
<td>0.967</td>
</tr>
<tr>
<td>0.50</td>
<td>0.50</td>
<td>0.978</td>
</tr>
</tbody>
</table>
Again, the results show that when hypothesis testing on the large scale time trend parameters the method of modeling observations on the small scale separately at each large scale time point in order to obtain estimates of the large scale behavior of the data performs comparably to using non-overlapping averages, with the additional benefit of providing more information on the behavior of the data on the small scale as it changes from one large scale time point to the next. It is also seen that for the smaller sample sizes, the proposed method of modeling on both scales results in an inflated type I error rate. However, the popular method of using non-overlapping averages also resulted in an inflated type I error rate which was comparable to, and in fact slightly higher than, that of the proposed method.
CHAPTER 3

3.1 Application of Model to Weight Lifting Data

In the previous chapter, a model was described in Section 2.1, namely through Equations 2.1.4, 2.1.5, and 2.1.6. That model was fit to observations made on a bivariate time series on two time scales; first on the small scale at each large scale time point and then the estimates resulting from this modeling were modeled on the large scale to successfully estimate large scale time trends. Through simulations, it was also seen that using these methods of modeling the data on both time scales performed as well as using non-overlapping averages to estimate the large scale time trends of the data while providing more information about the behavior of the data on the small scale. In this chapter, the model is applied to the weight lifting data described previously, which is made up of two variables (average power and peak power) occurring on two time scales (the smaller rep-to-rep within day scale and the larger day-to-day scale).

The Section 2.1 equations that describe the model, referenced above, are again given below:

\begin{align*}
\mathbf{z}_{s,t} &= \mathbf{\alpha}_t + \mathbf{\Gamma}_1 \mathbf{z}_{s-1,t} + \mathbf{\omega}_{s,t} \\
\hat{\mathbf{y}}_t &= (\mathbf{I}_2 - \hat{\mathbf{F}}_{1,t})^{-1} \hat{\mathbf{\alpha}}_t \\
\mathbf{y}_t &= \mathbf{\theta}_0 + \mathbf{\theta}_1 t + \Phi_1 \mathbf{y}_{t-1} + \mathbf{\epsilon}_t
\end{align*}

where (3.1.1) refers to the modeling of the data on the small scale at each large scale time point \( t \) (\( t = 1 \) to \( T \)), (3.1.2) allows for the estimation of each \( \mathbf{y}_t \), and (3.1.3) models the data on the large scale. The assumptions of stability and the distribution of error terms remain the same as those outlined in Sections 2.1 and 1.4. After using (3.1.1) to model the data at each large scale time point and then using (3.1.2) to obtain each \( \hat{\mathbf{y}}_t \), these estimates for the
\(y_t\)'s are entered into (3.1.3) to model the data and obtain parameter estimates on the large scale. The underlying large scale time trend parameters, \(\mu_0\) and \(\mu_1\), are then estimated using (1.5.6) with standard errors being calculated through the use of the multivariate delta method using numerical approximation for partial derivatives, as was discussed in Chapter 2.

As mentioned in Chapter 1, when collecting the weight lifting data during workouts, each athlete follows a personalized workout plan and each exercise may have a different objective so that each athlete and exercise need to be considered separately. In order to see how the model performs, the data collected from the athletes were de-identified by replacing each athlete's name with an identification number. Using a random number generator, one athlete (11) was selected at random. The model was then fit to the data collected from athlete 11 while performing bench press lifts over the course of the 2012 winter training period.

For simplicity, in Chapter 2 it was assumed that \(s = 1\) to \(S\) for all \(t = 1\) to \(T\) where \(S\) was the number of small scale observations at large scale time \(t\) and \(T\) was the number of large scale observations. Thus, it was assumed that \(S\) was constant for all \(t (t = 1\) to \(T)\) and so there were \(S\) small scale observations at each large scale time \(t\). In the context of the weight lifting data, \(S\) will refer to the number of reps occurring on a given day and \(T\) will refer to the number of days. However, the number of reps an athlete completes may differ from day to day. Let \(T\) still be the number of days for which data has been recorded, but let the number of reps occurring on day \(t\) be \(S_t\) so that \(S_t\) may differ for each \(t (t = 1\) to \(T)\). Using this notation, we now have \(t = 1\) to \(T\) and \(s = 1\) to \(S_t\) for day \(t\).

In the case of the bench press data collected from athlete 11 \(T = 7, S_1 = 32, S_2 = 24, S_3 = 16, S_4 = 23, S_5 = 16, S_6 = 10,\) and \(S_7 = 12\). To ensure that the model will perform adequately for sample sizes such as these, 1000 simulations were again run as described in
Section 2.4 using the same parameter values, the only exceptions being the sample sizes, which were this time set to \( T = 10 \) and \( S = 20 \) to better approximate the sample sizes seen here. As before, the evaluation was done using both the approach of modeling the observations on the small scale to obtain the large scale time point estimates and also using non-overlapping averages. For both methods, the type I error rate and power for hypothesis testing on the large scale time trend parameters were estimated as in Chapter 2 and the results are given in the tables below.

| Table 3.1.1 Estimated Type I Error Rate with Modeling on Small Scale, \( T = 10, S = 20 \) |
|---------------------------------|-----------------|------------------|
| \( \mu_{1,1} \) | \( \mu_{1,2} \) | Proportion Rejected |
| 0.00 | 0.00 | 0.078 |

| Table 3.1.2 Estimated Power with Modeling on Small Scale, \( T = 10, S = 20 \) |
|---------------------------------|-----------------|------------------|
| \( \mu_{1,1} \) | \( \mu_{1,2} \) | Proportion Rejected |
| 0.1 | 0.1 | 0.159 |
| 0.2 | 0.2 | 0.279 |
| 0.3 | 0.3 | 0.379 |
| 0.4 | 0.4 | 0.497 |
| 0.5 | 0.5 | 0.594 |
| 0.6 | 0.6 | 0.668 |
| 0.7 | 0.7 | 0.743 |
| 0.8 | 0.8 | 0.788 |
| 0.9 | 0.9 | 0.814 |
| 1.0 | 1.0 | 0.844 |
| 1.1 | 1.1 | 0.862 |
| 1.2 | 1.2 | 0.882 |
| 1.3 | 1.3 | 0.889 |
| 1.4 | 1.4 | 0.895 |
| 1.5 | 1.5 | 0.900 |

| Table 3.1.3 Estimated Type I Error Rate with Non-overlapping Averages, \( T = 10, S = 20 \) |
|---------------------------------|-----------------|------------------|
| \( \mu_{1,1} \) | \( \mu_{1,2} \) | Proportion Rejected |
| 0.00 | 0.00 | 0.082 |
Again it is seen that the proposed method of modeling observations on the small scale to estimate the behavior of the data at each large scale time point performs comparably to using non-overlapping averages, even for this smaller sample size. Also, the type I error rates for both methods are again inflated. However, as was the case with the simulation results given in Section 2.5, the proposed method’s rate is slightly lower. The proposed method was then applied to the selected bench press data collected from athlete 11 over the 2012 winter training period. The figure below shows a plot of the raw data, the estimated daily behavior of the data, and the estimated time trends of the data. In the context of the model given by Equations 3.1.1, 3.1.2, and 3.1.3, Average Power and Peak Power represent the bivariate observations $z_{s,t}$, Estimated Daily Average Power and Estimated Daily Peak Power represent $\hat{y}_t$, and Average Power Trend and Peak Power Trend represent the underlying large scale deterministic time trends, $\mu_t$, as first described in Section 1.5.

<table>
<thead>
<tr>
<th>$\mu_{1,1}$</th>
<th>$\mu_{1,2}$</th>
<th>Proportion Rejected</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.167</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>0.276</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3</td>
<td>0.392</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4</td>
<td>0.513</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0.589</td>
</tr>
<tr>
<td>0.6</td>
<td>0.6</td>
<td>0.679</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7</td>
<td>0.735</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
<td>0.783</td>
</tr>
<tr>
<td>0.9</td>
<td>0.9</td>
<td>0.810</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>0.840</td>
</tr>
<tr>
<td>1.1</td>
<td>1.1</td>
<td>0.859</td>
</tr>
<tr>
<td>1.2</td>
<td>1.2</td>
<td>0.877</td>
</tr>
<tr>
<td>1.3</td>
<td>1.3</td>
<td>0.886</td>
</tr>
<tr>
<td>1.4</td>
<td>1.4</td>
<td>0.899</td>
</tr>
<tr>
<td>1.5</td>
<td>1.5</td>
<td>0.906</td>
</tr>
</tbody>
</table>

Table 3.1.4  Estimated Power with Non-overlapping Averages, $T = 10, S = 20$
Figure 3.1.1 Plot of Athlete 11 Bench Press Data, Daily Estimates, and Trend Estimates

