Old Dominion University
ODU Digital Commons

Mathematics & Statistics Theses & Dissertations | Mathematics & Statistics

Winter 2010

Canonical Correlation Analysis for Longitudinal Data

Raymond McCollum
Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/mathstat_etds

 Part of the Applied Statistics Commons, and the Longitudinal Data Analysis and Time Series Commons

Recommended Citation
McCollum, Raymond. "Canonical Correlation Analysis for Longitudinal Data" (2010). Doctor of Philosophy (PhD), dissertation, Mathematics and Statistics, Old Dominion University, DOI: 10.25777/sx07-de18
https://digitalcommons.odu.edu/mathstat_etds/38

This Dissertation is brought to you for free and open access by the Mathematics & Statistics at ODU Digital Commons. It has been accepted for inclusion in Mathematics & Statistics Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

CANONICAL CORRELATION ANALYSIS FOR LONGITUDINAL DATA

by

Raymond McCollum
B.S. August 1996, Old Dominion University
M.S. May 2000, Kansas State University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirement for the Degree of

DOCTOR OF PHILOSOPHY

MATHEMATICS AND STATISTICS

OLD DOMINION UNIVERSITY
December 2010

Approved by:

Dr. Dayanand Naik (Director)

Dr. Rao Chaganty

Dr. Norou Diawara

Dr. Edward Markowski

ABSTRACT

CANONICAL CORRELATION ANALYSIS FOR LONGITUDINAL DATA

Raymond McCollum
Old Dominion University, 2010
Director: Dr. Dayanand Naik

Data (multivariate data) on two sets of vectors commonly occur in applications. Statistical analysis of these data is usually done using a canonical correlation analysis (CCA). Occurrence of these data at multiple occasions or conditions leads to longitudinal multivariate data for a CCA. We address the problem of canonical correlation analysis on longitudinal data when the data have a Kronecker product covariance structure. Using structured correlation matrices we model the dependency of repeatedly observed data. Recent work of Srivastava, Nahtman, and von Rosen (2008) developed an iterative algorithm to determine the maximum likelihood estimate of the Kronecker product covariance structure for one set of variables. We implement and generalize their method to estimate the covariance parameters in the context of canonical correlation analysis. We implemented unstructured and autoregressive covariance structures for the repeated measures. However, the developed methods can be easily implemented for other covariance structure. Testing of hypothesis problems using the likelihood ratio test statistics are explored. Bootstrap methods are adopted for calculating the p-values of the tests. Methods are illustrated on a data set obtained from NASA. Performance of the tests is explored using simulation experiments. Consequences of assuming the independence, between repeated measures and performing CCA at different time components, on the distribution of estimated canonical correlations is also explored. Certain simple tests to study the effect of repeated measures are provided here as well.

ACKNOWLEDGEMENTS

I am heartily thankful to my advisor Dr. Dayanand N. Naik for his support and guidance throughout the project. I would also like to thank my committee Dr. Rao Chaganty, Dr. Norou Diawara, and Dr. Edward Markowski for their time and expertise to better my work. My thanks also go to NASA Langley Research Center, Dr. Peter A. Parker for helping to find data relevant to the research topic and our many discussions about the process. I am grateful to Dr. Gao Chen for providing data and subject matter expert knowledge to help interpret and analyze the data from the Intercontinental Chemical Transport Experiment. Lastly I would like to acknowledge my family and express my appreciation for their continued support during the completion of the degree.

I dedicate this thesis to my children Ciaran and Kevin McCollum who have lived their entire early childhood knowing their father as a Ph.D. student, and my wife Julie McCollum for her patience and understanding.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Studying the relationship between two sets of multivariate vectors is an important problem in statistics. Canonical correlation coefficients are used to study these relationships. Canonical correlation analysis (CCA) is a general multivariate method that is mainly used to study the relationship when both the sets of variables are quantitative. The method was introduced by Hotelling (1936) and the analysis utilizes the variance covariance matrices of the two variables as well as the covariance matrix between the two variables. For example, the analysis is based on a matrix of the form:

$$\begin{bmatrix} \Sigma_y & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_x \end{bmatrix}.$$

Generalization of the method to more than two sets of variables was done by Kettenring (1971).

In this dissertation, we provide theory and application of CCA to repeatedly observed (over a time period) data on two sets of variables. The data here look like data on a multivariate time series, but the interest is to perform CCA on the two sets of vectors. One can put all the data, on the two vectors, corresponding to the time periods into large vectors and perform CCA using covariance matrices of these vectors. However, the procedure requires the covariance matrix between each set of vectors be calculated. Estimating the variance covariance matrix requires many data points. In fact, the number of variance covariance parameters in the matrix grows dramatically as the number of time periods increases. Further, when the longitudinal data are analyzed directly with no specific structure to the covariance matrix, the results of repeated CCA are hard to interpret. The resulting variables and coefficients are the linear combinations of variable one at time one and variable one at time two and so on make it difficult to form a complete picture of the data. Random variation of the parameter estimates increases the difficulty by making it seem one time period is more or less important than another. Assuming certain structures to the variance covariance matrices can alleviate this problem. This research addresses the longitudinal data structure and several possible model matrices will be explored

in an effort to model the data.

While CCA has been generalized in the literature by Kettenring (1971), there are not many papers directly dealing with CCA for longitudinal data. Srivastava (2007) and Srivastava and Naik (2008) considered this problem and provided partial solution. Modeling longitudinal CCA with a specific variance covariance structure, allows better fit between the model and the data. It also allows the analysis to be completed using less data points and allows the user to interpret the results more easily. This research uses a Kronecker product structure for performing a longitudinal CCA. The utilized matrix is

$$\begin{pmatrix} \Psi_y \otimes \Sigma_y & \Psi_{yx} \otimes \Sigma_{yx} \\ \Psi_{xy} \otimes \Sigma_{xy} & \Psi_x \otimes \Sigma_x \end{pmatrix}. \tag{1}$$

Kronecker product covariance matrices for analyzing multivariate repeated measures data have been utilized in the past. For example, see Naik and Rao (2001), Chaganty and Naik (2002), Roy and Khattree (2005), Srivastava, Nahtman, and von Rosen (2008) and others. The Kronecker product structure is fit to the data as an additional restriction on the CCA analysis. Hence all the existing assumptions of CCA are present and the Kronecker product covariance structure provides more subject matter expert insight to the model.

The objective of longitudinal CCA is the same as that in the standard CCA in that the researcher is trying to maximize the correlation between linear combinations of the variables. The added advantage of the Kronecker product covariance structure is that it makes the solution much more interpretable than it would be without the structure. Our work builds on previous work by Srivastava, Nahtman, and Von Rosen (2008) that solved the problem of determining the explicit forms of maximum likelihood estimates for a multivariate normal model using a Kronecker product covariance structure. They used a flip-plop iterative scheme to iterate between the two explicit forms of the estimators of the two matrices in the assumed Kronecker product. We carefully analyzed the existing solution and broken down the estimators into sum of several matrices. We showed that the maximum likelihood estimator can be actually written as a weighted linear combination of the smaller biased estimators. With the use of Hadamard product and other matrix theory results, weights on the linear combination of biased estimators are shown to be the inverse values from the

accompanying matrix in the Kronecker product structure. This insight and understanding of their solution was used as a basis for proposing an estimate of the cross product term in the Kronecker product variance covariance matrix. Rearranging the longitudinal component of the data allows the development of maximum likelihood estimators for some of the cross product terms. Details are discussed in Chapter III.

There are several possible structures for use in modeling the variance covariance matrix. In Chapter III, the merits of each is discussed in detail and the advantages of each are weighed against the others. Five different models are discussed in total. The first referred to as type I variance covariance matrix is the totally unstructured covariance matrix. This matrix assumes no information is available about the structure of the covariance. It also requires the most data. Type II covariance structure has three separate time matrices for the multivariate variables and another three matrices to model the time component. This matrix allows the researcher to explore different models in relation to the time correlation between the multivariate X and Y components. Type III covariance structure assumes either the X or the Y variable is not correlated in time while the other variable is assumed to be correlated in time. The cross product term is correlated appropriately as determined by the data. Type IV covariance structure assumes the time correlation component is identical for X, Y and the cross product between the two. This structure assumes that all correlation between the time points is the same. Through rearranging this matrix, it is possible to transform this estimator to get a maximum likelihood estimator of the covariance matrix for use in the CCA analysis. Finally, Type V covariance structure assumes there is no correlation between time points. In this case, it assumes that each set of multivariate data are independent of every other set collected within the X variables and similarly within the Y variables. The specific form of the time correlation is also analyzed. For example, an AR(1) correlation structure was assumed to represent the structure of the multivariate data between time points.

While the MLE was discovered for the likelihood with only one Kronecker product structure, the variance covariance matrix for longitudinal CCA is a partitioned matrix with four Kronecker product structures (see the structure given above in (1)). Through transformations, a maximum likelihood estimator was only available for 3 of the 5 covariance structures of interest. Maximum likelihood estimators of some matrices were developed for models I, IV, and V. Maximum likelihood estimates were not available for models II and III.

For the matrix structures that did not have an MLE solution, the exiting MLE solution was mimicked for each of the four partitions of the variance covariance matrix. The upper left and lower right partitions could use the existing MLE formula. The cross correlation matrix needed its own estimator and was the most difficult to attain. Using a combination of the Hadmard product, eigenvectors, the uniqueness of eigenvalues of a full rank matrix, the transformation of the cross product term, and the asymptotic properties of the multivariate T distribution final estimators are proposed. These components were combined to create an estimate of the partitioned cross product term. Previous work by Srivastava and Naik (2008) used the SAS optimization procedure to solve for the cross correlation term in the CCA matrix. Our work here attempted to investigate the theoretical solutions to the this problem. The solution developed gave some insight as to how a general solution could be developed. See details in Chapter II.

Several testing sequences are explored in Chapter III, in an effort to reduce the covariance structure down to the simplest matrix. Two prime hypothesis testing paths are given. One travels through a uniform time correlation while the other tests to determine if one of the variables does not truly have a correlation in time. Both paths lead to the multivariate data taken independently in time. The estimation procedures learned in Chapter II can be utilized to estimate the different covariance structures. To ensure the correct variance complexity level is chosen, the researcher must test the covariances of each model in an attempt to reduce the complexity. A series of hypothesis tests are conducted using the loglikelihood ratio statistics to reduce the complexity of the covariance structure at each step. The log likelihood ratio statistic is used even though the estimates of covariance matrices II and III are not exactly the maximum likelihood estimators.

In practice, the researcher will not know exactly which covariance matrix is the correct matrix. He/she will have to perform repeated hypothesis tests until one test rejects an attempt to reduce the matrix. The tests listed below show the same testing information. In Chapter III, the tests are shown for each situation. Hence, all the hypothesis tests are shown when the type II covariance matrix is true. The hypothesis are carried out again when the type III covariance matrix is true. Similarly this is done when type IV and type V matrices are true. The results show reasonable rejection rates for large samples but very high rejection rates for small samples. This is particularly true for the cases where the type II hypothesis is tested. The

high rejection rate when reducing the matrix from type I to type II has greater implications. It means the first step of the hypothesis testing sequence will also be the last because for small samples the test often rejects even when the null hypothesis is true. Both testing paths have the test of reducing from type I to type II as their first step. The large rejection rate will most likely show a rejection of the test and hence the researcher will not attempt to reduce the model to one of the simpler covariance matrices.

These high rejection rates for type II hypothesis were attributed to the fact that the type II hypothesis estimator, while performing well in MSE and bias metrics, was not actually an MLE. To address this, a parametric bootstrapping procedure was used to perform each test and determine if the hypothesis could be accepted. Parametric bootstrapping performed very well with all observed type I error rates at or near expected levels. Parametric bootstrapping also showed high power by rejecting the reduction to hypothesis that were too simple. Using the parametric bootstrapping gives a good probability of reducing the matrix down to the correct covariance structure and stopping the reduction of the covariance matrix at that correct structure. Hence the parametric bootstrapping combined with the estimation techniques in Chapter II could be used to create an estimation and testing procedure. To illustrate the procedure, several data sets were pursued from industry and academia. This resulted in several false starts but eventually a data set that matched the longitudinal CCA came from NASA.

To illustrate the procedure, NASA Langley in Hampton, Virginia allowed access to the Intercontinental Chemical Transport Experiment (INTEX). The purpose of the experiment was to measure airflow in and out of major metropolitan cities around the world, but specifically central America and the Asian subcontinent. Airflow data from Mexico City was analyzed for two aircrafts March 19th, 2006 from 18:34:15 to 19:15:15. Two aircrafts flew wing tip to wing tip through the Mexico City Megaplex pollution outflow. The aircrafts recorded multiple sensor readings repeatedly over time from many elements. The data contained many errors and missing values but data from three elements were complete enough for analysis. Ozone, water vapor content, and carbon monoxide levels were all measured simultaneously by each plane flying in close proximity. This allowed the estimation of the Kronecker product variance covariance structure and led itself to canonical correlation analysis. Analysis of the data showed the changes in measurements between the sensor readings were

most similar for ozone and water vapor content. Unfortunately, the carbon monoxide measurements between sensor readings did not match under canonical correlation analysis. Hence, it was concluded the sensor would be good to estimate ozone jointly for each plane. Water vapor changes were marginally aligned between the planes while the observed carbon monoxide relation was not aligned well and should be investigated.

Finally, Chapter IV aimed to provide a simple analysis methodology and test to confirm if the data is indeed one of the Kronecker product covariance matrices. This chapter creates a test based on large sample theory to address the inefficiencies in the large sample approximations attempted in Chapter II. Chapter IV also examines the consequences of ignoring the Kronecker product variance covariance structure and instead assuming independence of data among the time components. Assuming such independence among the time components still allows researchers to estimate the variance component for each time period. The usual canonical correlation analysis were examined under this assumption and the canonical variables were used to estimate the variance covariance matrix. This then leads to estimation of the canonical correlations and the asymptotic distribution of the canonical correlations. The asymptotic distribution involves the duplication matrix and the requirement to create the matrix quickly for large t. A simple algorithm is developed to generate the duplication matrix. The algorithm begins with a matrix of zeroes and cycles through the elements of the base matrix, placing ones in the appropriate places. The algorithm is simpler than most existing methods and gives insight into the structure of the duplication matrix.

Once the asymptotic distribution of the canonical correlations is attained, it is clear that under hypothesis II through V, the eigenvalues should all be identical. Hence a test is constructed over the t time periods to determine if the eigenvalues are similar. Due to the number of time intervals being directly proportional to the number of tests required, Bonferroni correction is required to ensure the significance level of the test remains constant. Tables are provided showing the resulting levels. The test offers a method using asymptotic statistics to determine if the matrix is totally unstructured or if one of the structures II through V should be used.

CHAPTER II

ESTIMATION OF COVARIANCE MATRICES

Canonical correlation analysis (CCA) is used to identify and characterize the relationship between two sets of random vectors. The procedure uses the variance covariance matrices of these vectors to establish the mathematical relationship between them. It is one of the most general methods of determining a relationship between two sets of vectors in that it does not make any distributional assumptions. The method of CCA was introduced by Hotelling (1936) for use in instructional research when studying the relationship between two sets of variables. CCA is found in many applications and throughout many fields. It is routinely discussed in multivariate statistical textbooks including Johnson and Wichern (2002), Khattree and Naik (2000), and Mardia, Kent and Bibby (1979).

II.1 CANONICAL CORRELATION ANALYSIS

Let $\mathbf{\Sigma}_x$ be the positive definite variance-covariance matrix of the $p \times 1$ vector X, $\mathbf{\Sigma}_y$ be the positive definite variance covariance matrix of $q \times 1$ vector Y, and let the covariance matrix between X and Y be the $p \times q$ matrix $\mathbf{\Sigma}_{xy} = cov(X, Y)$. The positive definite variance covariance matrix of X and Y then can be written as

$$\begin{bmatrix} \mathbf{\Sigma}_y & \mathbf{\Sigma}_{yx} \\ \mathbf{\Sigma}_{xy} & \mathbf{\Sigma}_x \end{bmatrix}.$$

The objective of canonical correlation analysis is to create a relationship between the X variables and the Y variables. There are many methods to define a relationship, but CCA attempts to find a $q \times 1$ vector 'a' and a $p \times 1$ vector 'b' such that the correlation between $a'Y$ and $b'X$ is maximum. Specifically, the problem here is to find 'a_1' and 'b_1' such that

$$Corr(a_1'Y, b_1'X) = \frac{a_1'\Sigma_{xy}b_1}{\sqrt{a_1'\Sigma_x a_1}\sqrt{b_1'\Sigma_y b_1}}$$

is maximized. In order to solve this problem the following restrictions on the variances are put, that is, $Var(a_1'Y) = 1$ and $Var(b_1'X) = 1$. Such obtained linear

combinations pair 'a_1' ,'b_1' is called as the first pair of canonical variables and the maximum correlation is called the first canonical correlation coefficient. Further, the linear combinations of Y and X which are uncorrelated with the first pair such that the correlations between them is the next maximum are also obtained under the restriction $Var(a_i'Y) = 1$ and $Var(b_i'X) = 1$ and such that

$$Corr(a_i'Y, b_i'X) = a_i'\Sigma_{xy}b_i$$

is maximized. As mentioned, all canonical pairs of different order are uncorrelated, that is,

$$Cov(a_i'Y, a_j'Y) = 0 \quad i \neq j,$$

$$Cov(b_i'X, b_j'X) = 0 \quad i \neq j,$$

$$Cov(a_i'Y, b_j'X) = 0 \quad i \neq j.$$

The solution to CCA problem lies in obtaining eigenvalues and eigenvectors of certain matrices. The first pair of canonical variables $(a_1'Y, b_1'X)$ is equal to

$$e_1'\Sigma_y^{-1/2}Y \text{ and } f_1'\Sigma_x^{-1/2}X.$$

Similarly, the i th pair of canonical variables $i = 2, 3, ..., g = min(p, q)$ is equal to

$$e_i'\Sigma_y^{-1/2}Y \text{ and } f_i'\Sigma_x^{-1/2}X.$$

The vectors $e_2', e_3', \ldots, e_g'$ are the corresponding g eigenvectors of

$$\Sigma_y^{-1/2}\Sigma_{yx}\Sigma_x^{-1}\Sigma_{xy}\Sigma_y^{-1/2}, \tag{2}$$

and $f_1', f_2', \ldots, f_g'$ are the corresponding g eigenvectors of

$$\Sigma_x^{-1/2}\Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx}\Sigma_x^{-1/2}. \tag{3}$$

The correlations $corr(a_i'Y, b_i'X)$ are the square roots of the (the same set of) eigenvalues of the matrix in (2) or (3).

Canonical correlation analysis has been generalized to more than two sets of vectors by Kettenring (1971). Not many approaches are available in the literature to perform CCA on longitudinal data. Srivastava (2007) and Srivastava and Naik (2008) have provided some directions and partial solutions. In this chapter we provide a systematic approach to CCA of longitudinally observed variables.

II.2 REPEATED CANONICAL CORRELATION ANALYSIS

Let X and Y be repeated measurements observations over time. For a fixed time t¿0, let x_i and y_i be vectors observed at the ith time period $i = 1, ...t$. Then define

$$Y = (y_1', y_2', \ldots, y_t') \text{ and } X = (x_1', x_2', \ldots, x_t').$$

The variance covariance matrix of Y and X can be represented by a Kronecker product structured matrix

$$D_{t(q+p)\times t(q+p)} = \begin{pmatrix} \Psi_y \otimes \Sigma_y & \Psi_{yx} \otimes \Sigma_{yx} \\ \Psi_{xy} \otimes \Sigma_{xy} & \Psi_x \otimes \Sigma_x \end{pmatrix}. \tag{4}$$

The positive definite matrices Ψ_Y, Ψ_X, and Ψ_{XY} respectively represent the correlation matrices of repeated measurements of Y, of X, and correlation matrix of repeated measurements between Y and X. Kronecker product structures have been successfully utilized to analyze multivariate normal repeated measures data in Naik and Rao (2001), Chaganty and Naik (2002), Roy and Khattree (2005), Srivastava, Nahtman, and von Rosen (2008), and Srivastava, Nahtman and von Rosen (2008). As in other canonical correlation problems, the objective is to determine linear functions $a'Y$ and $b'X$ to maximize

$$\frac{a'\Psi_{yx} \otimes \Sigma_{yx} b}{\sqrt{a'\Psi_y \otimes \Sigma_y a}\sqrt{b'\Psi_x \otimes \Sigma_x b}}.$$

As previously defined, the variance must be restricted, for example as,

$$Var(a'X) = a'CovXa = a'(\Psi_x \otimes \Sigma_x)a = 1,$$
$$Var(b'Y) = b'CovYb = b'(\Psi_y \otimes \Sigma_y)b = 1.$$

The covariance between $a'X$ and $b'Y$ is

$$Cov(a'X, b'Y) = a'Cov(X, Y)b = a'\Psi_{xy} \otimes \Sigma_{xy}b.$$

The vectors 'a' and 'b' must be chosen to maximize the correlation among all possible pairs of variables.

As in CCA, the vectors 'a' and 'b' that maximize the correlation are given by,

$$e_i'(\Psi_y^{-1/2} \otimes \Sigma_y^{-1/2})Y \text{ and } f_i'(\Psi_x^{-1/2} \otimes \Sigma_x^{-1/2})X,$$

where e_i is the i th eigenvector of,

$$\Psi_y^{-1/2}\Psi_{yx}\Psi_x^{-1}\Psi_{xy}\Psi_y^{-1/2} \otimes \Sigma_y^{-1/2}\Sigma_{yx}\Sigma_x^{-1}\Sigma_{xy}\Sigma_y^{-1/2} \tag{5}$$

and f_i is the i th eigenvector of

$$\Psi_x^{-1/2}\Psi_{xy}\Psi_y^{-1}\Psi_{yx}\Psi_x^{-1/2} \otimes \Sigma_x^{-1/2}\Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx}\Sigma_x^{-1/2}.$$

The i th pair of canonical variate $i = 1, 2, ..., min(p, q) = g$ are

$$e_i'(\Psi_y^{-1/2} \otimes \Sigma_y^{-1/2})Y \text{ and } f_i'(\Psi_x^{-1/2} \otimes \Sigma_x^{-1/2})X.$$

Both e_i and f_i can be decomposed into the product of two eigenvectors. For example, let the first eigenvector of

$$\Sigma_y^{-1/2}\Sigma_{yx}\Sigma_x^{-1}\Sigma_{xy}\Sigma_y^{-1/2}$$

be denoted by $e_{1\Sigma}$ and let the first eigenvector of

$$\Psi_y^{-1/2}\Psi_{yx}\Psi_x^{-1}\Psi_{xy}\Psi_y^{-1/2}$$

be denoted by $e_{1\Psi}$. Then the eigenvector of (5) is given by,

$$e_{1\Psi} \otimes e_{1\Sigma}.$$

This structure is important because it allows for ease in interpretation. It is easy to see how the time variance covariance and the variance covariance of each variable plays a role in determining the correlation. The canonical coefficients can be separated and identified for each time period and for each separate variable.

II.3 ESTIMATION OF MATRICES IN A KRONECKER PRODUCT

In practice, the components of matrix D in (4) are not known. Hence those need to be estimated based on the observed data. Previous work by Srivastava, Nahtman, and von Rosen (2008) (SNV (2008)) discusses the maximum likelihood estimation of the Kronecker product structure matrix of the form $\Psi \otimes \Sigma$. In order to apply a similar approach to obtain estimates of $\Psi_{xy} \otimes \Sigma_{xy}$ here, we will first review that work below.