The following table gives the parameter estimates as well as the standard errors for those estimates.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Power Intercept ($\theta_{1,0}$)</td>
<td>1.3223</td>
<td>0.0976</td>
</tr>
<tr>
<td>Peak Power Intercept ($\theta_{2,0}$)</td>
<td>0.7537</td>
<td>0.3368</td>
</tr>
<tr>
<td>Average Power Linear Time Trend ($\theta_{1,1}$)</td>
<td>-0.0236</td>
<td>0.0105</td>
</tr>
<tr>
<td>Peak Power Linear Time Trend ($\theta_{2,1}$)</td>
<td>-0.0820</td>
<td>0.0361</td>
</tr>
<tr>
<td>Autoregressive Effect: Average Power on Average Power ($\phi_{1,1,1}$)</td>
<td>0.0371</td>
<td>0.1283</td>
</tr>
<tr>
<td>Autoregressive Effect: Average Power on Peak Power ($\phi_{2,1,1}$)</td>
<td>0.7511</td>
<td>0.4427</td>
</tr>
<tr>
<td>Autoregressive Effect: Peak Power on Average Power ($\phi_{1,2,1}$)</td>
<td>-1.2840</td>
<td>0.1646</td>
</tr>
<tr>
<td>Autoregressive Effect: Peak Power on Peak Power ($\phi_{2,2,1}$)</td>
<td>-0.8072</td>
<td>0.5679</td>
</tr>
<tr>
<td>Average Power Underlying Deterministic Intercept ($\mu_{1,0}$)</td>
<td>0.5165</td>
<td>0.0501</td>
</tr>
<tr>
<td>Peak Power Underlying Deterministic Intercept ($\mu_{2,0}$)</td>
<td>0.6061</td>
<td>0.0221</td>
</tr>
<tr>
<td>Average Power Underlying Deterministic Linear Time Trend ($\mu_{1,1}$)</td>
<td>0.0232</td>
<td>0.0108</td>
</tr>
<tr>
<td>Peak Power Underlying Deterministic Linear Time Trend ($\mu_{2,1}$)</td>
<td>-0.0357</td>
<td>0.0059</td>
</tr>
</tbody>
</table>
By examining the plot and the resulting parameter estimates, it appears that the model does in fact produce appropriate results for the bench press weight lifting data for this athlete.

3.2 Small Scale Linear Time Trends

So far, in both Chapter 2 and in the previous section, the models of small scale observations at each large scale time point \( t \) as given by Equation 3.1.1 have contained only an intercept term in addition to the stochastic autoregressive term. This was done in Section 2.1 in order to first introduce the model in its simplest form: a bivariate VAR(1) model occurring on two time scales with an intercept term on the small scale that varies between each large scale time point \( t \) and linear deterministic time trend on the large scale. It also allowed for a more straightforward comparison with the method of using non-overlapping averages as that method does not allow for any modeling of the time series on the smaller scale from one large scale time point to the next. In addition, the model was left unchanged in the previous section to demonstrate that in its most basic form, it was able to successfully model the weight lifting data.

However, in the case of the weight lifting data, and in general, it is appropriate to also include a linear time trend parameter in the small scale models at each large scale time point. In the context of the weight lifting example, it seems intuitive that while completing several reps within a given a day an athlete would begin to fatigue, resulting in a gradual decrease in both average and peak power. Thus, even with improvement in performance over the course of the winter training session as indicated by positive linear time trends on the day-to-day scale, it is possible and even expected to see negative linear time trends on the rep-to-rep scale within each day. These negative within rep trends are in fact seen in
Figure 3.1.1 given above, as it appears that average power is exhibiting a slight downward trend within some days while still showing an upward trend across days.

A linear time trend parameter is now added to the small scale models at each large scale time point \( t \), which will result in some changes to the model as described in Section 2.1. The new model equations, adapted from those in Section 2.1, are given below.

\[
\begin{align*}
  z_{s,t} &= y_{t(s)} + w_{s,t} \quad \text{where} \\
  y_{t(s)} &= y_{0,t} + y_{1,t} S \quad \text{and} \\
  w_{s,t} &= \Gamma_{1,t} w_{s-1,t} + \omega_{s,t} \\
  \end{align*}
\]

(3.2.1)

\[
\begin{align*}
  z_{s,t} &= y_{0,t} + y_{1,t} S + \Gamma_{1,t} w_{s-1,t} + \omega_{s,t} \quad \text{or} \\
  z_{s,t} &= \alpha_{0,t} + \alpha_{1,t} S + \Gamma_{1,t} z_{s-1,t} + \omega_{s,t}
\end{align*}
\]

(3.2.4)

(3.2.5)

The modeling has been done as described in Section 1.5 for multivariate time series with deterministic time trends. All parameter descriptions and assumptions remain the same as those given in Section 2.1, with the only change being the addition of the linear time trends. Parameter estimates for the model given by (3.2.5) are obtained as described and demonstrated previously (i.e. ordinary least squares estimation which is equivalent to maximum likelihood estimation under the necessary assumptions). After obtaining these parameter estimates, the estimates for \( y_{0,t} \) and \( y_{1,t} \) are also found for each \( t \) by using the following set of equations as shown in Section 1.5:

\[
\begin{align*}
  \hat{y}_{0,t} &= (I_2 - \tilde{\Gamma}_{1,t})^{-1} \hat{\alpha}_{0,t} - (I_2 - \tilde{\Gamma}_{1,t})^{-1} \tilde{\Gamma}_{1,t} (I_2 - \tilde{\Gamma}_{1,t})^{-1} \hat{\alpha}_{1,t} \\
  \hat{y}_{1,t} &= (I_2 - \tilde{\Gamma}_{1,t})^{-1} \hat{\alpha}_{1,t}
\end{align*}
\]

(3.2.6)
3.3 Modeling with Small Scale Linear Time Trends

By now including a linear time trend in the small scale models at each large scale time point $t$ as described in the previous section, more information is obtainable in the case of the weight lifting data. There is now not only an intercept that varies across days being estimated from the small scale observations at each large scale time point, but also a linear time trend that varies across days.

Previously, the non-autoregressive portion of the small scale models, $y_t$, was strictly an intercept as described by Equation 2.1.1. Thus, estimation of this portion of the model resulted in only one estimate, $\hat{y}_0$, for each $t$ which was then entered into the model given by (2.1.6) in order to model the behavior and estimate the trends of the data on the large scale. Now, the non-autoregressive portion of the small scale models, $y_t^{(s)}$, contains both an intercept term, $y_0,t$, and the linear time trend term, $y_1,t$, as shown in Equation 3.2.2. Thus, estimating the non-autoregressive portion of the small scale models results in two estimates, both $\hat{y}_{0,t}$ and $\hat{y}_{1,t}$, as given by Equation 3.2.6, for each large scale time $t$.

Because $\hat{y}_{0,t}$ gives the estimated intercept for the small scale observations at each large scale time point $t$, these will be used as the summaries of the behavior of the data at each large scale time point and will be modeled over the large scale to examine the large scale trends of the data. In the context of the weight lifting problem, modeling these intercept estimates helps describe the change in the athletes’ strength over time by determining whether their average and peak power measures seem to follow a positive or negative trend over the course of the winter training period. Just as the $\hat{y}_t$’s were entered into Equation 2.1.6 in Chapter 2 and Equation 3.1.3 in Section 3.1 to model the behavior and trends of the data on the large scale, the $\hat{y}_{0,t}$’s are now entered into the equation given below in order to accomplish this estimation.
As was done in previous sections, the parameter estimates obtained for this model are then used to estimate the underlying large scale deterministic time trends of the average and peak power variables, now given by $\mu_{0,0}$ and $\mu_{0,1}$ as seen in Section 1.5.

In addition to entering the $\hat{y}_{0,t}$'s in the model given by (3.3.1) to look at the improvements in strength by examining the trends of average and peak power for each athlete, it is also useful to consider modeling the $\hat{y}_{1,t}$'s. As explained in the previous section, it makes sense for the reps the athletes complete within each day to exhibit a negative linear time trend as the athletes begin to fatigue. Because of interest in the athletes' fatigue, the estimates of the small scale linear time trends for each $t$ (i.e. the $\hat{y}_{1,t}$'s) may give a helpful measure of the athletes' conditioning. Negative slope values of higher magnitude indicate quicker fatigue and poorer conditioning, while negative slope values of lower magnitude (or nonnegative values) indicate slower (or no) fatigue and better conditioning. By modeling the $\hat{y}_{1,t}$'s over time, similar to the way the $\hat{y}_{0,t}$'s are modeled using (3.3.1), the large scale time trends of the athletes' conditioning are also estimated. Thus, both the strength and conditioning improvements of the athletes' are examined. Such modeling is done using the equation below.

$$y_{1,t} = \theta_{1,0} + \theta_{1,1} t + \Phi_{1,1} y_{1,t-1} + \epsilon_{1,t}$$  \hspace{1cm} (3.3.2)

Again, the parameter estimates obtained for this model are then used to estimate the underlying large scale deterministic time trends, now of the within day trends of the average and peak power variables, given by $\mu_{1,0}$ and $\mu_{1,1}$ similar to Section 1.5.
By creating a model that now includes a linear time trend parameter for the small scale models at each large scale time point, this modeling process will not only fit the true behavior of the data better, but also give another measure of the behavior of the data on the small scale from one large scale time point to the next. In the case of the weight lifting problem, allowing for both strength and conditioning evaluation.

3.4 Model Application to Weight Lifting Data

The model described in Sections 3.2 and 3.3, which now includes a linear time trend term for each of the small scale models at each large scale time point \( t \), is fit to the weight lifting data. Because bench press is the exercise that the athletes completed most frequently, the data collected during the bench press lift is used. As mentioned previously, each athlete completes an individually specialized weight lifting program so the model is fit to the weight lifting data of each athlete separately. For now, as in previous sections, only a VAR(1) model is considered for the small scale models at each large scale time point \( t \), and for the large scale models, for simplicity.