Let

$$z = \begin{pmatrix} z_{11} & z_{12} & \dots & z_{1t} \\ z_{21} & z_{22} & \dots & z_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ z_{p1} & z_{p2} & \dots & z_{pt} \end{pmatrix} = \begin{pmatrix} z_1' \\ z_2' \\ \vdots \\ z_p' \end{pmatrix}, \text{ where } z_i = (Z_{i1}, z_{i2}, \dots, z_{it})', i = 1, \dots, p$$

be a $p \times t$ random matrix such that $vec(z) \sim N_{pt}(\mu,\ \Psi \otimes \Sigma)$. That is,

$$f(vec(z), \mu, \Psi, \Sigma) = \frac{1}{(2\pi)^{(pt)/2}|\Psi \otimes \Sigma|^{1/2}} exp\left(\frac{-(vec(z) - \mu)'(\Psi \otimes \Sigma)^{-1}(vec(z) - \mu)}{2}\right). \tag{6}$$

Suppose $z_1, ..., z_N$ is a random sample of size N from the above multivariate normal distribution in (6). SNV (2008) obtained the maximum likelihood estimates of Σ and Ψ as follows. The MLE of an unrestricted positive definite matrix Σ is given as

$$\hat{\Sigma} = \frac{\sum_{i=1}^{N} z_{ic} \hat{\Psi}^{-1} z_{ic}'}{tN} \tag{7}$$

and similarly the MLE of an unrestricted matrix Ψ, except for the restriction that $\psi_{tt} = 1$, is given by

$$\hat{\Psi} = \frac{\sum_{i=1}^{N} z_{ic}' \hat{\Sigma}^{-1} z_{ic}}{pN} \tag{8}$$

and $\hat{\psi}_{tt} = 1$. Here

$$z_i = \begin{pmatrix} z_{i11} & z_{i12} & \cdots & z_{i1t} \\ z_{i21} & z_{i22} & \cdots & z_{i2t} \\ \vdots & \vdots & \ddots & \vdots \\ z_{ip1} & z_{ip2} & \cdots & z_{ipt} \end{pmatrix},$$

$$z_{ic} = z_i - \bar{z}, \tag{9}$$

where,

$$\bar{z} = \begin{pmatrix} \bar{z}_{11} & \bar{z}_{12} & \cdots & \bar{z}_{1t} \\ \bar{z}_{21} & \bar{z}_{22} & \cdots & \bar{z}_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{z}_{p1} & \bar{z}_{p2} & \cdots & \bar{z}_{pt} \end{pmatrix},$$

and

$$\bar{z}_{ij} = \frac{\sum_{i=1}^{n} z_{ijk}}{n}, j = 1, \ldots, p, k = 1, \ldots, t$$

The MLEs of Σ and Ψ are finally obtained by iterating between (7) and (8). The authors called this algorithm as a flip-flop algorithm. There are two structural arrangements of the random variables. Each arrangement of the variables results in the restructuring of the covariance matrix. The existing solutions take advantage of both of these forms simultaneously. The first arrangement $z_{.i}$ puts all variables that appear at the same time point in a vector. That is, for $i = 1, ..., t$,

$$z_{.i} = \begin{bmatrix} z_{1ic} \\ \vdots \\ z_{pic} \end{bmatrix}.$$

Using this notation, we have the following distribution:

$$\begin{bmatrix} z_{.1} \\ \vdots \\ z_{.t} \end{bmatrix} \sim N \left[\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \psi_{11}\Sigma & \cdots & \psi_{1t}\Sigma \\ \vdots & \ddots & \vdots \\ \psi_{t1}\Sigma & \cdots & \psi_{tt}\Sigma \end{pmatrix} \right].$$

The alternative arrangement is to list a single random variable across all time points. That is, for $j = 1, ..., p$ define,

$$z_{j\cdot} = \begin{bmatrix} z_{j1} \\ \vdots \\ z_{jt} \end{bmatrix}.$$

This results in the distribution:

$$\begin{bmatrix} z_{1\cdot} \\ \vdots \\ z_{p\cdot} \end{bmatrix} \sim N \left[\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \Psi\sigma_{11} & \dots & \Psi\sigma_{1p} \\ \vdots & \ddots & \vdots \\ \Psi\sigma_{p1} & \dots & \Psi\sigma_{pp} \end{pmatrix} \right].$$

II.4 EXAMINING THE SNV(2008) SOLUTION

The maximum likelihood estimators in (7) and (8) can be seen to be weighted sums of several estimators of the same matrix. The weights for each matrix are obtained from the inverse of the accompanying matrix. Hence, the weighting for the Σ estimate is based on estimated elements of Ψ^{-1} and the weighting to estimate Ψ is based on the Σ^{-1} estimate.

To better understand what the formula is doing, consider the inverse of estimated Σ matrix below:

$$\hat{\Sigma}^{-1} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1p} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2p} \\ \vdots & \ddots & & \vdots \\ \alpha_{p1} & \alpha_{p2} & \dots & \alpha_{pp} \end{pmatrix}.$$

The solution to the Ψ multiplies the z_{ic} matrix by the Σ inverse estimate,

$$\hat{\Psi} = \frac{\sum_{i=1}^{N} z'_{ic} \hat{\Sigma}^{-1} z_{ic}}{Np},$$

which can be written as:

$$\hat{\Psi} = \frac{\sum_{k=1}^{N} \alpha_{11} z_{k1\cdot} z'_{k1\cdot} + \alpha_{12} z_{k1\cdot} z'_{k2\cdot} + \ldots + \alpha_{pp} z_{kp\cdot} z'_{kp\cdot}}{pN}.$$

Note that this formula breaks down the estimate into the sum of individual biased estimates, $\sum_{k=1}^{N} z_{ki.} z'_{kj.}/N$ of Ψ. The expected value of each estimate is, $E(z_{i.} z'_{j.}) = \sigma_{ij}\Psi$. Hence the expected value of $\hat{\Psi}$ is

$$E(\hat{\Psi}) = \frac{\alpha_{11}\sigma_{11}\Psi + \alpha_{12}\sigma_{12}\Psi + \ldots + \alpha_{pp}\sigma_{pp}\Psi}{p} = \frac{(\alpha_{11}\sigma_{11} + \alpha_{12}\sigma_{12} + \ldots + \alpha_{pp}\sigma_{pp})\Psi}{p},$$

the numerator of which is Ψ times the the sum of the elements of Hadamard product (defined below) of Σ and Σ^{-1}.

Definition: Suppose $A = (a_{ij})$ and $B = (b_{ij})$ where $i = 1, \ldots, m$ and $j = 1, \ldots, n$ are two matrices of order $m \times n$ then the Hadamard product o of A and B is defined as $C = (c_{ij}) = AoB = (a_{ij}b_{ij})$.

An interesting property (Schott, 2007, p. 90) of Hadamard product is: $1'AoB1 = trace(AB') = \sum(a_{ij}b_{ij})$, where 1 is an $m \times 1$ vector of all ones.

When applied this result to $A = \Sigma$ and $B = \Sigma^{-1}$ we get $trace(I_p)$, which is the dimension of Σ, namely, p. However, since estimators of Σ and Σ^{-1} are used instead of the parameters, we get $E(\hat{\Psi}) \approx \Psi$, that is, the weighted sum of biased estimators is approximately unbiased.

Similarly, suppose

$$\hat{\Psi}^{-1} = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1t} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2t} \\ \vdots & \ddots & & \vdots \\ \gamma_{t1} & \gamma_{t2} & \cdots & \gamma_{tt} \end{pmatrix}.$$

Then

$$\hat{\Sigma} = \frac{\sum_{i=1}^{N} z_{ic}\hat{\Psi}^{-1}z'_{ic}}{Nt}$$

can be written as

$$\hat{\Sigma} = \frac{\sum_{k=1}^{N} \gamma_{11}z_{k.1}z'_{k.1} + \gamma_{12}z_{k.1}z'_{k.2} + \ldots + \gamma_{tt}z_{k.t}z'_{k.t}}{tN}. \tag{10}$$

As before the solution is weighted average of biased estimators of Σ. However, the combined estimator is approximately unbiased.

II.5 ESTIMATION OF AR(1) CORRELATION PARAMETER

Often in modeling the correlations, especially in the context of longitudinal data, certain simple correlation structures are useful. One popular and useful structure is the AR(1) structure, given by

$$\Psi = \Psi(\rho) = \begin{pmatrix} 1 & \rho & \dots & \rho^{t-2} & \rho^{t-1} \\ \rho & 1 & \dots & \rho^{t-3} & \rho^{t-2} \\ \vdots & \vdots & \ddots & \vdots & \\ \rho^{t-1} & \rho^{t-2} & \dots & \rho & 1 \end{pmatrix},$$

where ρ is the correlation coefficient between any two consecutive repeated measurements and $|\rho| \leq 1$. It is well known for this matrix that its determinant $|\Psi| = (1-\rho^2)^{t-1}$ and its inverse is:

$$\Psi^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & 0 & 0 & \dots & & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & \dots & & 0 \\ 0 & -\rho & 1+\rho^2 & \ddots & \dots & & 0 \\ 0 & \dots & \ddots & \ddots & \ddots & \dots & 0 \\ 0 & \dots & 0 & \ddots & 1+\rho^2 & -\rho & 0 \\ 0 & \dots & & 0 & -\rho & 1+\rho^2 & -\rho \\ 0 & \dots & & & 0 & -\rho & 1 \end{pmatrix}.$$

This can be expressed as $\Psi^{-1} = \frac{1}{1-\rho^2}(B_1 + \rho^2 B_3 - \rho B_2)$, where $B_1 = I_t = diag(1,1,...,1)$, $B_3 = diag(0,1,...,1,0)$, and B_2 is a tridiagonal matrix with 0 on the main diagonal and 1 on the upper and lower diagonals, that is,

$$\mathbf{B_2} = \begin{pmatrix} 0 & 1 & 0 & & \dots & 0 \\ 1 & 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & 0 & 1 \\ 0 & \dots & 0 & 0 & 1 & 0 \end{pmatrix}.$$

One can easily accommodate estimation of Σ and ρ when the covariance matrix is of Kronecker product $\Psi \otimes \Sigma$ structure, that is, when Ψ is of $AR(1)$ structure, $\Psi(\rho)$. Suppose $\hat{\Psi}$ is the estimator obtained using the equation (8). Then solving the

following cubic equation gives the maximum likelihood estimate of ρ for given $\hat{\Sigma}$, which is obtained using equation (7). Suppose $\hat{\rho}$ is the solution to

$$2(t-1)\rho^3 - tr(B_2\hat{\Psi})\rho^2 + 2(tr(B_1\hat{\Psi}) + tr(B_3\hat{\Psi}) - (t-1))\rho - tr(B_2\hat{\Psi}) = 0. \quad (11)$$

Then $\hat{\Psi} = \Psi(\hat{\rho})$. One needs to iterate between solving the cubic equation (11) and solving (7) and (8) to obtain estimates of Ψ and Σ.

II.6 ESTIMATING FOR REPEATED CCA

Now we will turn our attention to estimation of the matrices in our structure, namely,

$$D = \begin{pmatrix} \Psi_y \otimes \Sigma_y & \Psi_{yx} \otimes \Sigma_{yx} \\ \Psi_{xy} \otimes \Sigma_{xy} & \Psi_x \otimes \Sigma_x \end{pmatrix}.$$

The cross correlation matrix has two components Σ_{xy} and Ψ_{xy}. The Σ_{xy} matrix has no restrictions and accounts for the cross correlation between X and Y at each fixed time unit. The Ψ_{xy} matrix accounts for correlation across time units. The two structures of the Ψ_{xy} matrix that we have investigated are:

- AR(1)
- Unstructured correlation matrix.

In the AR(1) case, the model assumes the covariance matrix between X and Y differ by a multiple. The partitioned matrix illustrates how the X and Y values are correlated to each other and within time. It is important to note that each set of X variates have the same covariance Σ_x within a single point in time. The Y variates also have the same constant covariance structure Σ_y within a single time period.

The Ψ component of the covariance represents the correlation in time for each set of multivariate X and Y data. The X values at time i and the X values at time j have have a covariance value determined by multiplying the Variance of X by a constant. The Y values have a similar covariance structure. In the case where the time values have an AR(1) structure, the X vector taken at different time points is $Cov(X_i, X_j) = \rho^{|i-j|}\Sigma_x$. The Y values have a similar structure.

The cross correlation covariance has a similar structure. The covariance values within a time period can be expressed in a covariance matrix Σ_{xy}. Values across time points have a covariance that is multiplied by a constant. In the case of the

AR(1) structure the covariance components of two X vectors at time points i and j are $Cov(X_i, Y_j) = \rho^{|i-j|}\Sigma_{xy}$.

The previous work by Srivastava (2007) and Srivastava and Naik (2008) used the SAS optimization procedure to get estimates of the restricted covariance matrix with an AR(1) structure. One of the criticisms of that work was that no detailed exploration of the estimation procedure was given. This work attempts to explore the estimation of the the covariance matrices in D.

Let

$$Y = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1t} \\ y_{21} & y_{22} & \cdots & y_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ y_{q1} & y_{q2} & \cdots & y_{qt} \end{pmatrix}$$

be a $q \times t$ random matrix and

$$X = \begin{pmatrix} x_{11} & X_{12} & \cdots & x_{1t} \\ x_{21} & X_{22} & \cdots & x_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & X_{p2} & \cdots & x_{pt} \end{pmatrix}$$

be a $p \times t$ random matrix. Suppose

$$W = \begin{pmatrix} vec(Y) \\ vec(X) \end{pmatrix} \sim N_{t(p+q)}(\mu, D_{t(p+q),t(p+q)}),$$

that is, the pdf of W is

$$f(W, \mu, D) = \frac{1}{2\pi^{(pt+qt)/2}|D|^{1/2}} exp\left(\frac{-(W-\mu)'D^{-1}(W-\mu)}{2}\right) \tag{12}$$

$$\text{where }, D = \begin{pmatrix} \Psi_y \otimes \Sigma_y & \Psi_{yx} \otimes \Sigma_{yx} \\ \Psi_{xy} \otimes \Sigma_{xy} & \Psi_x \otimes \Sigma_x \end{pmatrix}.$$

Suppose Y_i and X_i are the random sample matrices of size N. Assume that Y_{ic} and X_{ic} are defined as before as how we had defined z_{ic} in (9). The block diagonal Kronecker product matrices $\Psi_y \otimes \Sigma_y$ and $\Psi_x \otimes \Sigma_x$ can be estimated using the work of SNV (2008). However the $\Psi_{xy} \otimes \Sigma_{xy}$ matrix is more difficult to estimate. The estimate of Σ_{xy} depends on Ψ_{xy}^{-1} which is attainable based on the structure of Ψ_{xy}. However, the estimate of Ψ_{xy} depends on Σ_{xy}^{-1}. There is no guarantee that an estimate of Σ_{xy}^{-1} exists.

However, estimators for Ψ_{xy} and Σ_{xy} on a similar lines as that of SNV (2008) can be constructed because the property of Hadamard matrix that we had used earlier to represent SNV (2008)'s estimator holds true not only for an inverse of a matrix, but also for a generalized inverse of a matrix, for example, for Moore-Penrose inverse.

As we know Moore-Penrose inverse of a matrix $\mathbf{A}$ (denoted by $\mathbf{A}^+$) is computed as $\mathbf{A}^+ = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$. and it satisfies the following properties (Ravishanker and Day, 2002, p. 81)

1. $\mathbf{AGA}=\mathbf{A}$, i.e., $\mathbf{G}$ is a generalized inverse of $\mathbf{A}$,
2. $\mathbf{GAG}=\mathbf{G}$, i.e., $\mathbf{A}$ is a generalized inverse of $\mathbf{G}$,
3. $(\mathbf{AG})'=\mathbf{AG}$, i.e., $\mathbf{AG}$ is symmetric,
4. $(\mathbf{GA})'=\mathbf{GA}$, i.e., $\mathbf{GA}$ is symmetric.

Now the following property of Hadamard Product holds true, that is,

$$(1,\ldots,1)\mathbf{A}o(\mathbf{A}^+)'\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix} = tr(\mathbf{A}\mathbf{A}^+) = rank(\mathbf{A}). \tag{13}$$

The equation in (13) holds true because of the fact that $\mathbf{A}\mathbf{A}^+$ is an idempotent matrix. Of course any other generalized inverse instead of Moore-Penrose inverse can be used here.

The Hadamard product multiplies each element of $\mathbf{A}$ matrix by the corresponding element in the $(\mathbf{A}^+)'$ matrix. After the multiplication the elements are summed. The end result of this procedure provides the rank of the matrix $\mathbf{A}$. As we have observed earlier, this shows that the SNV (2008)'s procedure creates a weighting for the individual estimates as in (10). The fact that the Hadamard product equals the rank shows SNV (2008)'s maximum likelihood estimator to be a weighted average. Using this logic, we create estimators for the cross correlation terms $\Psi_{xy}\otimes\Sigma_{xy}$. Direct attempts to maximize the likelihood in (12) did not produce a viable solution due to the mathematics being intractable. Using the same structure as in SNV (2008)'s solution, and taking advantage of the Hadamard product, an intuitive solution was create to approximate the Σ_{xy} matrix and the Ψ_{xy} matrix.

The Σ_{xy} estimate can use the inverse of the Ψ_{xy}.

$$\hat{\Sigma}_{xy} = \frac{\sum_{i=1}^{N} Y_{ic} \hat{\Psi}_{xy}^{-1} X_{ic}'}{Nt}.$$

However, to gain the weighting for the estimate of Ψ_{xy} an inverse is needed from the Σ_{xy} matrix. Since the Σ_{xy} cross correlation matrix may not be invertible or even square, there are many matrices that could qualify as the inverse. The solution below in (14) was created by substituting the Moore-Penrose inverse $\hat{\Sigma}_{xy}^{+}$ for the inverse of Σ_{xy}.

$$\hat{\Psi}_{xy} = \frac{\sum_{i=1}^{N} X_{ic}' \hat{\Sigma}_{xy}^{+} Y_{ic}}{min(p,q)N}. \tag{14}$$

Once Ψ_{xy} is estimated for each iteration, the general form of the solution must be forced into the proper form. In this case Ψ_{xy} is assumed to have an AR(1) structure and must conform to that structure. To create this estimate, the inverse covariance matrix can be written as a cubic equation as shown earlier.

The intuitive estimators (14) did not converge as well as expected. However, the maximum likelihood estimators of SNV (2008) always converged. The number of converging samples as a function of sample size is shown in Table 1. The ML estimators and the intuitive estimators are shown in Tables 2, 3, 4, 5, and 6.

II.7 SIMULATION METHODOLOGY

In order to illustrate the analysis discussed here, data was simulated according to the covariance matrix described above. First the Helmert matrix was used to generate the positive definite matrices. The general form of a Helmert matrix $\mathbf{H}_k$ of order k has $k^{-1/2}\mathbf{1}_k'$ for its first row, and each of its other $k-1$ rows for $i = 1, \ldots, k-1$ has the partitioned form

$$\left[\begin{array}{c|c|c} \mathbf{1}_i' & -i & 0 \end{array} \right] / \sqrt{a_i},$$

with $a_i = i(i+1)$. A Helmert matrix is an orthogonal matrix, that is, $\mathbf{H}'\mathbf{H} =$

$\mathbf{HH}' = \mathbf{I}_k$. For example, the 4^{th} order Helmert matrix is given by

$$\mathbf{H}_4 = \begin{bmatrix} \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{-3}{\sqrt{12}} \end{bmatrix}.$$

The spectral decomposition of a symmetric matrix, $\mathbf{A}$ is $\mathbf{A} = \sum \lambda_i \mathbf{u}_i \mathbf{u}_i'$, where the $\mathbf{u}_i$'s are the eigenvectors of $\mathbf{A}$. Now to generate a $k \times k$ positive definite matrix we take the kth order Helmert matrix, whose columns will give us the eigenvector of the desired matrix. Then choosing k positive eigenvalues and using the spectral decomposition property we can construct the desired $k \times k$ positive definite matrix. We will use thus constructed positive definite matrix as $\mathbf{\Sigma}$.

Partitioning $\mathbf{\Sigma}$ will give

$$\mathbf{\Sigma} = \begin{pmatrix} \Sigma_y & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_x \end{pmatrix},$$

and Σ_y, Σ_x and Σ_{yx} can be used as variance-covariance matrix for $\mathbf{y}$, $\mathbf{x}$ and covariance matrix between $\mathbf{y}$ and $\mathbf{x}$ respectively. Then by choosing $t \times t$ modeling matrix Ψ_y to associate with Σ_y, Ψ_x with Σ_x and Ψ_{yx} with Σ_{yx} we can construct the desired matrix

$$\mathbf{D} = \begin{bmatrix} \Psi_y \otimes \Sigma_y & \Psi_{yx} \otimes \Sigma_{yx} \\ \Psi_{xy} \otimes \Sigma_{xy} & \Psi_x \otimes \Sigma_x \end{bmatrix}.$$

Any number of observations can now be generated from the multivariate Normal $N(\mathbf{0}, \mathbf{D})$ and repeated canonical correlation analysis can be performed on them as discussed earlier.

For the simulation example, we chose three $\mathbf{y}$ components, two $\mathbf{x}$ components and three repeated measurements, that is, $q = 4$, $p = 2$, and $t = 3$. A Helmert matrix of order 6 is chosen and used to determine a 6×6 positive definite variance covariance matrix Σ. In addition, 12,1,8,6,4, and 2 were used as the eigenvalues which yielded the following positive definite matrix Σ by the method described earlier.

$$\Sigma =$$

$$\begin{bmatrix} 5.5 & 3.1754265 & -0.5 & 0.3535534 & 1.004158 & 1.5652476 \\ 3.1754265 & 6.5 & 3.1754265 & 2.2453656 & 1.7392527 & 1.4200939 \\ -0.5 & 3.1754265 & 7.5 & -0.353553 - 0.273861 & -0.223607 & \\ 0.3535534 & 2.2453656 & -0.353553 & 6.25 & 0.1936492 & 0.1581139 \\ 1.004158 & 1.7392527 & -0.273861 & 0.1936492 & 4.55 & 0.4490731 \\ 1.5652476 & 1.4200939 & -0.223607 & 0.1581139 & 0.4490731 & 2.7 \end{bmatrix}.$$

By partitioning Σ we get Σ_{yy}, Σ_{xx} and Σ_{yx} as:

$$\Sigma_{yy} = \begin{bmatrix} 5.5 & 3.1754265 & -0.5 & 0.3535534 \\ 3.1754265 & 6.5 & 3.1754265 & 2.2453656 \\ -0.5 & 3.1754265 & 7.5 & -0.353553 \\ 0.3535534 & 2.2453656 & -0.353553 & 6.25 \end{bmatrix},$$

$$\Sigma_{xx} = \begin{bmatrix} 4.55 & 0.4490731 \\ 0.4490731 & 2.7 \end{bmatrix} \text{ and } \Sigma_{yx} = \begin{bmatrix} 1.004158 & 1.5652476 \\ 1.7392527 & 1.4200939 \\ -0.273861 & -0.223607 \\ 0.1936492 & 0.1581139 \end{bmatrix}.$$

Assume $AR(1)$ structure for repeated modeling matrices Ψ_y, Ψ_x, and Ψ_{yx} with correlation parameter $\rho_y = 0.2$, $\rho_x = 0.4$, and $\rho_{yx} = 0.3$ respectively. Arranging all the matrices together we have

$$\mathbf{D} = \begin{bmatrix} \Psi_{yy} \otimes \Sigma_{yy} & \Psi_{yx} \otimes \Sigma_{yx} \\ \Psi_{xy} \otimes \Sigma_{xy} & \Psi_{xx} \otimes \Sigma_{xx} \end{bmatrix}.$$

One thousand simulations were observed for each sample of size 500, 350, 200,100, and 50 from the multivariate normal ($N(\mathbf{0}, \mathbf{D})$) distribution and the population parameters Σ_y, Σ_x, Σ_{yx}, ρ_y, ρ_x, and ρ_{yx} were estimated.