After the final estimation of (3.3.1), which refers to the model of the intercepts used to evaluate the large scale trends of an athlete’s strength, and (3.3.2), which refers to the model of slopes used to evaluate the large scale trends of an athlete’s conditioning, as explained in section 3.3, the equations given by (1.5.6) are used to estimate these large scale linear time trends for each, referred to now as \( \mu_{0,1} \) and \( \mu_{1,1} \) respectively, with the standard errors of these estimates being found by using the multivariate delta method with numerical approximations for the partial derivatives as has been previously discussed. These equations used for the estimation of \( \mu_{0,1} \) and \( \mu_{1,1} \) in the VAR(1) case are given below.

\[
\hat{\mu}_{0,1} = (I_2 - \tilde{\Phi}_{0,1})^{-1} \tilde{\theta}_{0,1}
\]  

(3.4.1)
These steps are followed for the bench press weight lifting data collected from each athlete. The relevant results, which are the estimates of these linear time trend parameters that describe the improvements in both strength and conditioning for each athlete, as well as the standard errors of these estimates, are reported.

3.5 Weight Lifting Results

There were a total of 89 athletes for which the weight lifting data was collected. Of those 89, 73 of the athletes had data collected over six days or more. Because of the number of parameters being fit to the models, only these 73 athletes with at least six days of data were considered. In addition, the model described by Sections 3.2, 3.3, and 3.4 that is being fit to the weight lifting data is still based upon the assumptions outlined in previous chapters. Along with the assumptions on the distributions of the error terms, another assumption is that there are no problems with a lack of stability as described in Section 1.4. After examining the models and the corresponding plots (as illustrated in Figure 3.1.1) for the 73 athletes, 50 athletes with data exhibiting stability and thus appropriate estimates were selected. Methods of dealing with lack of stability will be discussed later.

As explained in the previous section, the most relevant results are the estimates for $\mu_{0,1}$ and $\mu_{1,1}$, as well as the standard errors for those estimates. The results for the large scale (day-to-day) strength improvement trends ($\mu_{0,1}$) are given first in Table 3.5.1 below, where $\mu_{1,0,1}$ is the large scale linear time trend for average power and $\mu_{2,0,1}$ is the large scale linear time trend for peak power.
<table>
<thead>
<tr>
<th>Athlete</th>
<th>Average Power Strength Trend Estimate ($\mu_{1,0,1}$)</th>
<th>$\hat{\mu}_{1,0,1}$ Std. Error</th>
<th>Peak Power Strength Trend Estimate ($\mu_{2,0,1}$)</th>
<th>$\hat{\mu}_{2,0,1}$ Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0032</td>
<td>0.1411</td>
<td>0.0181</td>
<td>0.0554</td>
</tr>
<tr>
<td>2</td>
<td>0.2259</td>
<td>0.1347</td>
<td>0.0555</td>
<td>0.0264</td>
</tr>
<tr>
<td>3</td>
<td>0.0192</td>
<td>0.0118</td>
<td>0.0101</td>
<td>0.0179</td>
</tr>
<tr>
<td>4</td>
<td>0.0341</td>
<td>0.0994</td>
<td>-0.0884</td>
<td>0.0960</td>
</tr>
<tr>
<td>5</td>
<td>-0.0516</td>
<td>0.0949</td>
<td>0.0570</td>
<td>0.0192</td>
</tr>
<tr>
<td>6</td>
<td>-0.0483</td>
<td>0.0199</td>
<td>-0.0713</td>
<td>0.0592</td>
</tr>
<tr>
<td>7</td>
<td>0.0918</td>
<td>0.0505</td>
<td>-0.0007</td>
<td>0.0446</td>
</tr>
<tr>
<td>8</td>
<td>0.0151</td>
<td>0.0061</td>
<td>0.0391</td>
<td>0.0167</td>
</tr>
<tr>
<td>9</td>
<td>0.0237</td>
<td>0.0246</td>
<td>-0.0575</td>
<td>0.0694</td>
</tr>
<tr>
<td>10</td>
<td>-0.0242</td>
<td>0.0446</td>
<td>-0.0079</td>
<td>0.0592</td>
</tr>
<tr>
<td>11</td>
<td>0.0477</td>
<td>0.0126</td>
<td>0.0134</td>
<td>0.0089</td>
</tr>
<tr>
<td>12</td>
<td>0.1127</td>
<td>0.0201</td>
<td>-0.0116</td>
<td>0.0445</td>
</tr>
<tr>
<td>13</td>
<td>-0.0089</td>
<td>0.0113</td>
<td>-0.0058</td>
<td>0.0322</td>
</tr>
<tr>
<td>14</td>
<td>0.1925</td>
<td>0.0473</td>
<td>0.1711</td>
<td>0.0937</td>
</tr>
<tr>
<td>15</td>
<td>0.1867</td>
<td>0.1771</td>
<td>-0.0025</td>
<td>0.0157</td>
</tr>
<tr>
<td>16</td>
<td>-0.0642</td>
<td>0.1071</td>
<td>0.0019</td>
<td>0.0761</td>
</tr>
<tr>
<td>17</td>
<td>0.0694</td>
<td>0.0139</td>
<td>0.0440</td>
<td>0.0163</td>
</tr>
<tr>
<td>18</td>
<td>-0.0459</td>
<td>0.0393</td>
<td>-0.0191</td>
<td>0.0294</td>
</tr>
<tr>
<td>19</td>
<td>0.0104</td>
<td>0.0207</td>
<td>0.0005</td>
<td>0.0069</td>
</tr>
<tr>
<td>20</td>
<td>0.1295</td>
<td>0.2583</td>
<td>0.0026</td>
<td>0.3089</td>
</tr>
<tr>
<td>21</td>
<td>0.1016</td>
<td>0.1505</td>
<td>-0.0720</td>
<td>0.1717</td>
</tr>
<tr>
<td>22</td>
<td>-0.0277</td>
<td>0.1141</td>
<td>-0.2549</td>
<td>0.2132</td>
</tr>
<tr>
<td>23</td>
<td>0.0331</td>
<td>0.0465</td>
<td>-0.0151</td>
<td>0.0136</td>
</tr>
<tr>
<td>24</td>
<td>0.0150</td>
<td>0.0403</td>
<td>-0.0687</td>
<td>0.0585</td>
</tr>
<tr>
<td>25</td>
<td>-0.0849</td>
<td>0.0205</td>
<td>0.0084</td>
<td>0.0104</td>
</tr>
<tr>
<td>26</td>
<td>0.0614</td>
<td>0.0421</td>
<td>-0.0301</td>
<td>0.0513</td>
</tr>
<tr>
<td>27</td>
<td>-0.0235</td>
<td>0.0062</td>
<td>-0.0447</td>
<td>0.0026</td>
</tr>
<tr>
<td>28</td>
<td>0.0290</td>
<td>0.0402</td>
<td>0.0199</td>
<td>0.0181</td>
</tr>
<tr>
<td>29</td>
<td>-0.1718</td>
<td>0.1670</td>
<td>-0.1022</td>
<td>0.1378</td>
</tr>
<tr>
<td>30</td>
<td>0.0263</td>
<td>0.0017</td>
<td>-0.0026</td>
<td>0.0100</td>
</tr>
<tr>
<td>31</td>
<td>-0.0037</td>
<td>0.0159</td>
<td>-0.0055</td>
<td>0.0213</td>
</tr>
<tr>
<td>32</td>
<td>0.0104</td>
<td>0.0487</td>
<td>0.0180</td>
<td>0.0368</td>
</tr>
<tr>
<td>33</td>
<td>0.0116</td>
<td>0.1798</td>
<td>-0.0099</td>
<td>0.1311</td>
</tr>
<tr>
<td>34</td>
<td>-0.0055</td>
<td>0.0309</td>
<td>-0.0065</td>
<td>0.0030</td>
</tr>
<tr>
<td>35</td>
<td>0.0223</td>
<td>0.0620</td>
<td>0.0127</td>
<td>0.0363</td>
</tr>
<tr>
<td>36</td>
<td>0.0022</td>
<td>0.0294</td>
<td>-0.0046</td>
<td>0.0141</td>
</tr>
<tr>
<td>37</td>
<td>0.0275</td>
<td>0.0510</td>
<td>0.0255</td>
<td>0.0607</td>
</tr>
<tr>
<td>38</td>
<td>0.0286</td>
<td>0.0017</td>
<td>0.0134</td>
<td>0.0054</td>
</tr>
<tr>
<td>39</td>
<td>0.0412</td>
<td>0.0419</td>
<td>0.0344</td>
<td>0.0447</td>
</tr>
<tr>
<td>40</td>
<td>-0.0469</td>
<td>0.0054</td>
<td>0.0324</td>
<td>0.0290</td>
</tr>
<tr>
<td>41</td>
<td>0.0984</td>
<td>0.0634</td>
<td>0.0056</td>
<td>0.0181</td>
</tr>
</tbody>
</table>
The results for the large scale (day-to-day) conditioning improvement trends ($\mu_{1,1}$) are given in Table 3.5.2 below, where $\mu_{1,1}$ is the large scale linear time trend for average power slope and $\mu_{2,1,1}$ is the large scale linear time trend for peak power slope.

<table>
<thead>
<tr>
<th>Athlete</th>
<th>Average Power Conditioning Trend Estimate ($\bar{\mu}_{1,1}$)</th>
<th>$\bar{\mu}_{1,1}$ Std. Error</th>
<th>Peak Power Conditioning Trend Estimate ($\bar{\mu}_{2,1,1}$)</th>
<th>$\bar{\mu}_{2,1,1}$ Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0114</td>
<td>0.0012</td>
<td>-0.0042</td>
<td>0.0026</td>
</tr>
<tr>
<td>2</td>
<td>-0.0037</td>
<td>0.0028</td>
<td>-0.0004</td>
<td>0.0019</td>
</tr>
<tr>
<td>3</td>
<td>0.0026</td>
<td>0.0035</td>
<td>-0.0004</td>
<td>0.0028</td>
</tr>
<tr>
<td>4</td>
<td>-0.0013</td>
<td>0.0018</td>
<td>0.0032</td>
<td>0.0042</td>
</tr>
<tr>
<td>5</td>
<td>-0.0191</td>
<td>0.0603</td>
<td>-0.0082</td>
<td>0.0114</td>
</tr>
<tr>
<td>6</td>
<td>0.0018</td>
<td>0.0029</td>
<td>0.0095</td>
<td>0.0102</td>
</tr>
<tr>
<td>7</td>
<td>-0.0055</td>
<td>0.0027</td>
<td>0.0042</td>
<td>0.0015</td>
</tr>
<tr>
<td>8</td>
<td>-0.0007</td>
<td>0.0019</td>
<td>-0.0004</td>
<td>0.0053</td>
</tr>
<tr>
<td>9</td>
<td>-0.0065</td>
<td>0.0019</td>
<td>0.0057</td>
<td>0.0054</td>
</tr>
<tr>
<td>10</td>
<td>-0.0015</td>
<td>0.0026</td>
<td>0.0026</td>
<td>0.0020</td>
</tr>
<tr>
<td>11</td>
<td>-0.0030</td>
<td>0.0016</td>
<td>-0.0052</td>
<td>0.0016</td>
</tr>
<tr>
<td>12</td>
<td>-0.0158</td>
<td>0.0121</td>
<td>-0.0036</td>
<td>0.0067</td>
</tr>
<tr>
<td>13</td>
<td>0.0126</td>
<td>0.0422</td>
<td>-0.0001</td>
<td>0.0285</td>
</tr>
<tr>
<td>14</td>
<td>-0.0264</td>
<td>0.0021</td>
<td>-0.0191</td>
<td>0.0092</td>
</tr>
<tr>
<td>15</td>
<td>-0.0122</td>
<td>0.0068</td>
<td>-0.0003</td>
<td>0.0008</td>
</tr>
<tr>
<td>16</td>
<td>-0.0083</td>
<td>0.0183</td>
<td>-0.0043</td>
<td>0.0118</td>
</tr>
<tr>
<td>17</td>
<td>-0.0164</td>
<td>0.0031</td>
<td>-0.0120</td>
<td>0.0001</td>
</tr>
<tr>
<td>18</td>
<td>0.0048</td>
<td>0.0041</td>
<td>0.0022</td>
<td>0.0089</td>
</tr>
<tr>
<td>19</td>
<td>-0.0010</td>
<td>0.0012</td>
<td>0.0004</td>
<td>0.0009</td>
</tr>
<tr>
<td>20</td>
<td>-0.0057</td>
<td>0.0020</td>
<td>0.0112</td>
<td>0.0128</td>
</tr>
<tr>
<td>21</td>
<td>-0.0043</td>
<td>0.0015</td>
<td>-0.0020</td>
<td>0.0040</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>-0.0057</td>
<td>0.0137</td>
<td>-0.0082</td>
<td>0.0037</td>
</tr>
<tr>
<td>23</td>
<td>-0.0047</td>
<td>0.0010</td>
<td>-0.0018</td>
<td>0.0026</td>
</tr>
<tr>
<td>24</td>
<td>-0.0025</td>
<td>0.0015</td>
<td>-0.0023</td>
<td>0.0024</td>
</tr>
<tr>
<td>25</td>
<td>0.0121</td>
<td>0.0052</td>
<td>0.0003</td>
<td>0.0004</td>
</tr>
<tr>
<td>26</td>
<td>-0.0057</td>
<td>0.0092</td>
<td>-0.0009</td>
<td>0.0066</td>
</tr>
<tr>
<td>27</td>
<td>0.0007</td>
<td>0.0007</td>
<td>-0.0005</td>
<td>0.0004</td>
</tr>
<tr>
<td>28</td>
<td>-0.0060</td>
<td>0.0015</td>
<td>-0.0008</td>
<td>0.0010</td>
</tr>
<tr>
<td>29</td>
<td>0.0251</td>
<td>0.0302</td>
<td>0.0098</td>
<td>0.0103</td>
</tr>
<tr>
<td>30</td>
<td>-0.0057</td>
<td>0.0007</td>
<td>-0.0022</td>
<td>0.0052</td>
</tr>
<tr>
<td>31</td>
<td>0.0018</td>
<td>0.0014</td>
<td>-0.0015</td>
<td>0.0011</td>
</tr>
<tr>
<td>32</td>
<td>0.0041</td>
<td>0.0032</td>
<td>-0.0028</td>
<td>0.0034</td>
</tr>
<tr>
<td>33</td>
<td>0.0017</td>
<td>0.0030</td>
<td>0.0017</td>
<td>0.0005</td>
</tr>
<tr>
<td>34</td>
<td>-0.0014</td>
<td>0.0015</td>
<td>0.0015</td>
<td>0.0011</td>
</tr>
<tr>
<td>35</td>
<td>-0.0079</td>
<td>0.0097</td>
<td>-0.0057</td>
<td>0.0183</td>
</tr>
<tr>
<td>36</td>
<td>0.0018</td>
<td>0.0039</td>
<td>0.0010</td>
<td>0.0017</td>
</tr>
<tr>
<td>37</td>
<td>-0.0073</td>
<td>0.0095</td>
<td>-0.0018</td>
<td>0.0024</td>
</tr>
<tr>
<td>38</td>
<td>-0.0021</td>
<td>0.0011</td>
<td>-0.0004</td>
<td>0.0010</td>
</tr>
<tr>
<td>39</td>
<td>-0.0026</td>
<td>0.0003</td>
<td>-0.0002</td>
<td>0.0036</td>
</tr>
<tr>
<td>40</td>
<td>0.0039</td>
<td>0.0019</td>
<td>-0.0019</td>
<td>0.0017</td>
</tr>
<tr>
<td>41</td>
<td>-0.0110</td>
<td>0.0018</td>
<td>-0.0017</td>
<td>0.0017</td>
</tr>
<tr>
<td>42</td>
<td>0.0108</td>
<td>0.0100</td>
<td>-0.0059</td>
<td>0.0094</td>
</tr>
<tr>
<td>43</td>
<td>-0.0041</td>
<td>0.0013</td>
<td>-0.0018</td>
<td>0.0006</td>
</tr>
<tr>
<td>44</td>
<td>0.0271</td>
<td>0.1414</td>
<td>-0.0049</td>
<td>0.0247</td>
</tr>
<tr>
<td>45</td>
<td>-0.0024</td>
<td>0.0015</td>
<td>0.0004</td>
<td>0.0004</td>
</tr>
<tr>
<td>46</td>
<td>0.0036</td>
<td>0.0032</td>
<td>0.0031</td>
<td>0.0012</td>
</tr>
<tr>
<td>47</td>
<td>-0.0217</td>
<td>0.0162</td>
<td>0.0062</td>
<td>0.0123</td>
</tr>
<tr>
<td>48</td>
<td>-0.0024</td>
<td>0.0019</td>
<td>0.0027</td>
<td>0.0003</td>
</tr>
<tr>
<td>49</td>
<td>-0.0004</td>
<td>0.0004</td>
<td>0.0020</td>
<td>0.0004</td>
</tr>
<tr>
<td>50</td>
<td>-0.0032</td>
<td>0.0053</td>
<td>0.0017</td>
<td>0.0018</td>
</tr>
</tbody>
</table>

From these results, coaches now have a set of values that quantify each athlete's improvement in average power and peak power, in terms of both strength and conditioning, for the bench press. It is seen that 32 of the 50 athletes (64%) showed an improvement in average power strength over the course of the 2012 winter training period, as indicated by an estimated $\mu_{1,0,1}$ value that was greater than zero. Similarly, 23 of the 50 athletes (46%) showed an improvement in peak power strength, 16 of 50 athletes (32%) showed an improvement in average power conditioning, and 19 of 50 athletes (38%) showed an
improvement in peak power conditioning over the course of the 2012 winter training period.

As stated previously, athletes with data that resulted in estimation problems due to a lack of stability were not included in these results. Methods of dealing with the unstable data from these athletes will be discussed in the next chapter. Also, it may be possible to achieve models that better fit the data by considering VAR\((p)\) models for values of \(p\) other than one, which will also be explored further in the next chapter.
CHAPTER 4

4.1 Athletes with Unstable Data

As discussed in Section 1.4, when constructing a VAR model one important assumption in order to achieve reliable parameter estimates is that the stochastic autoregressive portion of the model, as given by Equation 1.4.1, is stable and thus stationary. The stability is investigated by checking the roots of Equation 1.4.3, given again below.

\[
\text{det}(I_n - \Phi_1 u - \Phi_2 u^2 - \ldots - \Phi_p u^p) = 0
\]  

If the roots of this equation have modulus that are less than one, then the VAR\((p)\) process is stable and thus also stationary. Otherwise the process is unstable, with unit roots (roots equal to one) being a particularly common problem in economic time series data (Hatemi-J, 2002).

As stated in Section 3.5, weight lifting data was collected from 89 athletes, 73 of which had at least six days of data and were modeled using the methods described in previous sections and chapters. After plotting and modeling the data, it was discovered that 50 of these 73 athletes produced stable data without any issues due to unit roots, and thus reliable parameter estimates. These results were given in the previous chapter. The other 23 athletes did exhibit potential problems with near unit roots and stability (typically on the small scale rep-to-rep series within a given day but also occasionally on the larger scale day-to-day series when modeling the daily intercept estimates to investigate large scale strength improvements) as determined by the above criteria and through the examination of data plots, and thus their results were omitted. In order to obtain parameter estimates for these 23 athletes, particularly the large scale linear time trend parameters that describe
the strength and conditioning improvements in both average and peak power as explained in the previous chapter, several potential fixes are considered.