II.7.1 Intuitive Estimation Results

For a sample size of 500, 1000 simulations showed 996 converged. The ML estimators had a much smaller MSE and bias than the intuitive estimator. Table 2 shows the results for 500 samples.

For a sample size of 350, 1000 simulations showed 975 converged. The ML estimators had a much smaller MSE and bias than the intuitive estimator. Table 3 shows the results for 350 samples.

Sample Size	Number of Values Converged
500	996
350	975
200	871
100	645
50	515

TABLE 1: Intuitive estimate convergence values from 1000 simulations.

Sample Size 500	Simulations 1000			
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta-\hat{\Theta})^2)^{1/2}$	$\lvert(\hat{\Theta}-\Theta)\rvert$
$\Sigma_y(11)$	5.50000	5.49091	0.19355	0.00909
$\Sigma_y(21)$	3.17543	3.16755	0.16785	0.00788
$\Sigma_y(22)$	6.50000	6.48315	0.23614	0.01685
$\Sigma_y(31)$	-0.50000	-0.50270	0.15944	0.00270
$\Sigma_y(32)$	3.17543	3.16691	0.19235	0.00852
$\Sigma_y(33)$	7.50000	7.48407	0.27176	0.01593
$\Sigma_y(41)$	0.35355	0.34927	0.14595	0.00428
$\Sigma_y(42)$	2.24537	2.23774	0.17092	0.00762
$\Sigma_y(43)$	-0.35355	-0.35161	0.17407	0.00195
$\Sigma_y(44)$	6.25000	6.23379	0.23151	0.01621
ρ_y	0.20000	0.19939	0.01497	0.00061
$\Sigma_x(11)$	4.55000	4.53781	0.18379	0.01219
$\Sigma_x(21)$	0.44907	0.44612	0.09253	0.00296
$\Sigma_x(22)$	2.70000	2.69447	0.10096	0.00553
ρ_x	0.40000	0.39907	0.01952	0.00093
$\Sigma 1_{xy}(11)$	1.00416	1.18419	0.45267	0.18003
$\Sigma 1_{xy}(12)$	1.56525	1.82970	0.61523	0.26446
$\Sigma 1_{xy}(22)$	1.73925	2.02502	0.67560	0.28576
$\Sigma 1_{xy}(23)$	1.42009	1.66430	0.57283	0.24420
$\Sigma 1_{xy}(31)$	-0.27386	-0.33653	0.21721	0.06267
$\Sigma 1_{xy}(32)$	-0.22361	-0.26029	0.16079	0.03669
$\Sigma 1_{xy}(41)$	0.19365	0.21089	0.16858	0.01724
$\Sigma 1_{xy}(42)$	0.15811	0.18202	0.14543	0.02391
$\rho_1 xy$	0.30000	0.34486	0.10577	0.04486

TABLE 2: Maximum Likelihood Solution for Σ_y, Σ_x, ρ_x, ρ_y, intuitive estimate of Σ_{xy} and Ψ_{xy}, 500 Samples

Sample Size 350	Simulations 1000			
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta - \hat{\Theta})^2)^{1/2}$	$\lvert(\hat{\Theta} - \Theta)\rvert$
$\Sigma_y(11)$	5.5	5.4887061	0.2409443	0.0112939
$\Sigma_y(21)$	3.1754265	3.1764355	0.2086801	0.001009
$\Sigma_y(22)$	6.5	6.4907984	0.275141	0.0092016
$\Sigma_y(31)$	-0.5	-0.485849	0.1916637	0.014151
$\Sigma_y(32)$	3.1754265	3.1693236	0.2281348	0.0061029
$\Sigma_y(33)$	7.5	7.4604265	0.3341316	0.0395735
$\Sigma_y(41)$	0.3535534	0.3495299	0.1844736	0.0040235
$\Sigma_y(42)$	2.2453656	2.2294648	0.2041458	0.0159008
$\Sigma_y(43)$	-0.353553	-0.352193	0.2100673	0.00136
$\Sigma_y(44)$	6.25	6.2122216	0.2747173	0.0377784
ρ_y	0.2	0.199912	0.017486	8.8E-05
$\Sigma_x(11)$	4.55	4.5288238	0.2142948	0.0211762
$\Sigma_x(21)$	0.4490731	0.4441403	0.1141971	0.0049328
$\Sigma_x(22)$	2.7	2.6888746	0.1247928	0.0111254
ρ_x	0.4	0.3991333	0.0232862	0.0008667
$\Sigma 1_{xy}(11)$	1.004158	1.3688871	1.8452039	0.3647291
$\Sigma 1_{xy}(12)$	1.5652476	2.1224049	2.5966623	0.5571573
$\Sigma 1_{xy}(21)$	1.7392527	2.3245121	2.6125382	0.5852594
$\Sigma 1_{xy}(22)$	1.4200939	1.9407226	2.4474881	0.5206287
$\Sigma 1_{xy}(31)$	-0.273861	-0.380666	0.6019791	0.106805
$\Sigma 1_{xy}(32)$	-0.223607	-0.282851	0.3035659	0.059244
$\Sigma 1_{xy}(41)$	0.1936492	0.219135	0.2961896	0.0254858
$\Sigma 1_{xy}(41)$	0.1581139	0.2095802	0.2945439	0.0514663
$\rho 1_{xy}$	0.3	0.3677985	0.1486246	0.0677985

TABLE 3: Maximum Likelihood Solution for Σ_y, Σ_x, ρ_x, ρ_y, intuitive estimate of Σ_{xy} and Ψ_{xy}, 350 samples

Sample Size 200	Simulations 1000			
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta-\hat{\Theta})^2)^{1/2}$	$\lvert(\hat{\Theta}-\Theta)\rvert$
$\Sigma_y(11)$	5.5	5.4711093	0.3202831	0.0288907
$\Sigma_y(21)$	3.1754265	3.1640223	0.2735186	0.0114042
$\Sigma_y(22)$	6.5	6.4743773	0.3820472	0.0256227
$\Sigma_y(31)$	-0.5	-0.495453	0.262255	0.004547
$\Sigma_y(32)$	3.1754265	3.1623281	0.3299204	0.0130984
$\Sigma_y(33)$	7.5	7.465473	0.4487569	0.034527
$\Sigma_y(41)$	0.3535534	0.3534077	0.2348238	0.0001457
$\Sigma_y(42)$	2.2453656	2.2339085	0.2798036	0.0114571
$\Sigma_y(43)$	-0.353553	-0.345304	0.2741676	0.008249
$\Sigma_y(44)$	6.25	6.205943	0.3726603	0.044057
ρ_y	0.2	0.2006511	0.0246555	0.0006511
$\Sigma_x(11)$	4.55	4.5207644	0.2744799	0.0292356
$\Sigma_x(21)$	0.4490731	0.4449863	0.1441758	0.0040868
$\Sigma_x(22)$	2.7	2.6857411	0.163492	0.0142589
ρ_x	0.4	0.398861	0.0322768	0.001139
$\Sigma 1_{xy}(11)$	1.004158	1.7722421	3.9507414	0.7680841
$\Sigma 1_{xy}(12)$	1.5652476	2.7555405	5.7241202	1.1902929
$\Sigma 1_{xy}(21)$	1.7392527	3.0173272	6.3816367	1.2780745
$\Sigma 1_{xy}(22)$	1.4200939	2.524706	5.6338096	1.1046121
$\Sigma 1_{xy}(31)$	-0.273861	-0.474997	1.1771414	0.201136
$\Sigma 1_{xy}(32)$	-0.223607	-0.367542	1.0190742	0.143935
$\Sigma 1_{xy}(41)$	0.1936492	0.282505	0.9106433	0.0888558
$\Sigma 1_{xy}(42)$	0.1581139	0.2573953	0.629088	0.0992814
$\rho 1_{xy}$	0.3	0.3981963	0.2081228	0.0981963

TABLE 4: Maximum Likelihood Solution for Σ_y, Σ_x, ρ_x, ρ_y, intuitive estimate of Σ_{xy} and Ψ_{xy}, 200 samples

Sample Size 100	Simulations 1000			
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta-\hat{\Theta})^2)^{1/2}$	$\lvert(\hat{\Theta}-\Theta)\rvert$
$\Sigma_y(11)$	5.5	5.4525786	0.4581956	0.0474214
$\Sigma_y(21)$	3.1754265	3.1479364	0.4003749	0.0274901
$\Sigma_y(22)$	6.5	6.4140674	0.5559107	0.0859326
$\Sigma_y(31)$	-0.5	-0.475498	0.3650991	0.024502
$\Sigma_y(32)$	3.1754265	3.1443529	0.4558047	0.0310736
$\Sigma_y(33)$	7.5	7.4303896	0.6192218	0.0696104
$\Sigma_y(41)$	0.3535534	0.3716303	0.3494294	0.0180769
$\Sigma_y(42)$	2.2453656	2.2281972	0.4084596	0.0171684
$\Sigma_y(43)$	-0.353553	-0.359171	0.3896021	0.005618
$\Sigma_y(44)$	6.25	6.2055091	0.5261355	0.0444909
ρ_y	0.2	0.1991069	0.0358495	0.0008931
$\Sigma_x(11)$	4.55	4.5198069	0.3987957	0.0301931
$\Sigma_x(21)$	0.4490731	0.4486643	0.2064195	0.0004088
$\Sigma_x(22)$	2.7	2.6721157	0.2311758	0.0278843
ρ_x	0.4	0.39972	0.0427603	0.00028
$\Sigma 1_{xy}(11)$	1.004158	2.4483011	7.4717073	1.4441431
$\Sigma 1_{xy}(12)$	1.5652476	3.9285145	12.417629	2.3632669
$\Sigma 1_{xy}(21)$	1.7392527	4.33286	12.961319	2.5936073
$\Sigma 1_{xy}(22)$	1.4200939	3.5627714	11.956913	2.1426775
$\Sigma 1_{xy}(31)$	-0.273861	-0.739009	2.7171749	0.465148
$\Sigma 1_{xy}(32)$	-0.223607	-0.730624	3.6170008	0.507017
$\Sigma 1_{xy}(41)$	0.1936492	0.3510516	2.6102964	0.1574024
$\Sigma 1_{xy}(42)$	0.1581139	0.3349653	1.4479197	0.1768514
$\rho 1_{xy}$	0.3	0.442836	0.2805866	0.142836

TABLE 5: Maximum Likelihood Solution for Σ_y, Σ_x, ρ_x, ρ_y, intuitive estimate of Σ_{xy} and Ψ_{xy}, 100 samples

For a sample size of 200, 1000 simulations showed 871 converged. The ML estimators had a much smaller MSE and bias than the intuitive estimator. Table 4 shows the results for 200 samples.

For a sample size of 100, 1000 simulations showed 645 converged. The ML estimators had a much smaller MSE and bias than the intuitive estimator. Table 5 shows the results for 100 samples.

For a sample size of 50, 1000 simulations showed convergence. The ML estimators had a much smaller MSE and bias than the intuitive estimator. Table 6 shows the results for 50 samples.

Sample Size 50	Simulations 1000			
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta-\hat{\Theta})^2)^{1/2}$	$\lvert(\hat{\Theta}-\Theta)\rvert$
$\Sigma_y(11)$	5.50000	5.38572	0.63838	0.11428
$\Sigma_y(21)$	3.17543	3.12812	0.53214	0.04730
$\Sigma_y(22)$	6.50000	6.37219	0.75108	0.12781
$\Sigma_y(31)$	-0.50000	-0.48519	0.52173	0.01481
$\Sigma_y(32)$	3.17543	3.10831	0.63379	0.06712
$\Sigma_y(33)$	7.50000	7.35267	0.87776	0.14734
$\Sigma_y(41)$	0.35355	0.35696	0.46468	0.00340
$\Sigma_y(42)$	2.24537	2.20256	0.55177	0.04281
$\Sigma_y(43)$	-0.35355	-0.35449	0.55607	0.00094
$\Sigma_y(44)$	6.25000	6.12585	0.74061	0.12415
ρ_y	0.20000	0.19905	0.04838	0.00095
$\Sigma_x(11)$	4.55000	4.43166	0.56430	0.11834
$\Sigma_x(21)$	0.44907	0.44469	0.29620	0.00438
$\Sigma_x(21)$	2.70000	2.65309	0.32754	0.04691
ρ_x	0.40000	0.39930	0.06129	0.00070
$\Sigma 1_{xy}(11)$	1.00416	2.21826	4.54869	1.21410
$\Sigma 1_{xy}(12)$	1.56525	4.25959	10.03225	2.69434
$\Sigma 1_{xy}(21)$	1.73925	4.30226	9.20887	2.56301
$\Sigma 1_{xy}(22)$	1.42009	3.71180	8.74069	2.29170
$\Sigma 1_{xy}(31)$	-0.27386	-0.67614	4.06503	0.40228
$\Sigma 1_{xy}(32)$	-0.22361	-0.68435	4.28020	0.46074
$\Sigma 1_{xy}(41)$	0.19365	0.63663	3.26688	0.44298
$\Sigma 1_{xy}(42)$	0.15811	0.63454	3.27403	0.47643
$\rho 1_{xy}$	0.30000	0.46671	0.35035	0.16671

TABLE 6: Maximum Likelihood Solution for Σ_y, Σ_x, ρ_x, ρ_y, intuitive estimate of Σ_{xy} and Ψ_{xy}, 50 samples

II.8 USE OF TRANSFORMATIONS TO ESTIMATE $\Psi_{YX} \otimes \Sigma_{YX}$

The intuitive method did not always converge. Two main points were learned from the failures of the intuitive method. The first was that while the solution did not always converge, the first iteration always produced a result very close to the actual answer. The second was that the SVN (2008)'s solution always converged. These two facts were used to create a series of transformations to help solve the problem of estimation of Σ_{xy}. The proposed transformations move the cross product term into the upper left corner of the partitioned matrix. The work by SNV (2008) then can be used to estimate the matrix Ψ_{xy}. Once Ψ_{xy} is estimated, the intuitive method can be used for one iteration to gain an estimate of Σ_{xy}. Towards this end, let

$$\mathbf{Y} = (y_{11}, \ldots, y_{q1}, \ldots, y_{1t}, \ldots, y_{qt})'$$

and

$$\mathbf{X} = (x_{11}, \ldots, x_{p1}, \ldots, x_{1t}, \ldots, x_{pt})'$$

be random vectors such that the vector

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Y} \\ \mathbf{X} \end{pmatrix}_{(p+q)t \times 1} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \Psi_y \otimes \Sigma_y & \Psi_{yx} \otimes \Sigma_{yx} \\ \Psi_{xy} \otimes \Sigma_{xy} & \Psi_x \otimes \Sigma_x \end{bmatrix} \right). \tag{15}$$

Let $\mathbf{Z}_1, ..., \mathbf{Z}_N$ be a random sample from (15). Then the sample variance covariance matrix can be computed as $\frac{1}{N}\sum_{i=1}^{N} \mathbf{Z}\mathbf{Z}'$.

This sample variance covariance matrix can be partitioned into four block matrices so that each block can be used to estimate the corresponding Σ and Ψ matrices. Note that

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{Z}\mathbf{Z}' = \begin{pmatrix} \frac{1}{N}\sum_{i=1}^{N} \mathbf{Y}\mathbf{Y}' & \frac{1}{N}\sum_{i=1}^{N} \mathbf{Y}\mathbf{X}' \\ \frac{1}{N}\sum_{i=1}^{N} \mathbf{X}\mathbf{Y}' & \frac{1}{N}\sum_{i=1}^{N} \mathbf{X}\mathbf{X}' \end{pmatrix}.$$

The objective here is to estimate the matrices in the structure $\Psi_{xy} \otimes \Sigma_{xy}$ using

$$C = \frac{1}{N}\sum_{i=1}^{N} \mathbf{Y}\mathbf{X}'.$$

As mentioned the intuitive estimators in (14) did not produce reliable results. Hence to estimate the matrices in $\Psi_{xy} \otimes \Sigma_{xy}$ matrix we are adopting alternative

approach. The plan is to move this matrix to the upper left corner in the partition so that the method of SNV (2008) can be used. First, the upper right hand corner of the partitioned section can be isolated by using a series of transformations inspired by Tan (1973). Thus we first transform the data to make the diagonals of the covariance matrix an identity matrix as:

$$\begin{bmatrix} (\hat{\Psi}_y \otimes \hat{\Sigma}_y)^{-1/2} & 0 \\ 0' & (\hat{\Psi}_x \otimes \hat{\Sigma}_x)^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{Y} \\ \mathbf{X} \end{bmatrix}. \tag{16}$$

The normal distribution was used to approximate the asymptotic distribution of the transformation in (16). This results in a distribution that can be approximated as:

$$\underset{\sim}{\overset{approximately}{}} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} I & D \\ D' & I \end{pmatrix}\right].$$

In this, the matrix D equals

$$\hat{D} = (\hat{\Psi}_x \otimes \hat{\Sigma}_x)^{-1/2} C (\hat{\Psi}_y \otimes \hat{\Sigma}_y)^{-1/2}.$$

We once again use the transformation:

$$\begin{bmatrix} I & D \\ D' & I \end{bmatrix}^{-1/2} \begin{bmatrix} (\hat{\Psi}_y \otimes \hat{\Sigma}_y)^{-1/2} & 0 \\ 0' & (\hat{\Psi}_x \otimes \hat{\Sigma}_x)^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{Y} \\ \mathbf{X} \end{bmatrix}.$$

Multiplying the multivariate normal distribution vector by the inverse of its covariance matrix results in a multivariate T distribution (Tan, 1973). This converges asymptotically to a normal distribution with mean zero and the identity matrix as the variance covariance matrix.

$$\underset{\sim}{\overset{approximately}{}} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} \hat{I} & 0 \\ 0 & \hat{I} \end{pmatrix}\right].$$

The D matrix is then moved to the upper left corner with the transformation as:

$$\begin{bmatrix} 0 & \hat{D} \\ \hat{D}' & 0 \end{bmatrix} \begin{bmatrix} I & \hat{D} \\ \hat{D}' & I \end{bmatrix}^{-1/2} \begin{bmatrix} (\hat{\Psi}_y \otimes \hat{\Sigma}_y)^{-1/2} & 0 \\ 0' & (\hat{\Psi}_x \otimes \hat{\Sigma}_x)^{-1/2} \end{bmatrix} \begin{bmatrix} Y \\ X \end{bmatrix}$$

$$\underset{\sim}{\overset{approximately}{}} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}\begin{pmatrix} \widehat{DD}' & 0 \\ 0 & \widehat{D'D} \end{pmatrix}\right].$$

The solution converges to,

$$\widehat{D}_{\Psi_{xy}}\hat{D}'_{\Psi_{xy}} \otimes \widehat{D}_{\Sigma_{xy}}\hat{D}'_{\Sigma_{xy}}.$$

The spectral decomposition gives,

$$\widehat{D}_{\Psi_{xy}}\hat{D}'_{\Psi_{xy}} = \hat{U}\Delta^2\hat{U}'.$$

For a positive definite $\widehat{D}_{\Psi_{xy}}$, based on (Harville, 1997, p. 538), the eigenvalues of unique eigenvectors corresponding to positive definite matrices are unique. If non-unique eigenvalues are present then the sum of the matrices formed from non-unique eigenvectors corresponding to the same eigenvalues are unique. Hence the estimate

$$\widehat{D}_{\Psi_{xy}} = U\Delta U'$$

is unique. Then we have

$$\widehat{C}_{Base\Psi_{xy}} = \widehat{A}^{1/2}_{\Psi_y}\widehat{D}_{\Psi_{xy}}\widehat{B}^{1/2}_{\Psi_y}.$$

Note that the matrix $\widehat{C}_{Base\Psi_{xy}}$ can be either AR(1) or a positive definite matrix or any other structured matrix. If $\widehat{C}_{Base\Psi_{xy}}$ is a positive definite matrix then the solution is,

$$\hat{C}_{\Psi_{xy}} = Diag(\hat{C}_{Base\Psi_{xy}})^{-1/2}\hat{C}_{Base\Psi_{xy}}Diag(\hat{C}_{Base\Psi_{xy}})^{-1/2}.$$

Thus, we have found $\hat{C}_{\Psi_{xy}}$ as an alternative estimate of Ψ_{xy}. Shaping the resulting matrix into the AR(1) solution was discussed above.

Once the estimator of Ψ_{xy} is determined as above under the relevant restrictions, one iteration can produce an estimate of Σ_{xy} as:

$$\hat{\Sigma}_{xy} = \frac{\sum_{i=1}^{N} X_{ic}\hat{C}^{-1}_{\Psi_{xy}}Y'_{ic}}{Nt}.$$

II.8.1 Transformation Simulation Results

Using these transformations, all values converged. The tables below show simulation results. They include the intuitive estimates for comparison purposes.

In Table 7, the results of the transformation estimates are labeled $\Sigma_{xy}(ij)$. All the transformation results converged. The intuitive results based on the simulations that converged are once again denoted $\Sigma 1_{xy}(ij)$ and are included for comparison purposes. The resulting bias and MSE of the transformation estimate equal or exceed the bias and MSE of the existing MLE estimates.

Another goal for this analysis was to extend the results to small sample sizes. Tables 7, 8, 9, 10, and 11 show the results of this analysis. Based on the tables, it appears that the transformation solution equaled or exceeded the MLE bias and MSE.

Additional simulations are listed in tables 12, 13, 14, 15, 16, 17, 18, 19, 20, and 21.