The first option is to use non-overlapping averages, as has been done previously by several others as discussed in Section 1.3. For an unstable time series on the small scale, the small observations can be averaged at each large scale time point with a model then being fit to these averages to examine the large scale behavior. In the case of the weight lifting data, averaging means that rather than modeling an unstable series on the rep-to-rep scale within a given day, daily averages for each variable are calculated and a model is then fit to these daily averages to obtain estimates of the day-to-day linear time trend parameters. While this does not eliminate the possibility of the large scale series being unstable, it does at least take care of the problem of unstable small scale series. However, it also allows for less information on the small scale series at each of the large scale time points. For example, it was explained in the previous chapter that, for the weight lifting data, by fitting a model to the rep-to-rep data for each day, both a daily intercept and slope estimate could be obtained. The intercept estimates were then modeled on the day-to-day scale to examine strength improvements; the same was done with the slope estimates to examine conditioning improvements. By now using only daily averages of the variables and modeling these on the day-to-day scale, only the strength improvements are estimable.

Another option, instead of using non-overlapping averages, is to simply use the first small scale observation at each large scale time point and model these observations over the large scale to examine large scale time trends, a method also discussed in Section 1.3. Referring back to the weight lifting data, using the first observations means using only each variable’s measurement on the first rep of each day and then modeling these over the day-to-day scale rather than the daily averages of the variables. As with the daily averages, this method does not provide as much information on the rep-to-rep behavior within a day, but
it solves the problem of modeling an unstable series on the rep-to-rep scale. It also could still result in stability problems on the large scale, but in the cases that large scale stability is not an issue it at least allows for the evaluation of the strength improvements of an athlete through the estimation of the day-to-day linear time trend parameters. In addition, although the estimation of both an intercept and slope of the rep-to-rep observations for each day is still not possible, in the presence of a rep-to-rep linear time trend using the first rep measurements as a substitution for the intercept estimate may be more appropriate than using an average of the reps.

The most common solution when dealing with an unstable and thus non-stationary time series is to use the approach of first differencing as described by Nason et al. (2000). By modeling \( \Delta z_{s,t} = z_{s,t} - z_{s,1,t} \) rather than \( z_{s,t} \) itself the stationarity issues will typically be corrected. For the weight lifting data, first differencing means that the differences from rep-to-rep within each day are modeled as described in previous chapters. Because the original rep-to-rep observations within each day were assumed to be autoregressive with a non-autoregressive intercept and linear time trend, the differences are modeled with an autoregressive portion and a non-autoregressive portion. However, the non-autoregressive portion now contains only an intercept due to the fact that the original observations were assumed to exhibit only a linear time trend so the differences between observations should be nearly constant without depending on time. These intercept estimates for each day are now essentially the slope estimates for each day had the original observations been modeled. Thus, by modeling these daily intercept estimates on the first differenced data on the day-to-day scale (without differencing the estimates across days), the large scale linear time trends of the original daily slopes are still estimable. In the case of the weight lifting data, by differencing the rep-to-rep data within each day, although the day-to-day strength improvements can no longer be examined, the day-to-day conditioning improvements now
can be. The ability to examine the conditioning improvements but not the strength improvements is the opposite scenario that occurs when using either the non-overlapping averages or first observations from each day as described previously.

In Chapter 3, it was explained that the daily intercept estimates were modeled on the day-to-day scale to examine strength improvements and the daily slope estimates were modeled on the day-to-day scale to examine the conditioning improvements for the athletes with stable data. For the athletes with unstable data, either the non-overlapping averages or first observations for each day can be modeled on the day-to-day scale to examine the strength improvements. These two methods are compared to see which performs better for strength improvement evaluation. To examine the conditioning improvements, the daily intercept estimates obtained from modeling the rep-to-rep differences within each day can be modeled on the day-to-day scale.

The final options for dealing with the non-stationary data are to use either a log or square root transformation on the data as described by Weigand (2014). The data are then modeled on the small scale to obtain both intercept and slope estimates at each large scale time point, with these estimates then being modeled on the large scale to examine the strength and conditioning improvements, respectively. However, the issue here is that by transforming the data, whether using a log or square root transformation, the parameter estimates that describe these large scale trends will no longer be on the original data scale and thus comparisons with the athletes that did not require transformations will be more difficult. Each of the possible solutions discussed in this section is tested and compared through the use of simulations, as described in the following section.
4.2 Comparison of Simulation Results

One thousand simulations were run using the model outlined by Equations 3.2.4, 3.2.5, 3.2.6, and 3.3.1 as well as the relationship given by Equation 1.5.6. Those equations are given again below.

\[
\begin{align*}
\mathbf{z}_{s,t} &= \mathbf{y}_{0,t} + \mathbf{y}_{1,s} \mathbf{S} + \mathbf{\Gamma}_{1,t} \mathbf{w}_{s-1,t} + \mathbf{\omega}_{s,t} \\
&= \mathbf{\alpha}_{0,t} + \mathbf{\alpha}_{1,s} \mathbf{S} + \mathbf{\Gamma}_{1,t} \mathbf{z}_{s-1,t} + \mathbf{\omega}_{s,t} \\
\hat{\mathbf{y}}_{0,t} &= (\mathbf{I}_2 - \mathbf{\Gamma}_{1,0})^{-1} \mathbf{\alpha}_{0,t} - (\mathbf{I}_2 - \mathbf{\Gamma}_{1,0})^{-1} \mathbf{\Gamma}_{1,t} (\mathbf{I}_2 - \mathbf{\Gamma}_{1,0})^{-1} \mathbf{\alpha}_{1,t} \\
\hat{\mathbf{y}}_{1,t} &= (\mathbf{I}_2 - \mathbf{\Gamma}_{1,0})^{-1} \mathbf{\alpha}_{1,t} \\
\mathbf{y}_{0,t} &= \mathbf{\theta}_{0,0} + \mathbf{\theta}_{0,1} t + \mathbf{\Phi}_{0,1} \mathbf{y}_{0,t-1} + \mathbf{\varepsilon}_{0,t} \\
\mathbf{\mu}_0 &= (\mathbf{I}_n - \sum_{j=1}^{P} \mathbf{\Phi}_j)^{-1} \mathbf{\theta}_0 - (\mathbf{I}_n - \sum_{j=1}^{P} \mathbf{\Phi}_j)^{-1} (\sum_{j=1}^{P} \mathbf{\Phi}_j) (\mathbf{I}_n - \sum_{j=1}^{P} \mathbf{\Phi}_j)^{-1} \mathbf{\theta}_1 \\
\mathbf{\mu}_1 &= (\mathbf{I}_n - \sum_{j=1}^{P} \mathbf{\Phi}_j)^{-1} \mathbf{\theta}_1
\end{align*}
\]

The parameter values used for each of the 1000 simulations are as follows:

\[
\begin{align*}
T &= 10 \\
\mathbf{\mu}_{0,0} &= \begin{bmatrix} 3.6 \\ 3.6 \end{bmatrix}, \quad \mathbf{\mu}_{0,1} = \begin{bmatrix} 0.02 \\ 0.02 \end{bmatrix}, \quad \mathbf{\Phi}_{0,1} = \begin{bmatrix} 0.7 & -0.2 \\ 0.2 & 0.7 \end{bmatrix}, \quad \mathbf{\Sigma} = \begin{bmatrix} 0.02 & 0.00 \\ 0.00 & 0.02 \end{bmatrix}
\end{align*}
\]

\[
\begin{align*}
S &= 20 \\
\mathbf{\gamma}_{1,t} &= \begin{bmatrix} 0.00 & -0.95 \\ 0.00 & 0.95 \end{bmatrix}, \quad \mathbf{\Sigma}_t = \begin{bmatrix} 0.02 & 0.00 \\ 0.00 & 0.02 \end{bmatrix} \text{ for all } t (t = 1 \text{ to } 10)
\end{align*}
\]

These parameter values (specifically the intercepts, slopes, and covariance matrices) were chosen because they result in simulated data that closely resemble the values and trends found in the weight lifting data, which typically range from about 0.2 to
0.9. The large scale intercept was increased to 3.6 to ensure that none of the simulated data would result in negative values for the purposes of examining the log and square root transformations (the weight lifting data is also strictly positive). Although in the weight lifting data, the within day slopes are not assumed to remain constant from day to day, and in fact the change in these slopes is modeled to examine conditioning improvement, they are assumed to be constant for all \( t \) in these simulations as it is only the stationarity of the small scale series and large scale series of intercept estimates that are in question and in need of correction in this case. In addition, \( \Gamma_{1,2} \) was chosen because it results in a near unit root as determined by Equation 1.4.3. Thus by running 1000 simulations, where each consists of ten small scale time series generated by these parameters, it is expected that many of the simulations should result in non-stationary data on the small scale for at least one large scale time \( t \) due to this near unit root. The parameters of \( \Phi_{0,1} \) do not result in near unit roots or roots having modulus less than one meaning that the simulated large scale series should be stationary. However, due to randomness, it is possible that over the course of 1000 simulations there may be a few simulations where large scale stationarity is an issue. The scenario described matches what is seen in the weight lifting data, as most of the 23 athletes with unstable data exhibited stationarity problems on the rep-to-rep scale within a given day and only a few exhibited stationary problems on the larger day-to-day scale.