Sample Size 500	Simulations 1000			
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta-\hat{\Theta})^2)^{1/2}$	$\lvert(\hat{\Theta}-\Theta)\rvert$
$\Sigma_y(11)$	5.50000	5.49091	0.19355	0.00909
$\Sigma_y(21)$	3.17543	3.16755	0.16785	0.00788
$\Sigma_y(22)$	6.50000	6.48315	0.23614	0.01685
$\Sigma_y(31)$	-0.50000	-0.50270	0.15944	0.00270
$\Sigma_y(32)$	3.17543	3.16691	0.19235	0.00852
$\Sigma_y(33)$	7.50000	7.48407	0.27176	0.01593
$\Sigma_y(41)$	0.35355	0.34927	0.14595	0.00428
$\Sigma_y(42)$	2.24537	2.23774	0.17092	0.00762
$\Sigma_y(43)$	-0.35355	-0.35161	0.17407	0.00195
$\Sigma_y(44)$	6.25000	6.23379	0.23151	0.01621
ρ_y	0.20000	0.19939	0.01497	0.00061
$\Sigma_x(11)$	4.55000	4.53781	0.18379	0.01219
$\Sigma_x(21)$	0.44907	0.44612	0.09253	0.00296
$\Sigma_x(21)$	2.70000	2.69447	0.10096	0.00553
ρ_x	0.40000	0.39907	0.01952	0.00093
$\Sigma_{xy}(11)$	1.00416	1.00466	0.13027	0.00050
$\Sigma_{xy}(12)$	1.56525	1.56436	0.10695	0.00088
$\Sigma_{xy}(21)$	1.73925	1.73518	0.15310	0.00408
$\Sigma_{xy}(22)$	1.42009	1.41771	0.11413	0.00238
$\Sigma_{xy}(31)$	-0.27386	-0.27980	0.14816	0.00594
$\Sigma_{xy}(32)$	-0.22361	-0.22382	0.11304	0.00022
$\Sigma_{xy}(41)$	0.19365	0.18905	0.13287	0.00460
$\Sigma_{xy}(42)$	0.15811	0.15158	0.10581	0.00654
ρ_{xy}	0.30000	0.30217	0.04504	0.00217
$\Sigma 1_{xy}(11)$	1.00416	1.18419	0.45267	0.18003
$\Sigma 1_{xy}(12)$	1.56525	1.82970	0.61523	0.26446
$\Sigma 1_{xy}(21)$	1.73925	2.02502	0.67560	0.28576
$\Sigma 1_{xy}(22)$	1.42009	1.66430	0.57283	0.24420
$\Sigma 1_{xy}(31)$	-0.27386	-0.33653	0.21721	0.06267
$\Sigma 1_{xy}(32)$	-0.22361	-0.26029	0.16079	0.03669
$\Sigma 1_{xy}(41)$	0.19365	0.21089	0.16858	0.01724
$\Sigma 1_{xy}(42)$	0.15811	0.18202	0.14543	0.02391
$\rho_1 xy$	0.30000	0.34486	0.10577	0.04486

TABLE 7: Maximum Likelihood Solution for Σ_y, Σ_x, $\rho_x = .4$, $\rho_y = .2$, transformation estimate of $\rho_{xy} = .3$ and Σ_{xy}, 500 samples

Sample Size 350	Simulations 1000	Converge 975		
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta-\hat{\Theta})^2)^{1/2}$	$\lvert(\hat{\Theta}-\Theta)\rvert$
$\Sigma_y(11)$	5.5	5.4887061	0.2409443	0.0112939
$\Sigma_y(21)$	3.1754265	3.1764355	0.2086801	0.001009
$\Sigma_y(22)$	6.5	6.4907984	0.275141	0.0092016
$\Sigma_y(31)$	-0.5	-0.485849	0.1916637	0.014151
$\Sigma_y(32)$	3.1754265	3.1693236	0.2281348	0.0061029
$\Sigma_y(33)$	7.5	7.4604265	0.3341316	0.0395735
$\Sigma_y(41)$	0.3535534	0.3495299	0.1844736	0.0040235
$\Sigma_y(42)$	2.2453656	2.2294648	0.2041458	0.0159008
$\Sigma_y(43)$	-0.353553	-0.352193	0.2100673	0.00136
$\Sigma_y(44)$	6.25	6.2122216	0.2747173	0.0377784
ρ_y	0.2	0.199912	0.017486	8.8E-05
$\Sigma_x(11)$	4.55	4.5288238	0.2142948	0.0211762
$\Sigma_x(21)$	0.4490731	0.4441403	0.1141971	0.0049328
$\Sigma_x(22)$	2.7	2.6888746	0.1247928	0.0111254
ρ_x	0.4	0.3991333	0.0232862	0.0008667
$\Sigma_{xy}(11)$	1.004158	1.0028332	0.1646205	0.0013248
$\Sigma_{xy}(12)$	1.5652476	1.5647384	0.1348201	0.0005092
$\Sigma_{xy}(21)$	1.7392527	1.736294	0.1856855	0.0029587
$\Sigma_{xy}(22)$	1.4200939	1.422322	0.1464912	0.0022281
$\Sigma_{xy}(31)$	-0.273861	-0.26664	0.1800344	0.007221
$\Sigma_{xy}(32)$	-0.223607	-0.217351	0.1380549	0.006256
$\Sigma_{xy}(41)$	0.1936492	0.1883952	0.1634804	0.005254
$\Sigma_{xy}(42)$	0.1581139	0.15349	0.1283671	0.0046239
ρ_{xy}	0.3	0.3013871	0.0505144	0.0013871
$\Sigma 1_{xy}(11)$	1.004158	1.3688871	1.8452039	0.3647291
$\Sigma 1_{xy}(12)$	1.5652476	2.1224049	2.5966623	0.5571573
$\Sigma 1_{xy}(21)$	1.7392527	2.3245121	2.6125382	0.5852594
$\Sigma 1_{xy}(22)$	1.4200939	1.9407226	2.4474881	0.5206287
$\Sigma 1_{xy}(31)$	-0.273861	-0.380666	0.6019791	0.106805
$\Sigma 1_{xy}(32)$	-0.223607	-0.282851	0.3035659	0.059244
$\Sigma 1_{xy}(41)$	0.1936492	0.219135	0.2961896	0.0254858
$\Sigma 1_{xy}(42)$	0.1581139	0.2095802	0.2945439	0.0514663
$\rho 1_{xy}$	0.3	0.3677985	0.1486246	0.0677985

TABLE 8: Maximum Likelihood Solution for Σ_y, Σ_x, $\rho_x = .4$, $\rho_y = .2$, transformation estimate of $\rho_{xy} = .3$ and Σ_{xy}, 350 samples

Sample Size 200	Simulations 1000	Converge		
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta - \hat{\Theta})^2)^{1/2}$	$\lvert(\hat{\Theta} - \Theta)\rvert$
$\Sigma_y(11)$	5.5	5.4711093	0.3202831	0.0288907
$\Sigma_y(21)$	3.1754265	3.1640223	0.2735186	0.0114042
$\Sigma_y(22)$	6.5	6.4743773	0.3820472	0.0256227
$\Sigma_y(31)$	-0.5	-0.495453	0.262255	0.004547
$\Sigma_y(32)$	3.1754265	3.1623281	0.3299204	0.0130984
$\Sigma_y(33)$	7.5	7.465473	0.4487569	0.034527
$\Sigma_y(41)$	0.3535534	0.3534077	0.2348238	0.0001457
$\Sigma_y(42)$	2.2453656	2.2339085	0.2798036	0.0114571
$\Sigma_y(43)$	-0.353553	-0.345304	0.2741676	0.008249
$\Sigma_y(44)$	6.25	6.205943	0.3726603	0.044057
ρ_y	0.2	0.2006511	0.0246555	0.0006511
$\Sigma_x(11)$	4.55	4.5207644	0.2744799	0.0292356
$\Sigma_x(21)$	0.4490731	0.4449863	0.1441758	0.0040868
$\Sigma_x(22)$	2.7	2.6857411	0.163492	0.0142589
ρ_x	0.4	0.398861	0.0322768	0.001139
$\Sigma_{xy}(11)$	1.004158	1.0062308	0.2075244	0.0020728
$\Sigma_{xy}(12)$	1.5652476	1.5662221	0.17157	0.0009745
$\Sigma_{xy}(21)$	1.7392527	1.7315366	0.2421632	0.0077161
$\Sigma_{xy}(22)$	1.4200939	1.4190792	0.1827922	0.0010147
$\Sigma_{xy}(31)$	-0.273861	-0.276378	0.2391721	0.002517
$\Sigma_{xy}(32)$	-0.223607	-0.221834	0.1912619	0.001773
$\Sigma_{xy}(41)$	0.1936492	0.1914619	0.2176764	0.0021873
$\Sigma_{xy}(42)$	0.1581139	0.1531945	0.1671407	0.0049194
ρ_{xy}	0.3	0.3031572	0.0601464	0.0031572
$\Sigma 1_{xy}(11)$	1.004158	1.7722421	3.9507414	0.7680841
$\Sigma 1_{xy}(12)$	1.5652476	2.7555405	5.7241202	1.1902929
$\Sigma 1_{xy}(21)$	1.7392527	3.0173272	6.3816367	1.2780745
$\Sigma 1_{xy}(22)$	1.4200939	2.524706	5.6338096	1.1046121
$\Sigma 1_{xy}(31)$	-0.273861	-0.474997	1.1771414	0.201136
$\Sigma 1_{xy}(32)$	-0.223607	-0.367542	1.0190742	0.143935
$\Sigma 1_{xy}(41)$	0.1936492	0.282505	0.9106433	0.0888558
$\Sigma 1_{xy}(42)$	0.1581139	0.2573953	0.629088	0.0992814
$\rho 1_{xy}$	0.3	0.3981963	0.2081228	0.0981963

TABLE 9: Maximum Likelihood Solution for Σ_y, Σ_x, $\rho_x = .4$, $\rho_y = .2$, transformation estimate of $\rho_{xy} = .3$ and Σ_{xy}, 200 sample

Sample Size 100	Simulations 1000	Converge		
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta-\hat{\Theta})^2)^{1/2}$	$\lvert(\hat{\Theta}-\Theta)\rvert$
$\Sigma_y(11)$	5.5	5.4525786	0.4581956	0.0474214
$\Sigma_y(21)$	3.1754265	3.1479364	0.4003749	0.0274901
$\Sigma_y(22)$	6.5	6.4140674	0.5559107	0.0859326
$\Sigma_y(31)$	-0.5	-0.475498	0.3650991	0.024502
$\Sigma_y(32)$	3.1754265	3.1443529	0.4558047	0.0310736
$\Sigma_y(33)$	7.5	7.4303896	0.6192218	0.0696104
$\Sigma_y(41)$	0.3535534	0.3716303	0.3494294	0.0180769
$\Sigma_y(42)$	2.2453656	2.2281972	0.4084596	0.0171684
$\Sigma_y(43)$	-0.353553	-0.359171	0.3896021	0.005618
$\Sigma_y(44)$	6.25	6.2055091	0.5261355	0.0444909
ρ_y	0.2	0.1991069	0.0358495	0.0008931
$\Sigma_x(11)$	4.55	4.5198069	0.3987957	0.0301931
$\Sigma_x(21)$	0.4490731	0.4486643	0.2064195	0.0004088
$\Sigma_x(22)$	2.7	2.6721157	0.2311758	0.0278843
ρ_x	0.4	0.39972	0.0427603	0.00028
$\Sigma_{xy}(11)$	1.004158	1.0004119	0.3007671	0.0037461
$\Sigma_{xy}(12)$	1.5652476	1.5498737	0.2434819	0.0153739
$\Sigma_{xy}(21)$	1.7392527	1.7357805	0.3349671	0.0034722
$\Sigma_{xy}(22)$	1.4200939	1.4032583	0.264244	0.0168356
$\Sigma_{xy}(31)$	-0.273861	-0.265667	0.3352885	0.008194
$\Sigma_{xy}(32)$	-0.223607	-0.227258	0.2607316	0.003651
$\Sigma_{xy}(41)$	0.1936492	0.2007334	0.3049106	0.0070842
$\Sigma_{xy}(42)$	0.1581139	0.1641078	0.2403161	0.0059939
ρ_{xy}	0.3	0.3040875	0.0706872	0.0040875
$\Sigma 1_{xy}(11)$	1.004158	2.4483011	7.4717073	1.4441431
$\Sigma 1_{xy}(12)$	1.5652476	3.9285145	12.417629	2.3632669
$\Sigma 1_{xy}(21)$	1.7392527	4.33286	12.961319	2.5936073
$\Sigma 1_{xy}(22)$	1.4200939	3.5627714	11.956913	2.1426775
$\Sigma 1_{xy}(31)$	-0.273861	-0.739009	2.7171749	0.465148
$\Sigma 1_{xy}(32)$	-0.223607	-0.730624	3.6170008	0.507017
$\Sigma 1_{xy}(41)$	0.1936492	0.3510516	2.6102964	0.1574024
$\Sigma 1_{xy}(42)$	0.1581139	0.3349653	1.4479197	0.1768514
$\rho 1_{xy}$	0.3	0.442836	0.2805866	0.142836

TABLE 10: Maximum Likelihood Solution for Σ_y, Σ_x, $\rho_x = .4$, $\rho_y = .2$, transformation estimate of $\rho_{xy} = .3$ and Σ_{xy}, 100 samples

Sample Size 50	Simulations 1000			
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta-\hat{\Theta})^2)^{1/2}$	$\lvert(\hat{\Theta}-\Theta)\rvert$
$\Sigma_y(11)$	5.50000	5.38572	0.63838	0.11428
$\Sigma_y(21)$	3.17543	3.12812	0.53214	0.04730
$\Sigma_y(22)$	6.50000	6.37219	0.75108	0.12781
$\Sigma_y(31)$	-0.50000	-0.48519	0.52173	0.01481
$\Sigma_y(32)$	3.17543	3.10831	0.63379	0.06712
$\Sigma_y(33)$	7.50000	7.35267	0.87776	0.14734
$\Sigma_y(41)$	0.35355	0.35696	0.46468	0.00340
$\Sigma_y(42)$	2.24537	2.20256	0.55177	0.04281
$\Sigma_y(43)$	-0.35355	-0.35449	0.55607	0.00094
$\Sigma_y(44)$	6.25000	6.12585	0.74061	0.12415
ρ_y	0.20000	0.19905	0.04838	0.00095
$\Sigma_x(11)$	4.55000	4.43166	0.56430	0.11834
$\Sigma_x(21)$	0.44907	0.44469	0.29620	0.00438
$\Sigma_x(21)$	2.70000	2.65309	0.32754	0.04691
ρ_x	0.40000	0.39930	0.06129	0.00070
$\Sigma_{xy}(11)$	1.00416	0.99012	0.40499	0.01404
$\Sigma_{xy}(12)$	1.56525	1.55870	0.34521	0.00655
$\Sigma_{xy}(21)$	1.73925	1.70115	0.46706	0.03810
$\Sigma_{xy}(22)$	1.42009	1.41299	0.36314	0.00710
$\Sigma_{xy}(31)$	-0.27386	-0.27898	0.48118	0.00512
$\Sigma_{xy}(32)$	-0.22361	-0.21468	0.36631	0.00893
$\Sigma_{xy}(41)$	0.19365	0.19743	0.43212	0.00378
$\Sigma_{xy}(42)$	0.15811	0.17390	0.34203	0.01579
ρ_{xy}	0.30000	0.30405	0.07828	0.00405
$\Sigma 1_{xy}(11)$	1.00416	2.21826	4.54869	1.21410
$\Sigma 1_{xy}(12)$	1.56525	4.25959	10.03225	2.69434
$\Sigma 1_{xy}(21)$	1.73925	4.30226	9.20887	2.56301
$\Sigma 1_{xy}(22)$	1.42009	3.71180	8.74069	2.29170
$\Sigma 1_{xy}(31)$	-0.27386	-0.67614	4.06503	0.40228
$\Sigma 1_{xy}(32)$	-0.22361	-0.68435	4.28020	0.46074
$\Sigma 1_{xy}(41)$	0.19365	0.63663	3.26688	0.44298
$\Sigma 1_{xy}(42)$	0.15811	0.63454	3.27403	0.47643
$\rho 1_{xy}$	0.30000	0.46671	0.35035	0.16671

TABLE 11: Maximum Likelihood Solution for Σ_y, Σ_x, $\rho_x = .4$, $\rho_y = .2$, transformation estimate of $\Psi_{xy} = .3$ and Σ_{xy}, 50 samples

Sample Size 500	Simulations 1000			
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta-\hat{\Theta})^2)^{1/2}$	$\lvert(\hat{\Theta}-\Theta)\rvert$
$\Sigma_y(11)$	5.5	5.497456	0.2102784	0.00909
$\Sigma_y(21)$	3.1754265	3.1824973	0.181729	0.00788
$\Sigma_y(22)$	6.5	6.4945398	0.2481227	0.01685
$\Sigma_y(31)$	-0.5	-0.490892	0.1682839	0.0027
$\Sigma_y(32)$	3.1754265	3.1716143	0.1982771	0.00852
$\Sigma_y(33)$	7.5	7.4754364	0.2738458	0.01593
$\Sigma_y(41)$	0.3535534	0.3537135	0.156369	0.00428
$\Sigma_y(42)$	2.2453656	2.2282898	0.1672169	0.00762
$\Sigma_y(43)$	-0.353553	-0.359163	0.1761539	0.00195
$\Sigma_y(44)$	6.25	6.2310437	0.2451567	0.01621
ρ_y	0.5	0.5001469	0.0122433	0.00061
$\Sigma_x(11)$	4.55	4.5418406	0.1979322	0.01219
$\Sigma_x(21)$	0.4490731	0.4462306	0.092643	0.00296
$\Sigma_x(21)$	2.7	2.6924514	0.121602	0.00553
ρ_x	0.7	0.6993468	0.0130086	0.00093
$\Sigma_{xy}(11)$	1.004158	1.0140237	0.1333523	0.0005
$\Sigma_{xy}(12)$	1.5652476	1.5760578	0.1274859	0.00088
$\Sigma_{xy}(21)$	1.7392527	1.7555498	0.1588138	0.00408
$\Sigma_{xy}(22)$	1.4200939	1.4338413	0.1353264	0.00238
$\Sigma_{xy}(31)$	-0.273861	-0.276152	0.1436449	0.00594
$\Sigma_{xy}(32)$	-0.223607	-0.221007	0.1139707	0.00022
$\Sigma_{xy}(41)$	0.1936492	0.1922868	0.133658	0.0046
$\Sigma_{xy}(42)$	0.1581139	0.1601105	0.0981006	0.00654
ρ_{xy}	0.6	0.6031658	0.0331045	0.00217

TABLE 12: Maximum Likelihood Solution for Σ_y, Σ_x, $\rho_x = .7$, $\rho_y = .5$, transformation estimate of $\Psi_{xy} = .6$, Σ_{xy}, 500 samples

Sample Size 350	Simulations 1000			
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta-\hat{\Theta})^2)^{1/2}$	$\lvert(\hat{\Theta}-\Theta)\rvert$
$\Sigma_y(11)$	5.5	5.4810299	0.2496884	0.0189701
$\Sigma_y(21)$	3.1754265	3.1720532	0.2122072	0.0033733
$\Sigma_y(22)$	6.5	6.4816379	0.2846793	0.0183621
$\Sigma_y(31)$	-0.5	-0.485198	0.1913265	0.014802
$\Sigma_y(32)$	3.1754265	3.1646824	0.2297575	0.0107441
$\Sigma_y(33)$	7.5	7.4496956	0.3436327	0.0503044
$\Sigma_y(41)$	0.3535534	0.3490526	0.1842461	0.0045008
$\Sigma_y(42)$	2.2453656	2.2263096	0.2053112	0.019056
$\Sigma_y(43)$	-0.353553	-0.351765	0.2099211	0.001788
$\Sigma_y(44)$	6.25	6.2034697	0.2865738	0.0465303
ρ_y	0.5	0.4992722	0.0144013	0.0007278
$\Sigma_x(11)$	4.55	4.5212693	0.2415551	0.0287307
$\Sigma_x(21)$	0.4490731	0.4434266	0.1146822	0.0056465
$\Sigma_x(21)$	2.7	2.6844134	0.1393211	0.0155866
ρ_x	0.7	0.6987151	0.0149236	0.0012849
$\Sigma_{xy}(11)$	1.004158	1.0084197	0.1698664	0.0042617
$\Sigma_{xy}(12)$	1.5652476	1.5730801	0.1510284	0.0078325
$\Sigma_{xy}(21)$	1.7392527	1.7456283	0.1990721	0.0063756
$\Sigma_{xy}(22)$	1.4200939	1.4298761	0.157933	0.0097822
$\Sigma_{xy}(31)$	-0.273861	-0.268075	0.1791795	0.005786
$\Sigma_{xy}(32)$	-0.223607	-0.218595	0.1371458	0.005012
$\Sigma_{xy}(41)$	0.1936492	0.1896067	0.1621896	0.0040425
$\Sigma_{xy}(42)$	0.1581139	0.1543918	0.1276023	0.0037221
ρ_{xy}	0.6	0.6032167	0.0371701	0.0032167

TABLE 13: Maximum Likelihood Solution for Σ_y, Σ_x, $\rho_x = .7$, $\rho_y = .5$, transformation estimate of $\rho_{xy} = .6$, Σ_{xy}, 350 samples

Sample Size 200	Simulations 1000			
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta-\hat{\Theta})^2)^{1/2}$	$\lvert(\hat{\Theta}-\Theta)\rvert$
$\Sigma_y(11)$	5.5	5.460057	0.3333726	0.039943
$\Sigma_y(21)$	3.1754265	3.1574783	0.2762595	0.0179482
$\Sigma_y(22)$	6.5	6.4612826	0.3939938	0.0387174
$\Sigma_y(31)$	-0.5	-0.494673	0.262293	0.005327
$\Sigma_y(32)$	3.1754265	3.1559783	0.3335234	0.0194482
$\Sigma_y(33)$	7.5	7.4505928	0.4667418	0.0494072
$\Sigma_y(41)$	0.3535534	0.3525436	0.2343228	0.0010098
$\Sigma_y(42)$	2.2453656	2.2292787	0.2803032	0.0160869
$\Sigma_y(43)$	-0.353553	-0.344566	0.2733578	0.008987
$\Sigma_y(44)$	6.25	6.1935118	0.3871119	0.0564882
ρ_y	0.5	0.4994802	0.0197859	0.0005198
$\Sigma_x(11)$	4.55	4.5080946	0.3134081	0.0419054
$\Sigma_x(21)$	0.4490731	0.4435843	0.144832	0.0054888
$\Sigma_x(21)$	2.7	2.6782526	0.1862663	0.0217474
ρ_x	0.7	0.697912	0.0207387	0.002088
$\Sigma_{xy}(11)$	1.004158	1.0141367	0.2139515	0.0099787
$\Sigma_{xy}(12)$	1.5652476	1.5782876	0.1883618	0.01304
$\Sigma_{xy}(21)$	1.7392527	1.7452219	0.2561211	0.0059692
$\Sigma_{xy}(22)$	1.4200939	1.429939	0.1942096	0.0098451
$\Sigma_{xy}(31)$	-0.273861	-0.278291	0.2382108	0.00443
$\Sigma_{xy}(32)$	-0.223607	-0.223631	0.1905037	2.4E-05
$\Sigma_{xy}(41)$	0.1936492	0.1928939	0.2163289	0.0007553
$\Sigma_{xy}(42)$	0.1581139	0.1541716	0.1661991	0.0039423
ρ_{xy}	0.6	0.6053189	0.044396	0.0053189

TABLE 14: Maximum Likelihood Solution for Σ_y, Σ_x, $\rho_x = .7$, $\rho_y = .5$, transformation estimate of $\rho_{xy} = .6$, Σ_{xy}, 200 samples