The model that was fit to the weight lifting data, as described in Sections 3.2, 3.3, and 3.4, as well as above, was fit to each of the two-scale bivariate time series resulting from these 1000 simulations. By examining the roots and plots of these series, as described in the previous section, it was determined that 430 of the simulations resulted in series with small scale non-stationarity problems for at least one large scale time \( t \). When fitting the large
scale model to the intercept estimates (Equation 3.3.1), 88 of the 1000 simulations resulted in series with large scale non-stationarity problems.

To determine the best method for obtaining parameter estimates for athletes exhibiting unstable data on either the small scale series within a given day or the large scale series of intercept estimates, the potential solutions discussed in the previous section were applied to the simulated data. The effectiveness of each method will be determined by the number of simulations still resulting in a non-stationary small scale or large scale series after the method has been applied. Because applying either the log or square root transformation to the data still allows for VAR models to be fit to both the small scale time series at each large scale time $t$, as well as the large scale time series of the resulting intercept estimates, they are considered first. The results are given in the table below.

<table>
<thead>
<tr>
<th>Potential Solution</th>
<th>Small Scale</th>
<th>Large Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original Data</td>
<td>430</td>
<td>88</td>
</tr>
<tr>
<td>Log</td>
<td>439</td>
<td>71</td>
</tr>
<tr>
<td>Square Root</td>
<td>433</td>
<td>80</td>
</tr>
</tbody>
</table>

These results are slightly better than those obtained without transforming the data for the large scales series, but not much improvement has been made. And in fact, modeling the data after utilizing these transformations actually resulted in more simulations with at least one unstable small scale series.

The next two options are to use either daily non-overlapping averages or first observations. Because for each of these methods a model is no longer being fit to the small scale observations at each large scale time $t$, less information will be obtained on the small scale behavior of the data at each large scale time point, but the stationarity of the data on the small scale is also no longer an issue because parameters are not being estimated. Now,
a VAR(1) model is only being fit to the large scale time series resulting from either the averages or first observations from each large scale time \( t \) so it is only the stationarity of this large scale series that is of concern. In the context of the weight lifting data, it allows for an estimate of the day-to-day strength improvement trends but not of the conditioning improvements because daily slope estimates are no longer attainable. The results of using these two methods on the simulated data are given in the table below.

<table>
<thead>
<tr>
<th>Potential Solution</th>
<th>Large Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original Data</td>
<td>88</td>
</tr>
<tr>
<td>Non-overlapping Averages</td>
<td>65</td>
</tr>
<tr>
<td>First Observations</td>
<td>40</td>
</tr>
</tbody>
</table>

It appears that these two methods, especially using the first observations at each large scale time point, dramatically reduce the number of simulated two-scale bivariate time series exhibiting problems with stationarity on the larger scale. Because of this result it seems that, although it does not allow for models to be constructed on the rep-to-rep observations within each day, in order to estimate the day-to-day trends in strength improvement for athletes with unstable data the best option is to model the first reps of each day.

As explained in the previous section, modeling the first differences is the most common solution to dealing with stationarity problems in VAR models. Once the observations have been differenced, however, the interpretations will change. In the case of the weight lifting data, by taking the differences between reps within a given day and modeling these differences, the resulting intercept estimate will approximate the within day slope of the original observations. Thus fitting a model to these daily intercept estimates will allow for the examination of the day-to-day trends in conditioning improvement that was not possible when using the first observations. First differencing was carried out for the
simulated data in order to see how effective differencing was in correcting the stationarity problems in the small scale observations at each large scale time $t$. Again, because the stationarity problems when looking at the conditioning improvements only occurred in the small scale series for the real data (and the simulations were run to mirror this situation) it is only the non-stationarity of the simulated small scale series that needs to be corrected through the use of differencing. The results are given in the table below.

<table>
<thead>
<tr>
<th>Potential Solution</th>
<th>Small Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original Data</td>
<td>430</td>
</tr>
<tr>
<td>First Differences</td>
<td>0</td>
</tr>
</tbody>
</table>

From these results it is seen that taking the first differences between small scale observations at each large scale time $t$ and fitting a VAR(1) model to these new series is quite effective as none of the 1000 simulations contained a single non-stationary small scale series. In order to examine the conditioning improvements by modeling the change in the daily slopes over the course of a training period for athletes with non-stationary rep-to-rep data within a given day, modeling the first differences is the clear solution as it removed all non-stationary small scale series from the original simulated data, which was much better than either the log or square root transformations.

From these results it appears that for the athletes with unstable data, for which the original model is not a reliable option, fitting a VAR model to the first reps of each day in order to obtain estimates of the day-to-day strength improvement trends, and fitting a VAR model to the differences between reps for each day and then fitting a VAR model to the resulting daily intercept estimates to obtain estimates of the day-to-day conditioning improvement trends is the best choice.
4.3 Identifying Trends

Although many of the methods discussed thus far can be applied to any bivariate time series occurring on two time scales in general, the ultimate goal in this specific case is to apply these methods to the weight lifting data collected from athletes in order to evaluate their improvements in strength and conditioning over the course of a training period. The basic idea is to be able to identify whether or not an athlete is in fact improving i.e. whether or not the day-to-day time trends of both the daily intercept and slope estimates are positive or negative.

Because of this ultimate goal, it is appropriate to also compare the potential solutions to dealing with the unstable data based on which are most successful at correctly identifying these long term trends as either positive or negative. As shown in the previous section, when attempting to fit a model to the \( \mu_{1,1} \) parameters (which denote the day-to-day improvements in the conditioning of the athletes) in the presence of unstable data, modeling the first differences is the clear choice as it provides much more improvement in reducing the number of unstable series than do the log or square root transformations, both of which actually increased the number of simulations with at least one unstable small scale series. Thus, it is only necessary to compare the potential solutions based on their abilities to correctly identify the \( \mu_{0,1} \) parameters, which denote the day-to-day improvements in the strength of the athletes, in the presence of unstable data either on the small or large scale.

The potential solutions for modeling these parameters were the log and square root transformations, as well as using the “daily” averages or first observations of the data. Based on each solution's ability to reduce the number of simulations with unstable data, it was discovered in the previous section that using the first observations seemed to perform best. This was because both the log and square root transformations still require a model to be fit to the small scale observations at each large scale time point \( t \), several of which were
still unstable, and also because using the first observations at each $t$ resulted in the greatest reduction in the number of simulations with unstable large scale series.

Simulations were again run using the parameter values given in Section 4.2, so that several simulations include at least one unstable small scale series and possibly an unstable large scale series, with one change: 1000 simulations were run with $\mu_{0,1} = \begin{bmatrix} 0.04 \\ 0.04 \end{bmatrix}$ and another 1000 simulations were run with $\mu_{0,1} = \begin{bmatrix} -0.04 \\ -0.04 \end{bmatrix}$. The four possible corrections were then applied to these 2000 simulations and the proportions of simulations which were correctly identified as containing either a positive or negative trend were recorded. The results are given in the table below.

<table>
<thead>
<tr>
<th>Potential Solution</th>
<th>$\mu_{1,0,1}$ Identified Correctly</th>
<th>$\mu_{2,0,1}$ Identified Correctly</th>
<th>$\mu_{1,0,1}$ and $\mu_{2,0,1}$ Both Identified Correctly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log</td>
<td>0.6670</td>
<td>0.6535</td>
<td>0.3390</td>
</tr>
<tr>
<td>Square Root</td>
<td>0.6735</td>
<td>0.6515</td>
<td>0.3425</td>
</tr>
<tr>
<td>Averages</td>
<td>0.7465</td>
<td>0.7705</td>
<td>0.5500</td>
</tr>
<tr>
<td>First Observations</td>
<td>0.8095</td>
<td>0.8010</td>
<td>0.6525</td>
</tr>
</tbody>
</table>

From these results, it can again be seen that using the first observation at each large scale time $t$ provides the best solution for the estimate of the large scale strength improvement scales when dealing with data that may be unstable on either the small or large scale. Using the first observations results in a much higher proportion of simulations with large scale trends correctly identified as either positive or negative, whether considering only $\mu_{1,0,1}$ (the trend for the first variable), only $\mu_{2,0,1}$ (the trend for the second variable), or both $\mu_{1,0,1}$ and $\mu_{2,0,1}$. This conclusion agrees with the results from the previous section, which identified using the first observations as the best method for reducing the number of simulations exhibiting unstable data. It also shows that fitting a VAR model to the
first reps of each day in order to obtain estimates of the day-to-day strength improvement trends is the best choice for the athletes with unstable data, for which the original model is not a reliable option. The same is true of fitting a VAR model to the differences between reps for each day and then fitting a VAR model to the resulting daily intercept estimates to obtain estimates of the day-to-day conditioning improvement trends for such athletes.