Sample Size 100	Simulations 1000			
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta-\hat{\Theta})^2)^{1/2}$	$\lvert(\hat{\Theta}-\Theta)\rvert$
actual mean mse $\Sigma_y(11)$	5.5	5.4260702	0.4784037	0.0739298
$\Sigma_y(21)$	3.1754265	3.13262	0.4068378	0.0428065
$\Sigma_y(22)$	6.5	6.3827564	0.5812693	0.1172436
$\Sigma_y(31)$	-0.5	-0.473144	0.3636832	0.026856
$\Sigma_y(32)$	3.1754265	3.1288726	0.4600244	0.0465539
$\Sigma_y(33)$	7.5	7.3937446	0.6431054	0.1062554
$\Sigma_y(41)$	0.3535534	0.3699242	0.3478751	0.0163708
$\Sigma_y(42)$	2.2453656	2.2174955	0.4118784	0.0278701
$\Sigma_y(43)$	-0.353553	-0.357346	0.3879035	0.003793
$\Sigma_y(44)$	6.25	6.1751332	0.548247	0.0748668
ρ_y	0.5	0.4973483	0.0290149	0.0026517
$\Sigma_x(11)$	4.55	4.5048372	0.4448621	0.0451628
$\Sigma_x(21)$	0.4490731	0.4469569	0.2074079	0.0021162
$\Sigma_x(21)$	2.7	2.6629995	0.2590287	0.0370005
ρ_x	0.7	0.6978319	0.0275209	0.0021681
$\Sigma_{xy}(11)$	1.004158	1.0114324	0.3063041	0.0072744
$\Sigma_{xy}(12)$	1.5652476	1.5671074	0.2590982	0.0018598
$\Sigma_{xy}(21)$	1.7392527	1.7548912	0.3538823	0.0156385
$\Sigma_{xy}(22)$	1.4200939	1.4184933	0.2751917	0.0016006
$\Sigma_{xy}(31)$	-0.273861	-0.267722	0.3345054	0.006139
$\Sigma_{xy}(32)$	-0.223607	-0.229939	0.261472	0.006332
$\Sigma_{xy}(41)$	0.1936492	0.203133	0.3045123	0.0094838
$\Sigma_{xy}(42)$	0.1581139	0.1658289	0.2404585	0.007715
ρ_{xy}	0.6	0.6070024	0.0529681	0.0070024

TABLE 15: Maximum Likelihood Solution for Σ_y, Σ_x, $\rho_x = .7$, $\rho_y = .5$, transformation estimate of $\rho_{xy} = .6$, Σ_{xy}, 100 samples

Sample Size 50	Simulations 1000			
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta-\hat{\Theta})^2)^{1/2}$	$\lvert(\hat{\Theta}-\Theta)\rvert$
$\Sigma_y(11)$	5.5	5.3364285	0.6661971	0.1635715
$\Sigma_y(21)$	3.1754265	3.0993952	0.5392218	0.0760313
$\Sigma_y(22)$	6.5	6.3134047	0.7774967	0.1865953
$\Sigma_y(31)$	-0.5	-0.480815	0.517831	0.019185
$\Sigma_y(32)$	3.1754265	3.0789193	0.6342282	0.0965072
$\Sigma_y(33)$	7.5	7.2843427	0.9035013	0.2156573
$\Sigma_y(41)$	0.3535534	0.3533904	0.4607988	0.000163
$\Sigma_y(42)$	2.2453656	2.1826858	0.555356	0.0626798
$\Sigma_y(43)$	-0.353553	-0.351376	0.5509576	0.002177
$\Sigma_y(44)$	6.25	6.070414	0.773375	0.179586
ρ_y	0.5	0.4953265	0.0392319	0.0046735
$\Sigma_x(11)$	4.55	4.4019057	0.632911	0.1480943
$\Sigma_x(21)$	0.4490731	0.4419227	0.2960075	0.0071504
$\Sigma_x(21)$	2.7	2.6348693	0.3653719	0.0651307
ρ_x	0.7	0.6953083	0.0394868	0.0046917
$\Sigma_{xy}(11)$	1.004158	1.0037164	0.4097876	0.0004416
$\Sigma_{xy}(12)$	1.5652476	1.579705	0.3640922	0.0144574
$\Sigma_{xy}(21)$	1.7392527	1.725738	0.4851759	0.0135147
$\Sigma_{xy}(22)$	1.4200939	1.4318691	0.3780229	0.0117752
$\Sigma_{xy}(31)$	-0.273861	-0.280552	0.4813859	0.006691
$\Sigma_{xy}(32)$	-0.223607	-0.216387	0.3660882	0.00722
$\Sigma_{xy}(41)$	0.1936492	0.2001897	0.4343178	0.0065405
$\Sigma_{xy}(42)$	0.1581139	0.1755556	0.3422631	0.0174417
ρ_{xy}	0.6	0.6080735	0.0581636	0.0080735

TABLE 16: Maximum Likelihood Solution for Σ_y, Σ_x, $\rho_x = .7$, $\rho_y = .5$, transformation estimate of $\rho_{xy} = .6$, Σ_{xy}, 50 samples

Sample Size 500	Simulations 1000			
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta-\hat{\Theta})^2)^{1/2}$	$\lvert(\hat{\Theta}-\Theta)\rvert$
$\Sigma_y(11)$	5.5	5.4778795	0.2159619	0.0221205
$\Sigma_y(21)$	3.1754265	3.1699448	0.179746	0.0054817
$\Sigma_y(22)$	6.5	6.4827151	0.2542428	0.0172849
$\Sigma_y(31)$	-0.5	-0.49089	0.1595155	0.00911
$\Sigma_y(32)$	3.1754265	3.1616817	0.1946784	0.0137448
$\Sigma_y(33)$	7.5	7.4640043	0.2961479	0.0359957
$\Sigma_y(41)$	0.3535534	0.3552423	0.1530363	0.0016889
$\Sigma_y(42)$	2.2453656	2.240337	0.1805828	0.0050286
$\Sigma_y(43)$	-0.353553	-0.353683	0.1789084	0.00013
$\Sigma_y(44)$	6.25	6.2275419	0.2485854	0.0224581
ρ_y	0.7	0.6992041	0.0089444	0.0007959
$\Sigma_x(11)$	4.55	4.5382513	0.2253272	0.0117487
$\Sigma_x(21)$	0.4490731	0.4497473	0.088836	0.0006742
$\Sigma_x(21)$	2.7	2.6966176	0.132359	0.0033824
ρ_x	0.9	0.8997312	0.0051027	0.0002688
$\Sigma_{xy}(11)$	1.004158	1.0309727	0.1393656	0.0268147
$\Sigma_{xy}(12)$	1.5652476	1.5995125	0.1478584	0.0342649
$\Sigma_{xy}(21)$	1.7392527	1.7765789	0.1847902	0.0373262
$\Sigma_{xy}(22)$	1.4200939	1.4588709	0.1491194	0.038777
$\Sigma_{xy}(31)$	-0.273861	-0.287845	0.1393342	0.013984
$\Sigma_{xy}(32)$	-0.223607	-0.220866	0.1128477	0.002741
$\Sigma_{xy}(41)$	0.1936492	0.1997921	0.1316753	0.0061429
$\Sigma_{xy}(42)$	0.1581139	0.1600122	0.0972805	0.0018983
ρ_{xy}	0.8	0.8057155	0.020209	0.0057155

TABLE 17: Maximum Likelihood Solution for Σ_y, Σ_x, $\rho_x = .9$, $\rho_y = .7$, transformation estimate of $\rho_{xy} = .8$, Σ_{xy}, 500 samples

Sample Size 350	Simulations 1000			
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta-\hat{\Theta})^2)^{1/2}$	$\lvert(\hat{\Theta}-\Theta)\rvert$
$\Sigma_y(11)$	5.5	5.4524711	0.2620899	0.0475289
$\Sigma_y(21)$	3.1754265	3.1483578	0.2145186	0.0270687
$\Sigma_y(22)$	6.5	6.4558712	0.3004566	0.0441288
$\Sigma_y(31)$	-0.5	-0.5018	0.2001472	0.0018
$\Sigma_y(32)$	3.1754265	3.1524237	0.2395107	0.0230028
$\Sigma_y(33)$	7.5	7.4525082	0.341371	0.0474918
$\Sigma_y(41)$	0.3535534	0.3447322	0.1849558	0.0088212
$\Sigma_y(42)$	2.2453656	2.2286763	0.2031017	0.0166893
$\Sigma_y(43)$	-0.353553	-0.350969	0.2071719	0.002584
$\Sigma_y(44)$	6.25	6.216812	0.2941629	0.033188
ρ_y	0.7	0.6988288	0.0104107	0.0011712
$\Sigma_x(11)$	4.55	4.5323967	0.2634224	0.0176033
$\Sigma_x(21)$	0.4490731	0.4481981	0.1084027	0.000875
$\Sigma_x(21)$	2.7	2.6811017	0.1553862	0.0188983
ρ_x	0.9	0.8992556	0.0061393	0.0007444
$\Sigma_{xy}(11)$	1.004158	1.0294186	0.1673844	0.0252606
$\Sigma_{xy}(12)$	1.5652476	1.5934765	0.1693836	0.0282289
$\Sigma_{xy}(21)$	1.7392527	1.7824754	0.2083688	0.0432227
$\Sigma_{xy}(22)$	1.4200939	1.4503627	0.166312	0.0302688
$\Sigma_{xy}(31)$	-0.273861	-0.287186	0.170476	0.013325
$\Sigma_{xy}(32)$	-0.223607	-0.230191	0.1305907	0.006584
$\Sigma_{xy}(41)$	0.1936492	0.1991799	0.154785	0.0055307
$\Sigma_{xy}(42)$	0.1581139	0.1634754	0.1225106	0.0053615
ρ_{xy}	0.8	0.8060741	0.0230246	0.0060741

TABLE 18: Maximum Likelihood Solution for Σ_y, Σ_x, $\rho_x = .9$, $\rho_y = .7$, transformation estimate of $\rho_{xy} = 8$, Σ_{xy}, 350 samples

Sample Size 200	Simulations 1000			
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta-\hat{\Theta})^2)^{1/2}$	$\lvert(\hat{\Theta}-\Theta)\rvert$
$\Sigma_y(11)$	5.5	5.4576495	0.3478309	0.0423505
$\Sigma_y(21)$	3.1754265	3.1632318	0.2795845	0.0121947
$\Sigma_y(22)$	6.5	6.4630494	0.4072938	0.0369506
$\Sigma_y(31)$	-0.5	-0.496903	0.252731	0.003097
$\Sigma_y(32)$	3.1754265	3.1432533	0.3217733	0.0321732
$\Sigma_y(33)$	7.5	7.4216936	0.4845413	0.0783064
$\Sigma_y(41)$	0.3535534	0.357276	0.2404264	0.0037226
$\Sigma_y(42)$	2.2453656	2.2318616	0.2704161	0.013504
$\Sigma_y(43)$	-0.353553	-0.348646	0.2732901	0.004907
$\Sigma_y(44)$	6.25	6.1761242	0.3955597	0.0738758
ρ_y	0.7	0.6982924	0.0143742	0.0017076
$\Sigma_x(11)$	4.55	4.5292582	0.336215	0.0207418
$\Sigma_x(21)$	0.4490731	0.4496254	0.1386017	0.0005523
$\Sigma_x(21)$	2.7	2.6774044	0.2034799	0.0225956
ρ_x	0.9	0.8994233	0.0076369	0.0005767
$\Sigma_{xy}(11)$	1.004158	1.0579392	0.2199889	0.0537812
$\Sigma_{xy}(12)$	1.5652476	1.6217339	0.2145326	0.0564863
$\Sigma_{xy}(21)$	1.7392527	1.8072079	0.2774193	0.0679552
$\Sigma_{xy}(22)$	1.4200939	1.4699227	0.2154127	0.0498288
$\Sigma_{xy}(31)$	-0.273861	-0.29672	0.2238377	0.022859
$\Sigma_{xy}(32)$	-0.223607	-0.230088	0.1753793	0.006481
$\Sigma_{xy}(41)$	0.1936492	0.2152343	0.2049283	0.0215851
$\Sigma_{xy}(42)$	0.1581139	0.1613138	0.1625379	0.0031999
ρ_{xy}	0.8	0.8088641	0.0268323	0.0088641

TABLE 19: Maximum Likelihood Solution for Σ_y, Σ_x, $\rho_x = .9$, $\rho_y = .7$, transformation estimate of $\rho_{xy} = .8$, Σ_{xy}, 200 samples

Sample Size 100	Simulations 1000			
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta - \hat{\Theta})^2)^{1/2}$	$\lvert(\hat{\Theta} - \Theta)\rvert$
$\Sigma_y(11)$	5.5	5.4084021	0.4954795	0.0915979
$\Sigma_y(21)$	3.1754265	3.1332523	0.4136412	0.0421742
$\Sigma_y(22)$	6.5	6.3884465	0.6026699	0.1115535
$\Sigma_y(31)$	-0.5	-0.487368	0.3614045	0.012632
$\Sigma_y(32)$	3.1754265	3.1156084	0.4633934	0.0598181
$\Sigma_y(33)$	7.5	7.3297653	0.6894374	0.1702347
$\Sigma_y(41)$	0.3535534	0.3494352	0.3453219	0.0041182
$\Sigma_y(42)$	2.2453656	2.2019964	0.4040923	0.0433692
$\Sigma_y(43)$	-0.353553	-0.342688	0.3935451	0.010865
$\Sigma_y(44)$	6.25	6.1174265	0.5647586	0.1325735
ρ_y	0.7	0.6964844	0.0208703	0.0035156
$\Sigma_x(11)$	4.55	4.4361779	0.5012065	0.1138221
$\Sigma_x(21)$	0.4490731	0.4456007	0.2091596	0.0034724
$\Sigma_x(21)$	2.7	2.6432643	0.3080681	0.0567357
ρ_x	0.9	0.8977719	0.0120884	0.0022281
$\Sigma_{xy}(11)$	1.004158	1.060472	0.3203957	0.056314
$\Sigma_{xy}(12)$	1.5652476	1.6330437	0.2852269	0.0677961
$\Sigma_{xy}(21)$	1.7392527	1.8163562	0.3851663	0.0771035
$\Sigma_{xy}(22)$	1.4200939	1.4926379	0.298534	0.072544
$\Sigma_{xy}(31)$	-0.273861	-0.281931	0.3311458	0.00807
$\Sigma_{xy}(32)$	-0.223607	-0.23692	0.2440435	0.013313
$\Sigma_{xy}(41)$	0.1936492	0.2026673	0.2885638	0.0090181
$\Sigma_{xy}(42)$	0.1581139	0.1734867	0.2262316	0.0153728
ρ_{xy}	0.8	0.8124478	0.0327053	0.0124478

TABLE 20: Maximum Likelihood Solution for Σ_y, Σ_x, $\rho_x = .9$, $\rho_y = .7$, transformation estimate of $\rho_{xy} = .8$, Σ_{xy}, 100 samples

Sample Size 50	Simulations 1000			
Parameter	Θ	$\hat{\Theta}$	$(E(\Theta-\hat{\Theta})^2)^{1/2}$	$\lvert(\hat{\Theta}-\Theta)\rvert$
$\Sigma_y(11)$	5.5	5.2449571	0.7141923	0.2550429
$\Sigma_y(21)$	3.1754265	3.0244643	0.5758214	0.1509622
$\Sigma_y(22)$	6.5	6.2175313	0.8461018	0.2824687
$\Sigma_y(31)$	-0.5	-0.508366	0.5187708	0.008366
$\Sigma_y(32)$	3.1754265	3.0246369	0.6474253	0.1507896
$\Sigma_y(33)$	7.5	7.1787759	0.9485797	0.3212241
$\Sigma_y(41)$	0.3535534	0.3542595	0.470308	0.0007061
$\Sigma_y(42)$	2.2453656	2.1597801	0.5567591	0.0855855
$\Sigma_y(43)$	-0.353553	-0.350286	0.5595164	0.003267
$\Sigma_y(44)$	6.25	5.9845576	0.806369	0.2654424
ρ_y	0.7	0.6925648	0.0292255	0.0074352
$\Sigma_x(11)$	4.55	4.3774264	0.6891424	0.1725736
$\Sigma_x(21)$	0.4490731	0.4422599	0.2910104	0.0068132
$\Sigma_x(21)$	2.7	2.5980262	0.4143961	0.1019738
ρ_x	0.9	0.8959648	0.0170313	0.0040352
$\Sigma_{xy}(11)$	1.004158	1.0703362	0.4358383	0.0661782
$\Sigma_{xy}(12)$	1.5652476	1.6251627	0.3743967	0.0599151
$\Sigma_{xy}(21)$	1.7392527	1.8176896	0.5232614	0.0784369
$\Sigma_{xy}(22)$	1.4200939	1.4814802	0.3902393	0.0613863
$\Sigma_{xy}(31)$	-0.273861	-0.300954	0.4596601	0.027093
$\Sigma_{xy}(32)$	-0.223607	-0.227072	0.3322915	0.003465
$\Sigma_{xy}(41)$	0.1936492	0.2146725	0.4206123	0.0210233
$\Sigma_{xy}(42)$	0.1581139	0.1685907	0.3344183	0.0104768
ρ_{xy}	0.8	0.8136635	0.0375795	0.0136635

TABLE 21: Maximum Likelihood Solution for Σ_y, Σ_x, $\rho_x = .9$, $\rho_y = .7$, transformation estimate of $\rho_{xy} = .8$, Σ_{xy}, 50 samples

CHAPTER III

HYPOTHESIS TESTING IN REPEATED CCA

III.1 FIVE HYPOTHESIS OF INTEREST

Estimating the covariance matrix with a Kronecker product covariance structure has many advantages. Most importantly, the Kronecker product covariance structure allows the data to conform to the subject matter expert understanding. An additional advantage is that using the structure greatly reduces the number of parameters required for estimation.

For comparison purposes, in the standard CCA with no time components, where the variance covariance matrix is

$$\begin{pmatrix} \Sigma_y & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_x \end{pmatrix},$$

the total number of parameters that require estimation of the covariance matrices is determined as follows.

- The Σ_{yy} matrix has $q(q+1)/2$ unique parameters.
- The Σ_{xx} matrix has $p(p+1)/2$ unique parameters.
- The Σ_{xy} matrix has pq unique parameters for the cross correlations.
- The μ_y vector has q parameters.
- and μ_x vector has p parameters.

This gives a total of $p(p+1)/2 + q(q+1)/2 + pq$ parameters for the covariance matrix and $p+q$ parameters for the mean.

When sets of variables are recorded over time the number of parameters required quickly increases. For t time periods and no assumed covariance structure, the number of parameters are as shown here in the matrix;

$$\begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,(p+q)t} \\ \sigma_{2,1} & \sigma_{2,2} & \cdots & \sigma_{2,(p+q)t} \\ \vdots & \ddots & \ddots & \vdots \\ \sigma_{(p+q)t,1} & \sigma_{(p+q)t,2)} & \cdots & \sigma_{(p+q)t,(p+q)t} \end{pmatrix}. \tag{17}$$

The number of parameters required for estimation of this covariance matrix are $\frac{(p+q)t(t(p+q)+1)}{2}$. An example is shown below to illustrated the amount of data required to estimate the parameters. The the X vector has $p = 2$ data values and, the Y vector has $q = 4$ values, with $t = 3$ time units. This results in a covariance matrix with $\frac{(2+4)3((2+4)3+1)}{2} = 171$ parameters. Estimating the mean increases it to 171+18=189 parameters. To have at least one degree of freedom for testing, 190 complete data sets are required with each data set having 2 X values and 4 Y values observed over 3 time periods.

A series of covariance structures are shown below. Each covariance matrix represents a potential hypothesis the researcher can test in an effort to reduce the model covariance structure. The covariances are numbered from I to V and provide a logical sequential choice for how to reduce the modeled covariance structure from the complete unrestricted covariance matrix labeled I down to the traditional covariance matrix assuming independence between the populations in time labeled V. Tests can be derived for matrices with a Kronecker product structure by using a similar methodology as that outlined below.

III.1.1 Variance Covariance Structures

Variance Covariance Matrices	
I	$\begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,(p+q)t} \\ \sigma_{2,1} & \sigma_{2,2} & \cdots & \sigma_{2,(p+q)t} \\ \vdots & \ddots & \ddots & \vdots \\ \sigma_{(p+q)t,1} & \sigma_{(p+q)t,2)} & \cdots & \sigma_{(p+q)t,(p+q)t} \end{pmatrix}$
II	$\begin{pmatrix} \Psi_y \otimes \Sigma_y & \Psi_{yx} \otimes \Sigma_{yx} \\ \Psi_{xy} \otimes \Sigma_{xy} & \Psi_x \otimes \Sigma_x \end{pmatrix}$
III	$\begin{pmatrix} \Psi_y \otimes \Sigma_y & \Psi_{yx} \otimes \Sigma_{yx} \\ \Psi_{xy} \otimes \Sigma_{xy} & I_x \otimes \Sigma_x \end{pmatrix}$, or , $\begin{pmatrix} I_y \otimes \Sigma_y & \Psi_{yx} \otimes \Sigma_{yx} \\ \Psi_{xy} \otimes \Sigma_{xy} & \Psi_x \otimes \Sigma_x \end{pmatrix}$
IV	$\begin{pmatrix} \Psi \otimes \Sigma_y & \Psi \otimes \Sigma_{yx} \\ \Psi \otimes \Sigma_{xy} & \Psi \otimes \Sigma_x \end{pmatrix}$
V	$\begin{pmatrix} I \otimes \Sigma_y & I \otimes \Sigma_{yx} \\ I \otimes \Sigma_{xy} & I \otimes \Sigma_x \end{pmatrix}$

The first matrix in the sequence is the full unrestricted covariance matrix given in (17). This matrix is used only if no other matrix will accommodate the data structure. Since the matrix has no special data structure, all model parameters of

the covariance matrix must be estimated. As we have illustrated, covariance matrix I would require the most data to estimate.

Besides having to estimate a large number of parameters, perhaps this structure is not very practical either. In many applications especially when the data are repeated, it is natural to have some structures as the covariances of the repeated measures. For examples, in multivariate time series and repeated measures data, the autoregressive and structure is common, and in that case the correlation with will likely decrease over time.

The covariance structures below integrate subject matter expert opinion into the modeling. These matrices are designed to use the correlation of multivariate observations over time. Taking advantage of these structures greatly reduces the number of covariance parameters that need to be estimated as well.

III.1.2 Variance Covariance Matrix II

$$\text{Variance Covariance Matrix II: } \begin{pmatrix} \Psi_y \otimes \Sigma_y & \Psi_{yx} \otimes \Sigma_{yx} \\ \Psi_{xy} \otimes \Sigma_{xy} & \Psi_x \otimes \Sigma_x \end{pmatrix}$$

shows the basic Kronecker product structure. It allows a different time element correlation for the X values, the Y values, and the cross correlation of the two. That is, the X values have their own time correlation and the Y values have their own time correlation. The XY cross correlation values also have their own separate time correlation. For example the time correlation for the X values at time one and the X values at time two may be 0.1, while the Y values at time one to time two correlation may be 0.3. The X values at time one and Y values at time two may have a 0.2 correlation.

Further, this structure has much fewer parameters than the full unrestricted covariance matrix I. Specifically,

- The $\Psi_y \otimes \Sigma_{yy}$ matrix has $[\frac{(t+1)t}{2} - 1] + \frac{q(q+1)}{2}$ unique parameters.
- The $\Psi_x \otimes \Sigma_{xx}$ matrix has $[\frac{(t+1)t}{2} - 1] + \frac{p(p+1)}{2}$ unique parameters.
- The $\Psi_{xy} \otimes \Sigma_{xy}$ matrix has $[\frac{(t+1)t}{2} - 1] + pq$ unique parameters for the cross correlations.
- This gives a total of $[\frac{(t+1)t}{2} - 1] + \frac{q(q+1)}{2} + [\frac{(t+1)t}{2} - 1] + \frac{p(p+1)}{2} + [\frac{(t+1)t}{2} - 1] + pq$ unique parameters to estimate.