4.4 Lag Selection

So far, for the small scale models at each large scale time point \( t \), and for the large scale model, only a VAR(1) model has been considered for simplicity. A VAR(1) model was fit to the simulated data in Chapter 2 and Section 3.1, as well as to the initial weight lifting data in Section 3.5. However, because the data is no longer being simulated and the model has now been adjusted to include a linear time trend on the small scale at each large scale time point to better fit the real weight lifting data, it is also appropriate to consider increasing the lag of the model on the small scale at each large scale time point beyond \( p = 1 \) for the athletes exhibiting stable data for which modeling of the data on the small scale is possible. Because of the sample sizes that are expected for the weight lifting data of each athlete \( (S_1, S_2, ..., S_T, T) \), for the small scale models both VAR(1) and VAR(2) models are considered for each large scale time point \( t \). For the large scale model only a VAR(1) is considered as \( T \) generally is either only six or seven for each athlete. Thus the equations given in the previous sections, now modified to specify this modeling process, are given below.

\[
\begin{align*}
z_{s,t} &= y_{t(s)} + w_{s,t} \quad \text{where} \\
y_{t(s)} &= y_{0,t} + y_{1,t}S \\w_{s,t} &= \Gamma_{1,t}w_{s-1,t} + \cdots + \Gamma_{p-1,t}w_{s-p+1,t} + \omega_{s,t}
\end{align*}
\]
\[ z_{s,t} = y_{0,t} + y_{1,t}s + \Gamma_{1,t}w_{s-1,t} + \ldots + \Gamma_{p,t}w_{s-p,t} + \omega_{s,t} \quad \text{or} \]
\[ z_{s,t} = \alpha_{0,t} + \alpha_{1,t}s + \Gamma_{1,t}w_{s-1,t} + \ldots + \Gamma_{p,t}w_{s-p,t} + \omega_{s,t} \quad (4.4.5) \]

\[ \hat{y}_{0,t} = (I_2 - \sum_{j=1}^{p} \hat{\Gamma}_{j,t})^{-1} \hat{\alpha}_{0,t} - (I_2 - \sum_{j=1}^{p} \hat{\Gamma}_{j,t})^{-1} \sum_{j=1}^{p} j \hat{\Gamma}_{j,t} (I_2 - \sum_{j=1}^{p} \hat{\Gamma}_{j,t})^{-1} \hat{\alpha}_{1,t} \quad (4.4.6) \]

\[ \hat{y}_{1,t} = (I_2 - \sum_{j=1}^{p} \hat{\Gamma}_{j,t})^{-1} \hat{\alpha}_{1,t} \]

\[ y_{0,t} = \theta_{0,0} + \theta_{0,1}t + \Phi_{0,1}y_{0,t-1} + \varepsilon_{0,t} \quad (4.4.7) \]
\[ y_{1,t} = \theta_{1,0} + \theta_{1,1}t + \Phi_{1,1}y_{1,t-1} + \varepsilon_{1,t} \quad (4.4.8) \]

As explained above, these models are implemented for \( p = 1, 2 \) in (4.4.5) for each \( t \) \((t = 1 \text{ to } T)\). After estimating each of these models for each \( p \), the Schwarz-Bayesian Information Criterion (BIC) as outlined by Zivot and Wang (2006) is calculated for the large scale series, as it is the estimation of the large scale parameters that are of greatest interest for the weight lifting data. The BIC formula is given below where \( \Sigma \) is the maximum likelihood estimate of the error variance-covariance matrix for the large scale model as given by Equation (1.6.4).

\[ \text{BIC} = \ln |\hat{\Sigma}| + \frac{\ln(T)}{T}pn^2 \quad (4.4.9) \]

For each athlete, this BIC is calculated for both \( p = 1 \) and \( p = 2 \) on the small scale and \( p = 1 \) on the large scale. The small scale \( p \) that results in the minimum BIC value can be considered \( p_{\text{optimal}} \) and is used in the final estimation of the models.

Another option of choosing \( p_{\text{optimal}} \) for the small scale models in this particular case is to estimate the models given by (4.4.7) and (4.4.8) in order to obtain estimates of \( \mu_{0,1} \) and \( \mu_{1,1} \), which are the large scale linear time trends used to evaluate athletes’ strength and
conditioning improvements, using the equations given by (1.5.6), as well as the standard errors of those estimates for both $p = 1$ and $p = 2$. Because the estimation of these parameters is the ultimate goal in the analysis of the weight lifting data, the $p$ which results in smaller standard errors for these parameter estimates can be considered $p_{\text{optimal}}$.

Originally, there were 73 athletes for whom data had been collected over enough days to consider fitting the model. Fifty of those 73 athletes exhibited stationary data and thus the original VAR(1) models were fit on both the small and large scales without issue. However, when attempting to fit VAR(2) models on the rep-to-rep scale within a given day, it was discovered that the resulting matrix of explanatory variables was singular for at least one large scale time $t$ for four of the 73 athletes, including two of the 50 athletes exhibiting stable data on both the small and large time scales. Thus fitting VAR(2) models to the rep-to-rep data is only possible for 69 athletes, 48 of which show no stationarity problems.

When comparing the BIC values for the 48 athletes originally identified as having stable data, it was found that 35 out of 48 (72.9%) indicated a better fit with $p = 1$ for the small scale models than with $p = 2$ for the small scale models. When comparing the standard errors of the estimates of $\mu_{0,1}$ (the day-to-day strength improvement parameters), it was found that fitting VAR(1) models on the small scale resulted in smaller standard errors 66.7% of the time. Similarly, when comparing the standard errors of the estimates of $\mu_{1,1}$ (the day-to-day conditioning improvement parameters), it was found that fitting VAR(1) models on the small scale resulted in smaller standard errors 60.4% of the time.

Based on these results, it appears that whether using BIC as the criterion for model selection or comparing the standard errors of the parameters of primary interest in the case of the weight lifting data, fitting VAR(1) models to the rep-to-rep data within each day provide superior results to fitting VAR(2) models to the rep-to-rep data within each day the majority of the time. Thus, for the 50 athletes exhibiting stable data, the modeling as done in
Section 3.5 stands as the preferred method, with the resulting parameter estimates (also given in Section 3.5) as the final results. For the 23 athletes whose data exhibit either small scale or large scale stability problems, the optimal corrections as described and identified in Sections 4.1, 4.2, and 4.3 are applied. The results are given in the following section.

4.5 Weight Lifting Results for Athletes with Unstable Data

As investigated and explained previously in this chapter, the best option for obtaining estimates for the large scale strength trend parameters for athletes with unstable data is to enter the first reps completed each day into the large scale model as given by Equation 3.3.1. The best option for obtaining estimates for the large scale conditioning trend parameters for athletes with unstable data is to calculate the first differences between reps within a given day and then enter those values into the small scale model as given by Equation 3.2.4. The resulting daily intercept estimates resulting from those \( t \) models can then be entered into the large scale model as given by Equation 3.3.2. Completing this process should result in stable large scale series which allows for reliable parameter estimates for both models given by Equations 3.3.1 and 3.3.2, which in turn allows for reliable estimates of \( \mu_{0,1} \) and \( \mu_{1,1} \), as well as the standard errors of those estimates, to be calculated as shown by Equation 1.5.6 and demonstrated previously.

These steps were followed to obtain these parameter estimates for the 23 athletes with unstable data. Fitting the large scale model to the first reps of each day resulted in a stable large scale series, and reliable large scale strength trend parameters, for 22 of the 23 athletes. Modeling the first differences between reps within each day on the small scale and the resulting daily intercept estimates on the large scale resulted in stable small scale and large scale series, and reliable large scale conditioning trend parameters, for 21 of the 23 athletes. Thus applying these corrections resulted in estimates for all trend parameters...
being obtained for 20 of the 23 athletes that originally exhibited unstable data and to which the original modeling process could not be applied.

The parameter estimates, along with their standard errors, for these 20 athletes are given below. The results for the large scale (day-to-day) strength improvement trends ($\mu_{0,1}$) are given first in Table 4.5.1 below, where $\mu_{1,0,1}$ is the large scale linear time trend for average power and $\mu_{2,0,1}$ is the large scale linear time trend for peak power.

<table>
<thead>
<tr>
<th>Athlete</th>
<th>Average Power Strength Trend Estimate ($\hat{\mu}_{1,0,1}$)</th>
<th>$\hat{\mu}_{1,0,1}$ Std. Error</th>
<th>Peak Power Strength Trend Estimate ($\hat{\mu}_{2,0,1}$)</th>
<th>$\hat{\mu}_{2,0,1}$ Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.0488</td>
<td>0.0342</td>
<td>-0.0394</td>
<td>0.0110</td>
</tr>
<tr>
<td>2</td>
<td>-0.0055</td>
<td>0.0324</td>
<td>0.0235</td>
<td>0.0084</td>
</tr>
<tr>
<td>3</td>
<td>-0.0491</td>
<td>0.0397</td>
<td>-0.2165</td>
<td>0.0561</td>
</tr>
<tr>
<td>4</td>
<td>-0.0220</td>
<td>0.0193</td>
<td>-0.0103</td>
<td>0.0213</td>
</tr>
<tr>
<td>5</td>
<td>0.0459</td>
<td>0.0292</td>
<td>0.0269</td>
<td>0.0444</td>
</tr>
<tr>
<td>6</td>
<td>-0.0090</td>
<td>0.0224</td>
<td>0.0135</td>
<td>0.0115</td>
</tr>
<tr>
<td>7</td>
<td>0.0030</td>
<td>0.0229</td>
<td>0.0322</td>
<td>0.0381</td>
</tr>
<tr>
<td>8</td>
<td>-0.0127</td>
<td>0.0357</td>
<td>-0.0589</td>
<td>0.0140</td>
</tr>
<tr>
<td>9</td>
<td>-0.0010</td>
<td>0.0077</td>
<td>0.0204</td>
<td>0.0090</td>
</tr>
<tr>
<td>10</td>
<td>0.0333</td>
<td>0.0021</td>
<td>0.0007</td>
<td>0.0006</td>
</tr>
<tr>
<td>11</td>
<td>-0.0181</td>
<td>0.0143</td>
<td>-0.0437</td>
<td>0.0146</td>
</tr>
<tr>
<td>12</td>
<td>0.0244</td>
<td>0.0182</td>
<td>-0.0133</td>
<td>0.0072</td>
</tr>
<tr>
<td>13</td>
<td>-0.0874</td>
<td>0.0817</td>
<td>-0.0332</td>
<td>0.0236</td>
</tr>
<tr>
<td>14</td>
<td>0.1155</td>
<td>0.2185</td>
<td>-0.0631</td>
<td>0.0311</td>
</tr>
<tr>
<td>15</td>
<td>-0.0384</td>
<td>0.0102</td>
<td>-0.0088</td>
<td>0.0053</td>
</tr>
<tr>
<td>16</td>
<td>0.0227</td>
<td>0.0230</td>
<td>0.0371</td>
<td>0.0323</td>
</tr>
<tr>
<td>17</td>
<td>0.0191</td>
<td>0.0321</td>
<td>0.0141</td>
<td>0.0091</td>
</tr>
<tr>
<td>18</td>
<td>0.0313</td>
<td>0.0148</td>
<td>0.0067</td>
<td>0.0232</td>
</tr>
<tr>
<td>19</td>
<td>0.0007</td>
<td>0.0048</td>
<td>-0.0632</td>
<td>0.0573</td>
</tr>
<tr>
<td>20</td>
<td>-0.0096</td>
<td>0.0270</td>
<td>-0.0162</td>
<td>0.0227</td>
</tr>
</tbody>
</table>