- If the AR(1) structure is assumed for the time structure, all $\frac{(t+1)t}{2}$ values above reduce to value of 1.

For the values $p = 4$, $q = 2$, and $t = 3$, with an AR(1) correlation matrix the covariance structure requires 24 parameters. This is 147 parameters less than the unrestricted model. As illustrated in simulation, the high number of degrees of freedom may contribute to inflating the significance level (probability of type 1 error) in the hypothesis tests using the likelihood ratio statistic, χ^2 approximation.

It is important to note that while the cross covariance values may vary, they must remain within a specified range in order for the resulting covariance matrix to be positive definite. Since Chapter II divides up the covariance matrix into partitioned blocks and estimates each block separately, it is possible that the reconstructed total variance covariance matrix is not positive definite. See the omitted covariance matrix below for consequences and an example of this problem.

III.1.3 Variance Covariance Matrix III

$$\text{Var - Cov Matrix III: } \begin{pmatrix} \Psi_y \otimes \Sigma_y & \Psi_{yx} \otimes \Sigma_{yx} \\ \Psi_{xy} \otimes \Sigma_{xy} & I_x \otimes \Sigma_x \end{pmatrix} \text{ or } \begin{pmatrix} I_y \otimes \Sigma_y & \Psi_{yx} \otimes \Sigma_{yx} \\ \Psi_{xy} \otimes \Sigma_{xy} & \Psi_x \otimes \Sigma_x \end{pmatrix}$$

retains the covariance structure in the X values but takes away the time covariance structure for Y values. The model makes intuitive sense since the XY cross correlation time covariance structure remains when only one value set of variables becomes independent. The model philosophy is that one of the data sets is dependent on time while the other is independent. Under the AR(1) time correlation the model contains $\frac{p(p-1)}{2} + \frac{q(q+1)}{2} + pq + 2$ parameters. Either the Y or the X variable can be correlated in time. Hence the two equations structures both have the same number of parameters that require estimation.

The next logical step from hypothesis III would seem to be variance covariance matrix

$$\begin{pmatrix} I_y \otimes \Sigma_y & \Psi_{xy} \otimes \Sigma_{yx} \\ \Psi_{xy} \otimes \Sigma_{xy} & I_x \otimes \Sigma_x \end{pmatrix}.$$

While the number of parameters $\frac{p(p+1)}{2} + \frac{q(q+1)}{2} + 1$ is 1 less than covariance matrix III, this covariance model is more complicated in the cross product correlation matrix. The model assumes that each individual Y and X at time i are independent of Y and X at time $j \neq i$. However, it assumes the cross product term is correlated in time.

Hence, the Y values at time one and the X values at time two are related but the X values at time one and the X values at time two are not related.

That is, this structure assumes that both the X and Y values have no time correlation structure but the cross correlation of the two have a correlation structure. In symbols this model assumes $\Psi_x = I$ and $\Psi_y = I$, while $\Psi_{xy} \neq I$. While this variance covariance may seem the next logical step to test, the resulting matrix under the null hypothesis is usually not positive definite because it results in negative eigenvalues. The fact that the covariance is unlikely to be positive definite is a sign that real data covariance structures will probably not assume this form of covariance.

III.1.4 Variance Covariance Matrix IV

Now, a hypothesis testing path is structure I $\Longrightarrow$ structure II $\Longrightarrow$ structure III $\Longrightarrow$ structure V. An alternative path is covariance I $\Longrightarrow$ structure II $\Longrightarrow$ structure IV $\Longrightarrow$ structure V. This path involves testing to determine if all AR(1) Ψ matrices are equal as in equation 18, that is

$$\rho_x = \rho_y = \rho_{xy}, \text{assuming unstructured or AR(1) structure.}$$

$$\text{Variance Covariance Matrix IV: } \begin{pmatrix} \Psi_t \otimes \Sigma_y & \Psi_t \otimes \Sigma_{yx} \\ \Psi_t \otimes \Sigma_{xy} & \Psi_t \otimes \Sigma_x \end{pmatrix} \tag{18}$$

Under covariance matrix IV, each set of X and Y variables have the same AR(1) correlation structure in time. This means only the ρ parameter is required to be estimated. The physical interpretation of this covariance structure is that all the data has the same time correlation component. Hence, time has the same effect on all observations simultaneously. This structure is most intuitive in that it assumes the time correlation component is constant. Hence the mechanisms that are occurring in time play the same role in the X values, the Y values, and the cross correlation of the two. The corresponding covariance matrix is shown in equation (18).

Estimating variance covariance IV can be performed using the maximum likelihood estimator defined above. First, the data must be rearranged such that the Y and X values are grouped by time. Let

$$z_{.i} = (y_{1i}, \ldots, y_{qi}, x_{1i}, \ldots, x_{pi})', i = 1, 2, ..., t.$$

Then arrange the $z_{.i}$ vectors as

$$\begin{bmatrix} z_{.1} \\ \vdots \\ z_{.t} \end{bmatrix} \sim N \left[\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \psi_{11}\Sigma & \psi_{12}\Sigma & \dots & \psi_{1t}\Sigma \\ \psi_{21}\Sigma & \psi_{22}\Sigma & \dots & \psi_{2t}\Sigma \\ \vdots & \vdots & \ddots & \dots \\ \psi_{t1}\Sigma & \psi_{t2}\Sigma & \dots & \psi_{tt}\Sigma \end{pmatrix} \right]$$

$$\sim N \left[\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \Psi \otimes \Sigma \end{pmatrix} \right]$$

where Ψ_{ij} is the i,j th element of Ψ_t and $\Sigma = \begin{pmatrix} \Psi_t \otimes \Sigma_y & \Psi_t \otimes \Sigma_{yx} \\ \Psi_t \otimes \Sigma_{xy} & \Psi_t \otimes \Sigma_x \end{pmatrix}$.

The parameters Ψ and Σ can be estimated using the previous work by SNV (2008). Note that the Ψ values all have the same subscript. Assuming an AR(1) structure for Ψ this solution has one less parameter than III, namely, $\frac{p(p+1)}{2} + \frac{q(q+1)}{2} + 1$.

III.1.5 Variance Covariance Matrix V

Regardless of which path the researcher takes, whether it is Covariance I $\Longrightarrow$ Covariance II $\Longrightarrow$ Covariance III $\Longrightarrow$ Covariance V or Covariance I $\Longrightarrow$ Covariance II $\Longrightarrow$ Covariance IV $\Longrightarrow$ Covariance V, the final reduction in the sequence is variance covariance matrix V.

A series of hypothesis can be tested to reduce the model to the simplest possible covariance structure. The simpler the covariance structure, the fewer parameters it will require and the more degrees of freedom left for testing. Fewer parameters also means less data required to estimate variance covariance parameters. Covariance V represents the simplest model. It has an identity matrix for the time covariance component. In this model the matrix Ψ equals the identity matrix throughout all four partitions,

$$\text{Covariance Matrix V} : \begin{pmatrix} I_y \otimes \Sigma_y & I_{yx} \otimes \Sigma_{yx} \\ I_{xy} \otimes \Sigma_{xy} & I_x \otimes \Sigma_x \end{pmatrix}$$

and it has the least number of parameters namely $\frac{p(p+1)}{2} + \frac{q(q+1)}{2} + pq$ that require estimation.

Covariance V shows no covariance structure across time units. The model assumes that what happens in time unit 1 does not influence time unit 2 or later. This simple

assumption will be discussed in more detail in Chapter IV, and will probably be the model used by a researcher when first faced with repeated data in canonical correlation analysis set up.

III.2 ASYMPTOTIC DISTRIBUTION IN TESTING

The log likelihood ratio statistics was used to determine if the model could be reduced in complexity and hence require less parameters. Let

$$f(Z,\mu,\Sigma) = \frac{1}{2\pi^{(p+q)/2}|\Sigma|^{1/2}} exp \frac{-(\mathbf{Z}-\mu)\Sigma^{-1}(\mathbf{Z}-\mu)}{2},$$

be the loglikelihood function, where

$$\mathbf{Z} = \begin{bmatrix} Y \\ X \end{bmatrix}.$$

Then the loglikelihood ratio statistic is defined by

$$\Lambda = -2\left(logf(Z,\hat{\mu}_o,\hat{\Sigma}_o) - logf(Z,\hat{\mu},\hat{\Sigma})\right) \sim \chi^2_{df}$$

where the degrees of freedom equal the difference in the number of parameters.

The Bartlett Correction for the likelihood ratio given by (Bakewell and Wit, 2007) is $\Lambda_{BC} = \frac{n-2}{n}\Lambda$. Simulation techniques as discussed in Chapter II were used to simulate data from variance covariance matrices II through V.

The results when using the loglikelihood statistic with the Bartlett Correction as per (Bakewell and Wit, 2007) are in the following tables. The left column of the table shows the covariance structure of the sample data. For example, Table 22 shows II which indicates that covariance structure II was assumed for the covariance matrix. The 2nd column denotes sample sizes for the tests. The following sample sizes were used; 500, 350, 200, 100, and, 50. The next five columns show the size of the corresponding tests. The top row shows the test performed in each case. I vs II is the likelihood ratio test of covariance I vs covariance II. Similarly, II vs III uses covariance III as the null hypothesis and determine whether or not the model can be reduced from covariance II to covariance III. This is continued for all tests.

It is important to note that in practice the user will not know the true covariance matrix associated with his data. Since the user of these hypothesis tests will not know what the true underlying covariance structure is, he will continue to test until

hypothesis testing indicates he should stop at a particular covariance. Possible testing paths include I $\Longrightarrow$ II $\Longrightarrow$ III $\Longrightarrow$ V or I $\Longrightarrow$ IV $\Longrightarrow$ V.

For each covariance structure, a simulation was constructed from a normal distribution. The covariance structure used in each simulation is shown in the model column. This was performed 1000 times for each sample size. Tests were performed using an α level of .05. The data was used to calculate the likelihood ratio statistic. The χ^2 distribution was used as an approximate distribution of the test statistics. The tables below indicates how often the null hypothesis was rejected, that is, it shows the observed size of the level of significance for each test. The null hypothesis changes for each column of the tables. Observed tests include variance covariance matrix I versus variance covariance matrix II (I vs II), variance covariance matrix II versus variance covariance matrix III (II vs III), variance covariance matrix III versus variance covariance matrix V (III vs V), or variance covariance matrix II versus variance covariance matrix IV (II vs IV), variance covariance matrix IV versus variance covariance matrix V (IV vs V).

The true matrix used to generate the data is listed in the first column of each table labeled as model. The table below shows the rejection rate for the hypothesis test shown in the top row. Hence the cell in table 22 corresponding to column labeled II vs IV and corresponding to the row with sample size 200 shows a rejection rate of .011. This means out of 1000 samples using the loglikelihood ratio test for testing variance covariance matrix II as the alternative hypothesis versus variance covariance matrix IV as the null hypothesis, there were 11 times the simulation showed a rejection of the null hypothesis. The null hypothesis for the test was variance covariance matrix IV but the data for all tests in table 22 was simulated from the variance covariance model II. Ideally all 1000 tests should have shown rejection of the null hypothesis since the data was generated from a variance covariance matrix of II.

Model	Sample Size	I vs II	II vs III	III vs V	II vs IV	IV vs V
II	500	.115	1.000	1.000	.782	1.000
II	350	.097	1.000	1.000	.290	1.000
II	200	.147	1.000	.999	.011	1.000
II	100	*	1.000	.993	0	1.000
II	50	*	.975	.982	0	1.000

TABLE 22: Rejection rates for 1000 samples, Model II

For covariance matrix II, the test of I vs. II should have shown consistent results.

The rejection proportions were 0.115 at 500 samples, dropped to 0.097 at 350 samples. These values are roughly double the desired 0.05 level for rejecting the null hypothesis. It would appear that the size of tests are biased as shown by the high rejection rates. The observed test sizes increases at 200 samples up to 0.147. There were not enough data points to estimate the type I covariance matrix for 100 data points or below.

A researcher using these tests will have an objective to reduce the covariance matrix to the simplest possible form. One advantage of increasing the matrix simplicity is that the matrix becomes easier to interpret and more intuitive to understand. Another advantage is of course the reduction in the number of parameters required for estimation. Since the practitioner will not know the true covariance matrix he is testing, it is thought he will test all matrices throughout the chain until the statistical testing indicates he should stop reducing the matrix structure.

To illustrate this, Table 22 above has the observed rejection proportions for all tests given the null hypothesis was covariance II. However, even if the true covariance matrix is structure II, the practitioner will continue to test until his attempts at matrix simplification are rejected. Hence, Table 22 shows the rejection proportion for all tests of interest. The table shows the observed proportion of times the testing procedure will prevent the researcher from reducing the matrix too far.

Table 22 shows the proportion of the tests rejected when testing to move from unrestricted covariance I to covariance II. The values range from .097 to .147 under covariance matrix II. This indicates a good chance of moving to the correct covariance matrix structure. After reducing to variance covariance II, the researcher will attempt to reduce the variance further down to covariance matrix III or IV. Attempts to reduce the variance covariance matrix to III result in rejection proportion 1.00 to .975, which indicates a low probability of reducing the covariance matrix to the incorrect structure III. If the matrix is incorrectly reduced to III, testing results in a rejection proportion 1.000 to .975 to avoid reducing the matrix to covariance hypothesis V. The IV path has a higher probability of incorrect reduction. For large samples, the tests show rejection proportions of .782 to .011 when attempting to reduce the matrix to form IV. Small samples yielded 0 proportion rejection when testing II vs IV. Hence if the sample size is 50, there is a high probability the matrix will be incorrectly assumed to be of form IV. However, the proportion of tests allowing the further incorrect reduction to V from IV was very small. No estimates were reduced from IV to V.

Reducing the model from II to IV shows the most drastic dependency on sample size. For 500 samples, the observed probability of rejection was 78.2%. Hence there is a good chance the model will not be reduced to the incorrect covariance matrix IV. However, as the sample size decreased, the variability in the estimates increased and the test was not powerful enough to detect the lack of fit from assuming covariance structure IV. This evidence shows small samples have a high chance of incorrectly reducing the covariance matrix.

Attempting to reduce the model to covariance V showed better results. There were two possible reasons for this. One is the estimate of IV and V are the maximum likelihood estimates. The other is that hypothesis V is so restricting that the estimates clearly show it is an incorrect reduction of the covariance matrix.

Model	Sample Size	I vs II	II vs III	III vs V	II vs IV	IV vs V
III	500	.137	.109	1.000	1.000	1.000
III	350	.114	.105	1.000	1.000	1.000
III	200	.157	.094	1.000	.994	1.000
III	100	*	.084	1.000	.340	.965
III	50	*	.063	.996	.005	.732

TABLE 23: Rejection rates for 1000 samples, Model III

Table 23 shows the hypothesis tests under the covariance structures III. For large samples, the test shows rejection proportions between 0.114 to 0.157. This indicates that large samples will allow reduction of the covariance matrix from I to II but the rejection rate is almost 3 times the required .05 Testing from II to III shows rejection proportions from 0.105 to 0.063. Attempts to reduce the model further from III to V yielded rejection rate of 1.000 to .996 or 1000 rejections out of 1000 simulations and 996 rejections out of 1000 simulations for sample size 50. Hence the testing procedure has a good chance of reducing the model to III but will most likely stop before it reduces the model all the way down to total independence hypothesis V. For large samples, the likelihood ratio statistic will also prevent reduction of the covariance to IV, but small samples may allow the incorrect reduction.

Table 24 shows the hypothesis tests when the true covariance matrix is IV. The first test was to reduce the model from the unrestricted variance covariance matrix to matrix II. This test showed .113 to .145 for sample sizes 500 to 200. Testing to reduce from covariance II to III showed rejection of the null hypothesis. Hence, it is unlikely the covariance matrix will be reduced to the III hypothesis. Attempts to

Model	Sample Size	I vs II	II vs III	III vs V	II vs IV	IV vs V
IV	500	.113	1.000	.917	0	1.000
IV	350	.145	1.000	.923	0	1.000
IV	200	.145	1.000	.928	0	1.000
IV	100	*	1.000	.889	0	1.000
IV	50	*	1.000	.775	0	1.000

TABLE 24: Rejection rates for 1000 samples, Model IV

reduce the hypothesis from II to IV resulted in no rejections indicating a very high likelihood that hypothesis testing for covariance reduction will end in covariance IV. All hypothesis testing for all samples sizes showed rejection of the null hypothesis when testing reduction from IV to V. Hence it is very likely to reduce the hypothesis to IV and stop reducing the covariance matrix. It should also be noted that the estimate for IV is the maximum likelihood estimate.

Model	Sample Size	I vs II	II vs III	III vs V	II vs IV	IV vs V
V	500	.122	.099	.021	0	.032
V	350	.091	.089	.015	0	.018
V	200	.142	.095	.020	0	.053
V	100	*	.080	.015	0	.038
V	50	*	.051	.023	0	.027

TABLE 25: Rejection rates for 1000 samples, Model V

Table 25 shows rejection proportions for the hypothesis testing sequence when hypothesis V was true. The test indicated rejection proportions between 0.091 and 0.142 for rejecting covariance matrix II. Once again the probability of rejecting the II hypothesis is 2 to 3 times higher than expected. Hence for small samples, it is unlikely the testing procedure will make it past the first reduction of covariance I versus covariance II. However, once the testing procedure proceeds past this initial test, the results are more promising that the procedure will reduce down to the correct covariance structure. Testing from covariance II to covariance III shows 0.050 to 0.099 rejection rate for all samples sizes. Reducing from III to V shows a much less rejection of 0.015 to 0.023. No rejections resulted when testing for reduction from II to IV. Hence this hypothesis was never rejected when the true covariance structure was hypothesis V i.e. the values were independent. Finally, the hypothesis test IV to V was rejected 0.018 to 0.053. Testing hypothesis IV versus V was a

true maximum likelihood test because all the estimates were gained using maximum likelihood functions.

III.3 BOOTSTRAPPING

The hypothesis tests above showed a higher rejection rate than the expected .05. This was most likely due to differences between the MLE and the estimates. While the estimates mentioned in Chapter II perform well estimates only in IV and V are maximum likelihood estimators and hence will yield a χ^2 asymptotic distribution. Since the estimates shown above may not be the maximum likelihood estimates, another methodology was used to approximate confidence intervals. Bootstrapping was used to create tests that give more accurate rejection probabilities. Parametric bootstrapping stimulations based on (Efron and Tibshirani, 1993) theory were used to create hypothesis tests. For each possible covariance structure, a set of 100 simulations were performed. Each of the 100 simulations consisted of 100 bootstrap samples.

The following sequence was followed for the bootstrapping simulations. The sequence stated here shows the parametric bootstrap procedure when testing covariance I vs covariance II. Other tests followed the same pattern. For covariance II, a covariance matrix was assumed as

$$\begin{pmatrix} \Psi_y \otimes \Sigma_y & \Psi_{yx} \otimes \Sigma_{yx} \\ \Psi_{xy} \otimes \Sigma_{xy} & \Psi_x \otimes \Sigma_x \end{pmatrix}.$$

The covariance matrix was used to simulate one initial sample data set termed the initial sample data.

$$\begin{pmatrix} y_{111} & \cdots & y_{p11} & \cdots & y_{1t1} & \cdots & y_{pt1} & x_{111} & \cdots & x_{q11} & \cdots & x_{1t1} & \cdots & x_{qt1} \\ \vdots & & \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ y_{11n} & \cdots & y_{p1n} & \cdots & y_{1tn} & \cdots & y_{ptn} & x_{11n} & \cdots & x_{q1n} & \cdots & x_{1tn} & \cdots & x_{qtn} \end{pmatrix}' \tag{19}$$

The third subscript in (19) denotes the sample number. The sample data was used for two purposes. The first was to estimate the test statistic based on the likelihood ratio. The likelihood ratio statistics was used as a test statistic for the bootstrap samples as well. The loglikelihood ratio statistic is defined by

$$\lambda = -2\left(log(L(Y, X, \mu, \Sigma_{H_0})) - log(L(Y, X, \mu, \Sigma_{H_{(A)}}))\right).$$

This statistic was compared to the statistics obtained using bootstrap samples generated next. The next step was to use the initial sample to estimate the bootstrap variance covariance matrix.

$$\Phi = \begin{pmatrix} \hat{\Psi}_y \otimes \hat{\Sigma}_y & \hat{\Psi}_{yx} \otimes \hat{\Sigma}_{yx} \\ \hat{\Psi}_{xy} \otimes \hat{\Sigma}_{xy} & \hat{\Psi}_x \otimes \hat{\Sigma}_x \end{pmatrix}$$

The bootstrap variance covariance matrix was used to generate B bootstrap samples.

$$y_1^*, y_2^*, \ldots, y_b^*$$
$$x_1^*, x_2^*, \ldots, x_b^*$$

Each bootstrap sample was used to create an estimate of the sample variance covariance matrix.

$$\Phi^* = \begin{pmatrix} \Psi_y^* \otimes \Sigma_y^* & \Psi_{yx}^* \otimes \Sigma_{yx}^* \\ \Psi_{xy}^* \otimes \Sigma_{xy}^* & \Psi_x^* \otimes \Sigma_x^* \end{pmatrix}.$$

Each bootstrap sample was also used to get a test statistics.

$$\lambda_b^* = -2\left(log(L(Y, X, \mu, \Sigma_{H_0^*})) - log(L(Y, X, \mu, \Sigma_{H_{(A)}^*}))\right).$$

λ was compared to the B=100 bootstrap samples of λ_b^*. If λ was greater than 95% of the λ_b^* values the null hypothesis was rejected. This procedure was performed S=100 times to gain an observed size for the bootstrap tests.

The bootstrap procedure defined above was repeated for samples sizes 50, 100, 200, 350, and 500. Null hypothesis structures tested were II, III, IV, and V. A series of tests were performed for each hypothesis. A researcher testing the hypothesis would not know the true underlying hypothesis. Hence he will continue to test starting with I versus II until an attempt to simplify the matrix fails. The bootstrap test table below shows all 5 test rejection rates in an attempt to reduce the null hypothesis from covariance I down to covariance V.

Table 26 above shows the observed size for the bootstrap tests. The data for all tests was generated under the covariance matrix II. In this case, the test to reject the null hypothesis covariance matrix II ranged from 0.08 to 0.06. These values are well within sampling error for the desired 5% value that was expected. Hence the bootstrap procedure seems to give a test of the appropriate size. The next step is to

Model	Sample Size	I vs II	II vs III	III vs V	II vs IV	IV vs V
II	500	.08	1.0	1.0	1.0	1.00
II	350	.07	1.0	1.0	.99	1.00
II	200	.06	1.0	1.0	.98	1.00
II	100	*	.92	1.0	.92	1.00
II	50	*	.95	.96	.66	1.00

TABLE 26: Rejection rates for 100 samples, Model II

see if the testing procedures will further reduce the test below hypothesis II. Since the true underlying distribution used to generate the data was covariance II, it is undesirable to reduce the assumed hypothesis any further. Attempts to reduce the covariance from II to III resulted in rejection proportions from 0.92 to 1.00. This indicates a strong power in the test to detect the incorrect reduction. Attempts to reduce the covariance structure from III to V resulted in rejection proportions from 0.96 up to 1.00. Attempts to reduce the covariance from II to IV for samples 100 and above resulted in rejection rates 0.92 and above. However, small samples results only showed 0.66 chance of detecting the incorrect reduction. Tests to reduce the test from IV to V were always rejected. Hence, it is very likely that this testing procedure will correctly reduce the model from covariance structure I to covariance structure II and no further.

Model	Sample Size	I vs II	II vs III	III vs V	II vs IV	IV vs V
III	500	.06	.05	1.00	1.00	1.00
III	350	.1	.06	1.00	1.00	1.00
III	200	.05	.01	1.00	1.00	1.00
III	100	*	.06	1.00	1.00	.96
III	50	*	.02	1.00	1.00	.77

TABLE 27: Rejection rates for 100 samples, Model III

Table 27 rejection rates are generated from covariance III. Testing to reduce the covariance matrix from I to II shows rejection rates from 0.05 to 0.10. The range is slightly higher than 0.05 target value. However, there is still a good chance of correctly reducing the covariance matrix from I to II. Furthermore, hypothesis testing from II to III has rejection rates of 0.01 to 0.06. Hence, it is likely that the testing procedure will result in a reduction to covariance III. Testing to reduce hypothesis III to V showed 1.00 rejection proportions. Testing to reduce the covariance matrix

from II to IV also resulted in a 1.00 rejection rate. Hence, if the true hypothesis is independent across time intervals, this test is likely to detect that independence.

Model	Sample Size	I vs II	II vs III	III vs V	II vs IV	IV vs V
IV	500	.05	1.00	.94	.04	1.0
IV	350	.1	1.00	.89	.07	1.0
IV	200	.03	1.00	.94	.06	1.0
IV	100	*	1.00	.88	.06	1.0
IV	50	*	1.00	.88	.04	1.0

TABLE 28: Rejection rates for 100 samples, Model IV

Table 28 shows the rejection rates under the hypothesis IV. Testing from I to II gives rejection proportions from 0.03 to 0.10. Testing from II to III gave 1.00 rejection rates. Hence it is unlikely to reduce to the incorrect hypothesis III. Testing hypothesis II vs IV gives expected rejection rate from 0.04 to 0.07. Continued testing from IV to V showed 100% rejection for all samples. Hence, once the testing gives the IV hypothesis, it is unlikely to reduce below that if the true covariance matrix is IV.

Model	Sample Size	I vs II	II vs III	III vs V	II vs IV	IV vs V
V	500	.05	.08	.1	.1	.07
V	350	.10	.08	.03	.07	.07
V	200	.03	.08	.09	.11	.06
V	100	*	.05	.08	.1	.04
V	50	*	.05	.05	.06	.06

TABLE 29: Rejection rates for 100 samples, Model V

Table 29 shows the rejection rates under the covariance V. The initial test reducing from the unstructured covariance I to II shows rejection rates from 0.03 to 0.1. It is important to note that the rejection rates for small samples are not larger than the rejection rates for large samples. For covariance V, it is very likely the researcher will reduce the model from the unstructured case down to covariance II. Further testing from covariance II to covariance III shows rejection rates of 0.05 to 0.08. Once again this is in the acceptable range and yields a good probability the researcher will reduce the covariance structure to III. Testing from III to V yields rejection rates of 0.03 to 0.09. These are close to the desired rejection rates. Taking the other path also yields promising results. Rejection rates for testing II to IV gives rejection rates from 0.06 to 0.11. Hence, it is likely the researcher will reduce the hypothesis to IV. Finally

testing from IV to V gives rejection values of 0.04 to 0.07. Hence using parametric bootstrapping gives reasonable rejection rates for hypothesis V.

Overall parametric bootstrapping combined with the estimator above showed good results. The bootstrap samples above were based on 100 boot strap iterations. Using the normal approximation, a 95% confidence interval for the rejection region when the null hypothesis is true runs from 0.006 to 0.094. Most of the values observed above fell in that range. Furthermore, when the null hypothesis was false, the observed power of the test was high, thus ensuring a low probability of the researcher incorrectly reducing the data to a hypothesis that is too simple and incorrect.

III.4 NASA DATA

Data was acquired from NASA Langley to illustrate the method. The data was used from the Intercontinental Chemical Transport Experiment Phase B (INTEX-B) project. The INTEX-B* was an experiment that aimed to understand gases and aerosol transport on a transcontinental / intercontinental scale. The project also aimed to assess the scale and the impact of such gases on air quality and the overall climate effect. The primary constituents of the experimentation were ozone, aerosols, long-lived greenhouse gases, and all their precursors.

The INTEX experiment was a two phase plan. The first phase INTEX-A was completed in the summer 2004. The second phase INTEX B was completed in the spring of 2006. INTEX experimentation was implemented in coordination with MIRAGE-Mex and, DLR/IMPACT studies. All three projects were performed in coordination with satellite observation studies also performed at NASA.

The data for this report came from the second half of the study INTEX-B. The INTEX-B field study was performed from March 1 to April 30, 2006 carried out approximately 8 weeks. It was performed in two parts. The first part was performed during March 1-31 and focused on Mexico City pollution outflow. The second part (INTEX-B/Part II) was performed during April 1-30 and focused on the Asian pollution inflow. Due to the location, several national and international partners joined in the INTEX-B study. Data used in this dissertation came primarily from March 19 2006 18:34:15 to 19:15:15.

Two aircraft were used in the experiment. One was a NASA DC-8 flown out of Houston, Texas and the other was an NSF/NCAR C-130 from Tampico, Mexico. Both planes had air routes designed to measure pollution outflow from Mexico city.

However, from 18:34:15 to 19:15:15, the two aircraft flew close together in formation, with sensors on each taking atmospheric readings. Preliminary readings of interest were Carbon (CO), Ozone (O3), and Water (H20). Both craft took continuous readings for the 41 minutes. Figure 1 shows the entire flight path of each plane as well as the joint flight path where measurements were taken from each plain simultaneously.

The estimation procedures created in Chapter II were used to estimate the variance covariance matrix of these variables. The objective was to determine if the the covariance structure of the flight data would match subject matter expectations. The data were taken simultaneously for each molecule, experts expected to see a covariance structure

$$\begin{pmatrix} \Psi_y \otimes \Sigma_y & \Psi_{yx} \otimes \Sigma_{yx} \\ \Psi_{xy} \otimes \Sigma_{xy} & \Psi_x \otimes \Sigma_x \end{pmatrix}.$$

The Ψ matrices in these were not AR(1) structure but instead represent the correlation matrix. Examining the data showed the aircraft changed altitude during the flight which drastically changed the element quantities. To compensate for this, data was used only when the aircraft was below 400 meters.

The aircrafts changed altitude roughly 700 seconds into the flight time wherein both planes were within the same airspace. Figure 2 shows the aircraft altitude over time. The change in aircraft altitude caused a large change in the reading level of the molecules of interest. The levels of O_3, CO, H_20, and the aircraft altitude were scaled to the values of zero to one. All Molecular measurements were taken in grams of substance per kilogram of atmosphere. Altitude was measured in meters.

Altitude	O3 DC-8	CO DC-8	H20 DC-8	O3 C-130	CO C-130	H2O C-130
313	35.0317	112.93	2.295522603	33.2	103.3845	3.90521
3992.3	81.85337	222.59	14.55818318	85.2	211.3836	16.5065

TABLE 30: Altitude and Molecule measurements before scaling

Figure 2 shows how the change in aircraft altitude affected the measurements. The H20 measurement dropped drastically as both aircraft increased altitude. CO measurements started around 150 and fell sharply to 112 as the aircraft approached 2500 feet. As the aircraft rose toward 4000 feet, the CO value rose sharply. The O3 values rose steadily every time the aircraft increased altitude.

Lag one differences were calculated for each molecule and the resulting data was analyzed. The lag one difference accounted for the change in molecules caused by

the change in altitude. The data was too close together in time to get an accurate measurement. After the data was thinned there was still an altitude oriented trend in the data. To remove this, a 4 th degree polynomial was fit to the data with change in altitude as the predictor variable and the change in each variable as the response. The residuals were then thinned and used in the canonical correlation analysis.

A thinning procedure was used to sub set the data for analysis. The data was divided into groups of 23 observations. The first, sixth, and eleventh observations were used in the analysis. This left a twelve observation gap between the time values used in each data set. Attempts to reduce the group size to 22 or fewer data points or to reduce the spacing of the observations closer resulted in a singular variance covariance matrix.

After the differencing, removal of the altitude trend, and data thinning, the resulting data may seem difficult to interpret. Physical interpretation of each can be thought of as follows. Differencing allows us to examine the change in each sensor measurement level with respect to the change in the other sensor levels. Removing the trend due to altitude changes simply removes some explainable variation from the calculations. Finally the data is highly correlated, with each second of data being very close to the next second. Thinning allows us to reduce the correlation between samples. The resulting canonical correlation analysis shows how changes in measurements of the DC-8 sensors relate to changes in the measurements of the C-130 sensors.

The covariance matrix testing sequence defined above was performed on the differenced values.

Hypothesis Test	Observed P-value
II $\Longrightarrow$ III	.00
II $\Longrightarrow$ IV	.23
IV $\Longrightarrow$ V	.007

TABLE 31: Testing Results for the NASA data

*Note that due to the low number of data points, the estimate of the unrestricted covariance I was singular and hence the likelihood could not be calculated.

The results using parametric bootstrapping show that hypothesis IV was the most likely covariance structure to have generated the data. Attempts to reduce the covariance to III and V resulted in observed p-values of .007 or less. Hence

the difference values were modeled using a IV matrix. As discussed previously, the covariance estimate for IV is a maximum likelihood estimator.

Σ_yDC-8 O3	Σ_yDC-8 CO	Σ_yDC-8 H20	Σ_xC-130 O3	Σ_x C-130 CO	Σ_xC-130 H20
0.2119047	-0.011894	0.0049709	0.2359842	0.1215785	0.0014622
-0.011894	0.5828666	-0.01056	0.1215785	7.2159748	0.0053734
0.0049709	-0.01056	0.0053191	0.0014622	0.0053734	0.011764

TABLE 32: Covariance IV Estimates for NASA DC-8 Σ_y and C-130 Σ_x

Σ_{xy} O3	Σ_{xy} CO	Σ_{xy} H20	Ψ O3	Ψ CO	Ψ H20
0.0371183	-0.056899	-0.00096	1	-0.131819	0.0173763
0.0284117	-0.0285	0.002261	-0.131819	1	-0.131819
-0.000511	0.0094509	6.8809E-6	0.0173763	-0.131819	1

TABLE 33: Covariance IV Estimates for NASA DC-8 and C-130 cross covariance Σ_{xy} and correlation Ψ

The Tables 32 and 33 show the maximum likelihood estimates of the variance covariance matrix. The values in Table 32 show the values for the data within a single time point. The values show the CO molecule as measured by the DC-8 aircraft has a much lower variance than the CO molecule as measured by the C-130 aircraft. Other variance and covariances are comparable. Using this estimation method, it was possible to estimate the 27 parameters of the variance covariance structure. If there were no covariance structures assumed, the variance covariance matrix would have required 171 parameters. Since the data set only contained 34 values after thinning, it would have been impossible to estimate a nonsingular variance covariance matrix.

NASA Analysis

Altitude	O3 DC-8	CO DC-8	H20 DC-8	O3 C-130	CO C-130	H2O C-130
313	35.0317	112.93	2.295522603	33.2	103.3845	3.90521
3992.3	81.85337	222.59	14.55818318	85.2	211.3836	16.5065

TABLE 34: Altitude and Molecule measurements before scaling

The advantage of the Kronecker product covariance structure in relation to the canonical correlation is the variables can be broken into two parts. The element correlation and the time correlation.

Canonical Correlation	
1*st* Canonical Correlation	0.201539
2*nd* Canonical Correlation	0.046524
3*rd* Canonical Correlation	0.02943

TABLE 35: Canonical Correlations Within Each Time Period

	1st Variable	2nd Variable	3rd Variable
Ozone DC-8	2.0194259995	-0.222118448	-0.835241353
Carbon Monoxide DC-8	0.4867624881	0.8116618203	0.9402128968
Water DC- 8	-3.115793454	12.605438	-5.528166362

TABLE 36: Raw Canonical Coefficients DC-8

Canonical correlation analysis on the data showed the combination of variables and how they were related. The first canonical correlation gives the highest correlation between the two sets of data. The best correlation between the two sets was 0.201. The first Canonical variable had Ozone from the DC-8 (OD) as the dominate attribute and Ozone from the C-130 (OC) also as the dominate attribute. Both values were positive and close in magnitude. This is an indicator that the variables for Ozone from each sensor are measured at about the same levels and are very closely correlated.

The OD had an 0.88 correlation with the DC-8 first Canonical variable (D1) from table 40 and 0.178 in table 42 with the first C-130 canonical variable (C1). This represents the majority of the correlation between the DC-8 and the C-130 data set.

This is an indication that most of the OD variability is represented in D1. Hence The DC-8 Ozone value is a strong driver for the best relationship with the C-130 data set. Similarly, the OC had a 0.92 correlation with the C1 and a 0.1860 correlation with D1 shown in tables 41 and 43. This means most of the OC is represented in C1.

The water coefficients in the first canonical variable are negative. The fact that both are negative shows they both effects are similar in how they relate to the D1 and C1. However it should be pointed out that the magnitude of that effect is different in each case. the H2O effect in the DC-10 (HD) had a coefficient of -3.11 while the

	1st Variable	2nd Variable	3rd Variable
Ozone C-130	1.9766923865	0.579679451	-0.18
Carbon Monoxide C-130	-0.141696141	0.3110932006	-0.151570441
Water C-130	-0.543495619	3.6870230035	8.4381687378

TABLE 37: Raw Canonical Coefficients C-130

	1st Variable	2nd Variable	3rd Variable
Ozone DC-8	0.9296	-0.1022	-0.3845
Carbon Monoxide DC-8	0.3716	0.6197	0.7178
Water DC- 8	-0.2272	0.9193	-0.4032

TABLE 38: Standardized Canonical Coefficients DC-8

H2O effect in the C-130 (HC) had a coefficient of -0.54 as shown in tables 36 and 37. This could be a sign that the scales of both variables are different in magnitude. However looking at the standardized coefficients which equalizes the scales shows a similar difference. Under the standardized scale HD is -0.2272 while HC is -0.0589 as shown in tables 38 and 39. Hence it would appear the HD plays a much larger role in D1 than HC plays in C1. In other words the water reading on the DC-8 is inversely proportional to the ozone reading on from the C-130 sensor values. While the water reading from the C-130 sensors is negatively correlated similar to the water reading from the DC-8 sensors, the correlation is not as strong.

The most surprising sensor readings were the carbon monoxide readings. The magnitude of the carbon coefficient was small in the first canonical correlations for both aircraft with D1 having a 0.487 and C1 having a -0.141. However the standard canonical coefficients showed 0.372 and -0.381 from tables 38 and 39 which is an indication the magnitude difference may have been caused by different scales used by each sensor. The unexpected characteristic is the sign of the coefficients. One is positive while the other is negative. Hence the DC-8 carbon (CD) measurement shows it is positively correlated with the C-130 data set while the C-130 carbon (CC) measurement shows it is negatively correlated with the DC-8 data set. This is an unfortunate find because both sensors are measuring the same elements in the same airspace, both correlations are expected to go the same direction. The two most likely causes are atmospheric variability and or sensor differences.

	1st Variable	2nd Variable	3rd Variable
Ozone C-130	0.9602	0.2816	-0.0899
Carbon Monoxide C-130	-0.3806	0.8357	-0.4072
Water C-130	-0.0589	0.3999	0.9152

TABLE 39: Standardized Canonical Coefficients C-130

	1st Variable	2nd Variable	3rd Variable
Ozone DC-8	0.8834	0.0129	-0.4685
Carbon Monoxide DC-8	0.3833	0.4488	0.8073
Water DC- 8	-0.1601	0.7867	-0.5962

TABLE 40: Correlation between DC-8 elements and their canonical Variables

The first is that the planes are not flying close enough together and that carbon element densities in the atmosphere vary with only a few meters distance between locations. For safety reasons, the planes are unlikely to fly any more closely together. If this is the case, it is unlikely to be resolved. However future tests may want to install both sets of sensors on the same aircraft. The second likely cause is that one or both of the sensors is not accurately measuring the carbon levels. In this case, a third sensor could be used to help calibrate the other two.

The two other canonical correlations are shown for completeness. However, since the correlation between the variables is 0.04 and 0.02, these were not considered to play an important role in the analysis.

In conclusion, the change in element measurements were modeled with a type II covariance structure from Chapter II. The structure showed each sensor's measurement of ozone matched up on both planes. Measurements of water matched slightly less and carbon measurements showed entirely different results between the planes.

	1st Variable	2nd Variable	3rd Variable
Ozone C-130	0.9231	0.3706	-0.1025
Carbon Monoxide C-130	-0.2923	0.8693	-0.3987
Water C-130	-0.0393	0.4231	0.9052

TABLE 41: Correlation between C-130 elements and their canonical Variables

	1st Variable	2nd Variable	3rd Variable
Ozone DC-8	0.1780	0.0006	-0.0138
Carbon Monoxide DC-8	0.0772	0.0209	0.0238
Water DC- 8	-0.0323	0.0366	-0.0175

TABLE 42: Correlation between DC-8 elements and the Canonical Variables of the C-130

	1st Variable	2nd Variable	3rd Variable
Ozone C-130	0.1860	0.0172	-0.0030
Carbon Monoxide C-130	-0.0589	0.0404	-0.0117
Water C-130	-0.0079	0.0197	0.0266

TABLE 43: Correlation between C-130 elements and the Canonical Variables of the DC-8

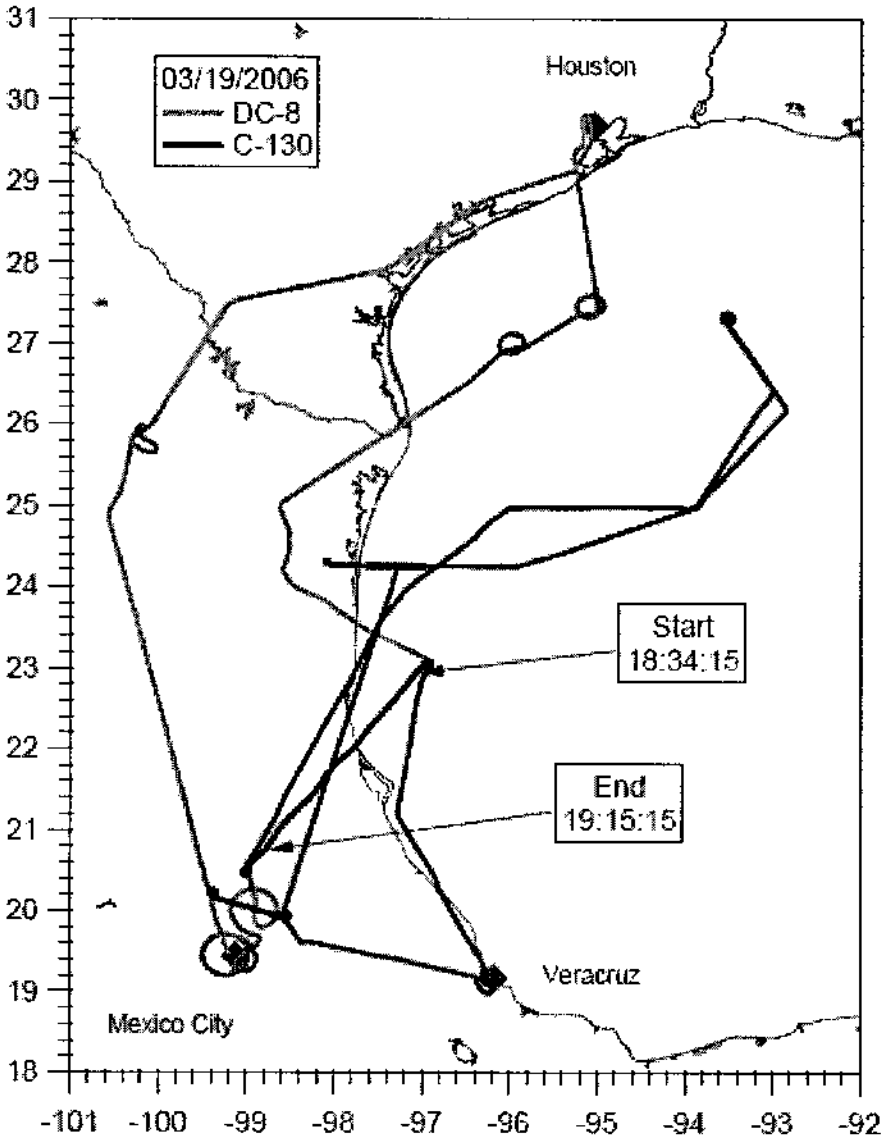

FIG. 1: INTEX-B Airtracks

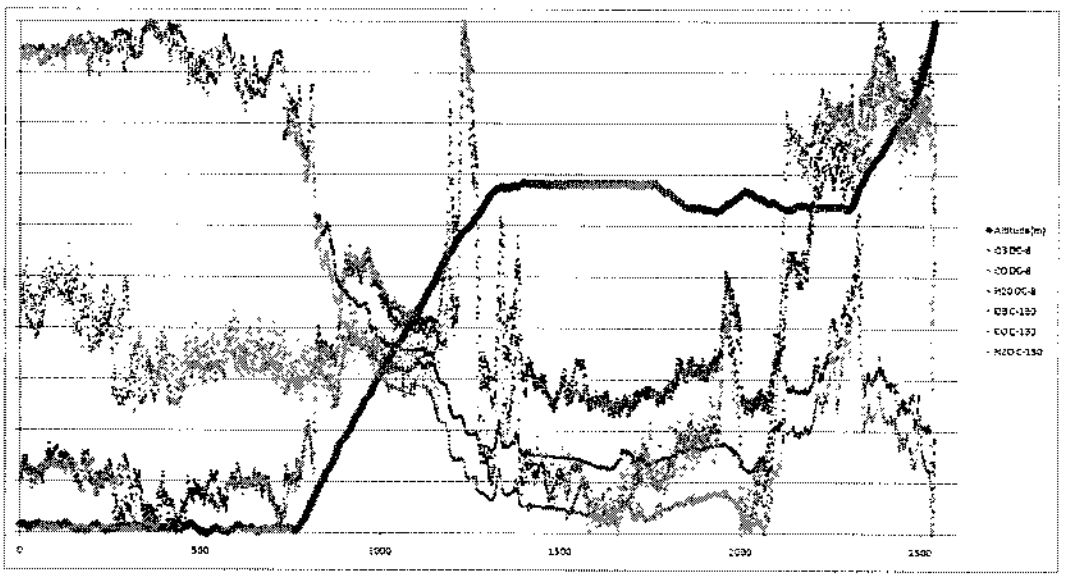

FIG. 2: INTEX-B Airtracks Altitude

CHAPTER IV

A SIMPLIFIED APPROACH

IV.1 ALTERNATIVE DATA STRUCTURE

In the previous chapters, we have successfully used Kronecker product covariance matrix and showed how to perform a CCA of longitudinal data. We also provided some tests for testing Kronecker product covariance structure. However for correctly implementing the tests, bootstrap computations were needed. Bootstrap computations can be quite involved, especially when the dimensions of data are large. In this chapter we look for alternative, but simpler methods for testing for Kronecker product covariance structure.