The results for the large scale (day-to-day) conditioning improvement trends ($\mu_{1,1}$) are given in Table 4.5.2 below, where $\mu_{1,1,1}$ is the large scale linear time trend for average power slope and $\mu_{2,1,1}$ is the large scale linear time trend for peak power slope.
From these results it is seen that 9 of the 20 athletes (45%) showed an improvement in average power strength over the course of the 2012 winter training period, as indicated by an estimated $\mu_{1,0,1}$ value that was greater than zero. Similarly, 9 of the 20 athletes (45%) showed an improvement in peak power strength, 8 of 20 athletes (40%) showed an improvement in average power conditioning, and 8 of 20 athletes (40%) showed an improvement in peak power conditioning over the course of the 2012 winter training period.

Combining these results with the results for the athletes exhibiting stable data, given in the previous chapter, trend parameter estimates have now been obtained for 70 of the 89 total athletes for which data was collected. (Of the 19 athletes for which parameter

| Table 4.5.2  Bench Press Conditioning Trend Estimates and Standard Errors |
|-----------------|-----------------|-----------------|-----------------|
| Athlete | Average Power Conditioning Trend Estimate ($\hat{\mu}_{1,1,1}$) | $\hat{\mu}_{1,1,1}$ Std. Error | Peak Power Conditioning Trend Estimate ($\hat{\mu}_{2,1,1}$) | $\hat{\mu}_{2,1,1}$ Std. Error |
| 1 | -0.0109 | 0.0326 | 0.0002 | 0.0390 |
| 2 | -0.0170 | 0.0423 | 0.0013 | 0.0285 |
| 3 | -0.0157 | 0.0586 | -0.0292 | 0.0834 |
| 4 | -0.0103 | 0.0008 | -0.0091 | 0.0107 |
| 5 | 0.0046 | 0.0102 | -0.0077 | 0.0028 |
| 6 | -0.0205 | 0.0057 | -0.0152 | 0.0070 |
| 7 | 0.0005 | 0.0152 | -0.0081 | 0.0180 |
| 8 | -0.0017 | 0.0250 | 0.0027 | 0.0069 |
| 9 | -0.0003 | 0.0031 | 0.0113 | 0.0084 |
| 10 | 0.0030 | 0.0522 | 0.0929 | 0.1267 |
| 11 | -0.0133 | 0.0875 | 0.0038 | 0.0229 |
| 12 | 0.0181 | 0.0213 | -0.0028 | 0.0091 |
| 13 | 0.0004 | 0.0161 | -0.0052 | 0.0349 |
| 14 | -0.0042 | 0.0218 | 0.0029 | 0.0205 |
| 15 | 0.0171 | 0.0020 | -0.0019 | 0.0013 |
| 16 | 0.0109 | 0.0068 | -0.0058 | 0.0108 |
| 17 | -0.0436 | 0.0091 | -0.0263 | 0.0073 |
| 18 | -0.0633 | 0.0737 | -0.0781 | 0.1217 |
| 19 | -0.0070 | 0.0214 | 0.0115 | 0.0416 |
| 20 | 0.0026 | 0.0050 | -0.0100 | 0.0075 |
estimates were not obtained, 16 had less than six days of bench press data while the other
three still were exhibiting signs of unstable data after the corrections were applied.) Overall,
41 of the 70 athletes (59%) showed improvement in average power strength, 32 of 70
(46%) showed improvement in peak power strength, 24 of 70 (34%) showed improvement
in average power conditioning, and 27 of 70 (39%) showed improvement in peak power
conditioning.

While these results illustrate the ability of the proposed method of modeling to
identify the large scale strength and conditioning trends of the weight lifting data, it is
important to note that by including other factors the evaluation of the athletes may be
improved. For example, athletes that have been in the program for different lengths of time
will likely be expected to show different rates of improvement. Athletes also may not be
completing each exercise in the same order, which can have an effect. Similarly, fatigue from
previous workouts, injuries, and illnesses can all play a role in the athletes’ performances as
well as the workout regimens to which they have been assigned.
CHAPTER 5

5.1 Discussion

As first introduced in Chapter 1, when working with time series data there may occasionally be more than one temporal scale to consider. Observations are collected on a certain time scale, but there may also be a larger scale for which the behavior of the data is of interest. These multiple time scales can occur whether dealing with the univariate or the multivariate case. In such situations, it is often an objective to not only describe the behavior of the data on the smaller observed scale, but also on the larger underlying scale in order to examine the long term trends.

Other methods of dealing with this situation, such as using non-overlapping averages or omitting data, allow for the examination of the data’s behavior on larger scales, but the data on the observable scale are typically assumed to be equally spaced and thus these methods essentially look at the behavior of the data as a single time series at varying degrees of resolution. In the case of the weight lifting data, there is also the issue of larger time spacing between observations from one large scale time point to the next, with consecutive reps within a given day likely being only seconds apart while consecutive reps across days are several hours apart. Because of this disparity, the method proposed here creates a separate model for the observations at each large scale time point in order to examine the small scale behavior of the rep-to-rep data within a given day. The resulting daily intercept and slope estimates are then modeled to examine the long term improvements in the athletes’ strength and conditioning on the larger day-to-day scale.

Through simulations, as outlined in Chapter 2, it is seen that this method performs comparably to the popular method of using non-overlapping averages when attempting to identify the long term time trends of the data as evidenced by the estimated power and type I error rates for hypothesis testing on these parameters. In addition, by fitting a model to
the data which includes a linear time trend as well as stochastic autoregressive terms, more information on the behavior of the data on the smaller scale is available for each large scale time point than is provided by an average.

As discussed in Chapter 3, fitting a separate model to the small scale observations at each large scale time point (i.e. to the reps within each day for the weight lifting data) also allows for the change in the daily parameter estimates to be monitored, specifically the daily slope estimates. In addition to fitting a model to the daily intercept estimates to evaluate the athletes’ improvement in strength, a model is also fit to the daily slope estimates to provide a measurement of the athletes’ improvement in conditioning. Together, these values provide coaches with an overall summary of each athlete’s performance over the course of a seasonal training period.

Because the modeling is done through the use of vector autoregression, one of the assumptions required for reliable parameter estimates is that the data be stationary. For the majority of the athletes the lack of stationarity was not a problem. However, for athletes that did exhibit non-stationary data, it was shown in Chapter 4 that by using the observations on the first reps within each day, and by fitting a model to the rep-to-rep differences within each day, most of these stationarity problems were overcome. In these cases, although the original data is no longer modeled on the rep-to-rep scale within each day and thus some of the information on the small scale behavior must be sacrificed, it is at least possible for the large scale time trend parameters that describe both the strength and conditioning improvements to be estimated.

5.2 Future Research

Beyond what has been presented here, several further extensions are possible. Although the models and methods developed are easily extendable to the general \( n \) variable
case, only the bivariate case has been explored thus far. It may be of interest to examine the performance of these methods, whether through simulation or application to real data, for more than two variables. Similarly, the methods could also be extended to more than two time scales. In the case of the weight lifting data specifically, this may mean looking at not only the trends of the data on the rep-to-rep scale within a given day and the day-to-day scale within a given seasonal training period, but also across several training periods.

Also, in the specific case of the weight lifting data, there are other factors that may be considered in order to provide more exact evaluation of the athletes' strength and conditioning improvements over the course of the training period. Some of these factors have been given at the end of the previous chapter, such as the length of time an athlete has been in the program, possible fatigue due to the order of the exercises, workout adjustment due to injuries, etc.

Other possibilities include fitting \( \text{VAR}(p) \) models where \( p \) may differ for each model and making comparisons for model selection in situations with large sample sizes; examining the change in small scale autoregressive parameters from one large scale time point to the next, as these parameters may be of greater interest in various applications; evaluating the ability of this method to forecast future observations, which is a popular application to financial multivariate time series data; and considering additional corrections for various possible assumption violations.

5.3 Conclusion

Overall, the proposed method provides a viable alternative to previous methods of dealing with multivariate time series on multiple temporal scales and also provides additional information by allowing for time trend and autoregressive parameters to be fit and estimated on both the small and large scales. While this approach can be applied to a
myriad of situations, it is seen to be especially useful in the particular case of using weight lifting data to evaluate the long term performance and improvements in the strength and conditioning of athletes.
REFERENCES


Coughlin, K., & Eto, J. H. (2010). *Analysis of Wind Power and Load Data at Multiple Time Scales*. Lawrence Berkeley National Laboratory, University of California, Berkeley, California, pp. 3-5.