Under the Kronecker product covariance structure, each $y_j = (y_{1,j}, y_{2,j}, \ldots, y_{q,j})'$, for $j = 1, ..., t$ has the same covariance matrix Σ_y. Similarly, each $x_j = (x_{1,j}, x_{2,j}, \ldots, x_{p,j})'$ for $j = 1, .., t$ has the same variance covariance matrix Σ_x. Further, the covariance matrix between x_j and y_j is the same as Σ_{xy} for all $j = 1, ..., t$. However, the covariance matrices between the x_j and x_j', between y_j and y_j' and between x_j and y_j' are affected by the matrices Ψ_x, Ψ_y, and Ψ_{xy} respectively. The Ψ matrices complicate the model and analysts may avoid calculating the full covariance involving the Ψ matrices in favor of a simpler model.

In this chapter, the canonical correlations are calculated ignoring the time correlation structure. The asymptotic distribution of the canonical correlations is derived which involves use of a duplication matrix. A simple algorithm is provided to generate the duplication matrix.

It is possible for the researcher to analyze the data ignoring the time covariance structure. In this case, the researcher will assume the $\Psi_{xy} = I$, $\Psi_y = I$, and $\Psi_x = I$. Ignoring the time covariance matrix will allow the researcher to conserve data points and simplify the analysis. When the researcher ignores the time covariance structure, he still retains the covariance structure within single time units. Hence the data points $(y_{1,j}, y_{2,j}, \ldots, y_{q,j})$ are still considered to have the same covariance structure within a single time unit. However, the covariance structure between the two time units is removed.

IV.2 USING THE UNSTRUCTURED ESTIMATE

When the time covariance matrix is ignored, the following is assumed for the analysis:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \\ X_1 \\ \vdots \\ X_n \end{bmatrix} \sim N \left[\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \left(\begin{array}{ccc|ccc} \Sigma_y & \dots & 0 & \Sigma_{xy} & \dots & 0 \\ & \ddots & & & \ddots & \\ 0 & \dots & \Sigma_y & 0 & \dots & \Sigma_{xy} \\ \hline \Sigma_{yx} & \dots & 0 & \Sigma_x & \dots & 0 \\ & \ddots & & & \ddots & \\ 0 & \dots & \Sigma_{yx} & 0 & \dots & \Sigma_x \end{array} \right) \right].$$

Without loss of generality, we have assumed means to be zero. The researcher will then pool the data together to increase data and gain better estimates. This method is much simpler than estimating the entire covariance matrix as discussed in Chapter II. However, this does not correctly model the data.

When taken individually, the covariances of each set of Y observations can be separated by time. In this way all Y data taken at time period one can be used to estimate Σ_y. Data taken at time period two can be used to estimate Σ_y and so on for all time points. The same is true for the X observations, where X values observed at time one can give an estimation of Σ_x and X values observed at time point two also give an estimate of Σ_x and so on.

Since this analysis is much simpler than the estimation technique discussed in Chapter II, it can be used as a preliminary method to determine whether or not the data needs to be modeled with the more complicated structure. To investigate this simplification, two values examined were the asymptotic distribution of the canonical variates and the asymptotic distribution of the canonical correlations.

If the true covariance matrices for X and Y (Σ_y, Σ_x, and Σ_{yx}) are known, then the true canonical variates and canonical correlations can be derived.

The t^{th} pair of canonical variate pairs are defined by:

$$U_t = e_t' \Sigma_y^{-1/2} Y \text{ and } V_t = f_t' \Sigma_x^{-1/2} X.$$

The vector e_t is the associated $(q \times 1)$ eigenvectors of

$$\Sigma_y^{-1/2} \Sigma_{yx} \Sigma_x^{-1} \Sigma_{xy} \Sigma_y^{-1/2},$$

and the vector f_t is the associated $(p \times 1)$ eigenvectors of

$$\Sigma_x^{-1/2}\Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx}\Sigma_x^{-1/2}.$$

It is important to note these are the same eigenvectors and eigenvalues e_Σ defined in Chapter II for longitudinal data. Chapter II covariance matrices had 2 parts, the time series correlation structure and the multivariate variable structure as

$$\Psi_y^{-1/2}\Psi_{yx}\Psi_x^{-1}\Psi_{xy}\Psi_y^{-1/2} \otimes \Sigma_y^{-1/2}\Sigma_{yx}\Sigma_x^{-1}\Sigma_{xy}\Sigma_y^{-1/2} \quad \text{and}$$
$$\Psi_x^{-1/2}\Psi_{xy}\Psi_y^{-1}\Psi_{yx}\Psi_x^{-1/2} \otimes \Sigma_x^{-1/2}\Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx}\Sigma_x^{-1/2}.$$

Ignoring the time correlation structure removed the $\Psi_y^{-1/2}\Psi_{yx}\Psi_x^{-1}\Psi_{xy}\Psi_y^{-1/2}$ component from the eigenvector calculations.

IV.3 REDUCING TO THE TIME COVARIANCE FUNCTION

Let

$$\mathbf{a}_k = e_k'\Sigma_y^{-1/2} \text{ and } \mathbf{b}_k = f_k'\Sigma_x^{-1/2}$$

where $\mathbf{a}_k$ and $\mathbf{b}_k$ are the vectors of canonical coefficients for the k th canonical variates. Let

$$a_k = \begin{pmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{qk} \end{pmatrix}, \text{ and } b_k = \begin{pmatrix} b_{1k} \\ b_{2k} \\ \vdots \\ b_{pk} \end{pmatrix}.$$

Under models II through V from Chapter III, it is important to note that each a_k and b_k canonical coefficient vector is identical when each time period is treated individually. The sample values will vary but the theoretical values remain equal.

Let the variables Y_j and X_j be defined before as,

$$\begin{aligned} Y_j &= (y_{1j}, y_{2j}, \ldots, y_{qj}),' \\ X_j &= (x_{1j}, x_{2j}, \ldots, x_{pj}).' \end{aligned} \tag{20}$$

The Y_j vector represents q values of the Y variable recorded at time j and the X_j vector represents the p values of the X variable recorded at time j. The canonical variates can be written as

$$\begin{pmatrix} I_t \otimes a_k' & 0 \\ 0 & I_t \otimes b_k' \end{pmatrix} \begin{pmatrix} Y \\ X \end{pmatrix} = \begin{bmatrix} a_k & \dots & 0 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & a_k & 0 & 0 & 0 \\ 0 & 0 & 0 & b_k & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & b_k \end{bmatrix} \begin{bmatrix} Y_1 \\ \vdots \\ Y_t \\ X_1 \\ \vdots \\ X_t \end{bmatrix}.$$

The transformation results in covariance matrix

$$\Theta = \begin{pmatrix} I_t \otimes a_k' & 0 \\ 0 & I_t \otimes b_k' \end{pmatrix} \begin{pmatrix} \Psi_y \otimes \Sigma_y & \Psi_{yx} \otimes \Sigma_{yx} \\ \Psi_{xy} \otimes \Sigma_{xy} & \Psi_x \otimes \Sigma_x \end{pmatrix} \begin{pmatrix} I_t \otimes a_k' & 0 \\ 0 & I_t \otimes b_k' \end{pmatrix}'$$

$$\Theta = \begin{pmatrix} a_k'\Sigma_y a_k & \dots & a_k'\rho_y^{n-1}\Sigma_y a_k & a_k'\Sigma_{xy} b_k & \dots & a_k'\rho_{xy}^{n-1}\Sigma_{xy} b_k \\ & \ddots & & & \ddots & \\ a_k'\rho_y^{n-1}\Sigma_y a_k & \dots & a_k'\Sigma_y a_k & a_k'\rho_{xy}^{n-1}\Sigma_{xy} b_k & \dots & a_k'\Sigma_{xy} b_k \\ b_k'\Sigma_{yx} a_k & \dots & b_k'\rho_{xy}^{n-1}\Sigma_{xy} a_k & b_k'\Sigma_x b_k & \dots & b_k'\rho_x^{n-1}\Sigma_x b_k \\ & \ddots & & & \ddots & \\ b_k'\rho_{xy}^{n-1}\Sigma_{xy} a_k & \dots & b_k'\Sigma_{yx} a_k & b_k'\rho_x^{n-1}\Sigma_x b_k & \dots & b_k'\Sigma_x b_k \end{pmatrix}.$$

Since $a_k'\Sigma_y a_k = 1$ and $b_k'\Sigma_x b_k = 1$ the matrix reduces to

$$\Theta = \begin{pmatrix} \Psi_y & \Psi_{yx}\lambda_k \\ \Psi_{xy}\lambda_k & \Psi_x \end{pmatrix}, \tag{21}$$

where

$$\lambda_k = a_k'\Sigma_{xy} b_k.$$

Equation (21) shows how canonical correlation analysis can be used to reduce the covariance matrix to a function of correlations between time units. It should be noted that while the researcher may choose to ignore the covariance structure across the time points, the true variance covariance matrix still shows the time dependency after the transformation.

IV.4 CREATING THE DUPLICATION MATRIX

The distribution of the canonical correlations can be derived from asymptotic distribution of the variance covariance matrix $\hat{\Theta}$, which is a 2t $\times$ 2t matrix of sample variance covariance matrix of the transformed data. This requires the use of the duplication matrix and accompanying matrix operations (Schott, 1997, p. 285; Tim, 2002, p. 90). The duplication matrix can be used to determine the asymptotic distribution of the vector vech($\hat{\Theta}$) using the asymptotic distribution of the $\hat{\Theta}$ matrix. The canonical correlations are subsets of the matrix $\hat{\Theta}$.

$$vech(\hat{\Theta}) \sim N\left(vech(\Theta), \left(\frac{2D_v^{+}\Theta \otimes \Theta D_v^{+'}}{n-1}\right)\right) \tag{22}$$

Equation 22 gives the asymptotic distribution of the vech($\hat{\Theta}$) vector. Vech($\hat{\Theta}$) is chosen over vec($\hat{\Theta}$) to ensure the corresponding distribution is nonsingular. The D_v^{+} matrix is the Moore-Penrose inverse of the Duplication matrix, determined as

$$D_v^{+} = (D_v' D_v)^{-1} D_v'.$$

The corresponding asymptotic variance is a function of the Kronecker product covariance matrix and the Moore-Penrose inverse of the duplication matrix. It is clear that the duplication matrix is a useful matrix and it would be helpful to have an algorithm to create this matrix easily. In the following the structure of the *vech* and *vec* operators will be used to create a simple algorithm to calculate the duplication matrix.

The matrix transformations shown next in equation (23) are the *vech* and *vec* operators. Equation (23) illustrates the *vech* and *vec* operators with an $n \times n$ matrix A. The *vech* operator takes an $n \times n$ symmetric matrix A and transforms it into a $\frac{(n+1)n}{2}$ vector that does not include the duplicate entries from the original matrix. The procedure used transforms the matrix A to take the first column of the matrix and create the first n rows of the vector. Take the second column of the original matrix, remove the first value, and attach the remaining $n-1$ values to the new vector. Continue in this way until the last column of the matrix is reached. When the last column of the A matrix is appended, only one value from matrix A, $A_{n,n}$ is appended to the column of vector $vech(A)$.

The *vec* operator uses an $n \times n$ matrix A to create an $n^2 \times 1$ vector by stacking the columns of A. The first value in the vector is the row 1 column 1 from the original

matrix $a_{1,1}$. The second value is the matrix entry from row 2 column 1 of the original matrix $a_{2,1}$. This continues for all elements of the matrix.

$$\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}; \mathrm{vech(A)} = \begin{bmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{n,1} \\ a_{2,2} \\ a_{3,2} \\ \vdots \\ a_{n,2} \\ \vdots \\ a_{n-1,n-1} \\ a_{n,n-1} \\ a_{n,n} \end{bmatrix}; \mathrm{vec(A)} = \begin{bmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{n,1} \\ a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{n,2} \\ \vdots \\ a_{1,n-1} \\ \vdots \\ a_{n,n-1} \\ a_{1,n} \\ \vdots \\ a_{n,n} \end{bmatrix} \tag{23}$$

The duplication matrix D_v converts an $\frac{n(n+1)}{2} \times 1$ vector into an $n^2 \times 1$ vector. The matrix D_v allows the transformation from $vech(A)$ into $vec(A)$. The transformation matrix between the $vec(A)$ and $vech(A)$ is shown below.

$$D_v vech(A) = vec(A),$$

where

$$D_v = \begin{bmatrix}
 & & & \mathbf{I}_n & & & \mathbf{0} & & \dots & & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
0 & 1 & 0 & 0 & \dots & 0 & 0 & & \dots & & 0 & \dots & 0 & 0 & 0 \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & & & \mathbf{I}_{n-1} & & & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
0 & 0 & 1 & 0 & \dots & 0 & 0 & & \dots & & 0 & \dots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 & 0 & \dots & 0 & \dots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & & \dots & & \mathbf{0} & \dots & \mathbf{I}_2 & & \mathbf{0} \\
0 & 0 & 0 & 0 & \dots & 1 & 0 & & \dots & & 0 & \dots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 1 & \dots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & \dots & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & 1
\end{bmatrix} \tag{24}$$

The D_v matrix must be created automatically to accommodate different number of time periods. An algorithm to create the D_v matrix is included below. The D_v matrix dimensions are $n^2 \times \frac{n(n+1)}{2}$. The first step is to create a matrix of zeros corresponding to those dimensions.

An algorithm was created to automate the creation of the duplication matrix. The algorithm takes advantage of the duplication matrix structure. Each row of the duplication matrix is used to recreate a row of the original matrix from (vech) vector. What follows is a simple algorithm to create the duplication matrix. The algorithm is simpler than another approach in (Harville, 1997, p. 352). The steps below show the logical sequence used to develop the algorithm.

The first rows of the duplication matrix form an identity matrix I_n. These simply replicate rows 1 through n of the original matrix.

$$\begin{bmatrix} & \mathbf{I}_n & & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{n,1} \\ a_{2,2} \\ a_{3,2} \\ \vdots \\ a_{n,2} \\ \vdots \\ a_{n-1,n-1} \\ a_{n,n-1} \\ a_{n,n} \end{bmatrix} = \begin{bmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{n,1} \\ 0 \\ \vdots \end{bmatrix}$$

Since $a_{1,2} = a_{2,1}$, the *vech* matrix does not contain a value for the matrix element $a_{1,2}$. Hence the duplication matrix uses a previous row from $vech(A)$ to add the correct value into a row to add back in the duplication matrix.

$$\begin{bmatrix} & \mathbf{I}_n & & & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 & \dots & 0 & \dots & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{n,1} \\ a_{2,2} \\ a_{3,2} \\ \vdots \\ a_{n,2} \\ \vdots \\ a_{n-1,n-1} \\ a_{n,n-1} \\ a_{n,n} \end{bmatrix} = \begin{bmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{n,1} \\ a_{1,2} \\ 0 \\ \vdots \end{bmatrix}$$

Matrix values $A_{2,2}$ to $A_{n,2}$ can be pulled directly from the $vech(A)$ vector rows n+1 to 2n-1. This creates another identity matrix with the dimensions I_{n-1}.

$$\left[\begin{array}{ccccccccccccc} & & & \mathbf{I}_n & & & \mathbf{0} & \ldots & & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 0 & 1 & 0 & 0 & \ldots & 0 & 0 & \ldots & & 0 & \ldots & 0 & 0 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & & \mathbf{I}_{n-1} & & & \ldots & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array}\right] \left[\begin{array}{c} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{n,1} \\ a_{2,2} \\ a_{3,2} \\ \vdots \\ a_{n,2} \\ \vdots \\ a_{n-1,n-1} \\ a_{n,n-1} \\ a_{n,n} \end{array}\right] = \left[\begin{array}{c} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{n,1} \\ a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{n,2} \\ 0 \\ \vdots \end{array}\right]$$

Rows $vec(A)_{2n+1}$ and $vec(A)_{2n+2}$ are elements $A_{1,3}$ and $A_{2,3}$ of the original matrix. These values are not directly contained in the *vech* vector because the A matrix is symmetric. To obtain these rows, the duplication matrix takes them from previous rows of $vech(A)$ vector.

$$\left[\begin{array}{ccccccccccccc} & & & \mathbf{I}_n & & & \mathbf{0} & \ldots & & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 0 & 1 & 0 & 0 & \ldots & 0 & 0 & \ldots & & 0 & \ldots & 0 & 0 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & & \mathbf{I}_{n-1} & & & \ldots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 0 & 0 & 1 & 0 & \ldots & 0 & 0 & \ldots & & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 1 \quad 0 & \ldots & 0 & \ldots & 0 & 0 & 0 \end{array}\right] \left[\begin{array}{c} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{n,1} \\ a_{2,2} \\ a_{3,2} \\ \vdots \\ a_{n,2} \\ \vdots \\ a_{n-1,n-1} \\ a_{n,n-1} \\ a_{n,n} \end{array}\right] = \left[\begin{array}{c} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{n,1} \\ a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{n,2} \\ a_{1,3} \\ a_{2,3} \\ 0 \\ \vdots \end{array}\right]$$

This pattern is repeated to gain the relationship $Dvech(A) = vec(A)$. The matrix in Equation (25) is separated with vertical and horizontal lines to illustrate

the different subsections. These subsections were used to automate the duplication matrix.

$$\left[\begin{array}{cccccc|ccccc|cc|c}
 & & & \mathbf{I}_n & & & \mathbf{0} & & \dots & & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\hline
0 & 1 & 0 & 0 & \dots & 0 & 0 & & \dots & & 0 & \dots & 0 & 0 & 0 \\
\hline
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & & & \mathbf{I}_{n-1} & & & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\hline
0 & 0 & 1 & 0 & \dots & 0 & 0 & & \dots & & 0 & \dots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 & 0 & \dots & 0 & \dots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\hline
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & & \dots & & \mathbf{0} & \dots & \mathbf{I}_2 & & \mathbf{0} \\
\hline
0 & 0 & 0 & 0 & \dots & 1 & 0 & & \dots & & 0 & \dots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \dots & 0 & 0 & & \dots & 0 & 1 & \dots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \dots & 0 & 0 & & \dots & & 0 & \dots & 0 & 1 & 0 \\
\hline
0 & 0 & 0 & 0 & \dots & 0 & 0 & & \dots & & 0 & \dots & 0 & 0 & \mathbf{I}_1
\end{array}\right]
\left[\begin{array}{c} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{n,1} \\ a_{2,2} \\ a_{3,2} \\ \vdots \\ a_{n,2} \\ \vdots \\ a_{n-1,n-1} \\ a_{n,n-1} \\ a_{n,n} \end{array}\right]
=
\left[\begin{array}{c} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{n,1} \\ a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{n,2} \\ \vdots \\ a_{1,n-1} \\ \vdots \\ a_{n,n-1} \\ a_{1,n} \\ \vdots \\ a_{n,n} \end{array}\right] \tag{25}$$

The columns of the duplication matrix can be labeled to correspond to the cells of the original matrix and the rows of the *vech* vector. The rows of the duplication matrix can be labeled to represent the cells of the original matrix and the rows of the $vec(A)$ vector.

row/column	11	21	...	*n*1	22	32	...	*n*2	...	*nn*
11	1	0	...	0	0	0	...	0	...	0
21	0	1	...	0	0	0	...	0	...	0
⋮										
*n*1	0	0	...	1	0	0	...	0	...	0
12	0	1	...	0	0	0	...	0	...	0
⋮										
nn	0	0	...	0	0	0	...	0	...	1

(26)

An algorithm was created to populate the duplication matrix. First the columns

of the matrix A are indexed from 1 to $\frac{n(n+1)}{2}$ and the rows are labeled from 1 to n^2.

Index		1	2				...			$\frac{n(n+1)}{2}$
	row/column	$1,1$	$2,1$	...	$n,1$	$2,2$	$3,2$	...	$n,2$	...	n,n
1	$1,1$	1	0	...	0	0	0	...	0	...	0
2	$2,1$	0	1	...	0	0	0	...	0	...	0
$\vdots$	$\vdots$		...		...		...		...		...
n	$n,1$	0	0	...	1	0	0	...	0	...	0
$n+1$	$1,2$	0	1	...	0	0	0	...	0	...	0
$\vdots$	$\vdots$		...		...		...		...		...
n^2	n,n	0	0	...	0	0	0	...	0	...	1

The algorithm cycles through all rows and columns of the original matrix. For each row and column, the duplication matrix will have a 1 in the position defined in equation (27) or (28) below.

The duplication matrix represents each element in the original matrix. Hence to create the algorithm for D_v, it must cycle through every element in the A matrix. Let

When cycling through the row and column indexes, use one of the two formulas below (27) or (28). Begin the algorithm with cell row index (RI)=1 and column index (CI)=1. Then move to RI =1 and CI=2. Cycle through the entire matrix in this way. While cycling through the matrix if the row index is greater than or equal to the column index use equation (27). The value n is the total number of rows or the total number of columns in the original matrix. Formula (27) or (28) gives the duplication row index (DRI) and the duplication column index (DCI) for each cell in the matrix that will become 1. The remaining cells are 0.

$$\begin{gathered}\text{Duplication Matrix Row Index and Column Index} = \\ (DRI, DCI) = \\ \Big((n(CI-1)+RI, n(CI-1) - \tfrac{(CI-2)(CI-1)}{2} + (1+RI-CI)\Big).\end{gathered} \tag{27}$$

If the column index is greater than the row index use equation (28).

$$\begin{gathered}\text{Duplication Row and Column index} = \\ (DRI, DCI) = \\ \Big((n(CI-1)+\text{row}, n(RI-1) - \tfrac{(RI-2)(RI-1)}{2} + (1+CI-RI)\Big).\end{gathered} \tag{28}$$

After cycling the matrix row and column indexes and placing a one in the corresponding location of the D_v matrix the result will be the duplication matrix such that $D_v vech(A) = vec(A)$.

To gain the large sample distribution for the canonical correlations, specific values are required from the covariance matrix. The canonical correlations are estimated in the covariance matrix Θ. The equation (33) shows a large sample approximation for the distribution of Θ. Only specific values of from the distribution are required to estimate the canonical correlation distribution.

$$vech(\hat{\Theta}) \sim N\left(vech(\Theta), \frac{2D^{+}\Theta \otimes \Theta D^{+'}}{n-1}\right) \tag{29}$$

$$vech(\Theta) =$$

$$vech\begin{pmatrix} a_k'\Sigma_y a_k' & \cdots & a_k'\rho_y^{n-1}\Sigma_y a_k & a_k'\Sigma_{xy} b_k & \cdots & a_k'\rho_{xy}^{n-1}\Sigma_{xy} b_k \\ & \ddots & & & \ddots & \\ a_k'\rho_y^{n-1}\Sigma_y a_k & \cdots & a_k'\Sigma_y a_k & a_k'\rho_{xy}^{n-1}\Sigma_{xy} b_k & \cdots & a_k'\Sigma_{xy} b_k \\ b_k'\Sigma_{yx} a_k & \cdots & b_k'\rho_{xy}^{n-1}\Sigma_{xy} a_k & b_k'\Sigma_x b_k & \cdots & b_k'\rho_x^{n-1}\Sigma_x b_k \\ & \ddots & & & \ddots & \\ b_k'\rho_{xy}^{n-1}\Sigma_{xy} a_k & \cdots & b_k'\Sigma_{yx} a_k & b_k'\rho_x^{n-1}\Sigma_x b_k & \cdots & b_k'\Sigma_x b_k \end{pmatrix} =$$

$$\begin{pmatrix} a_k'\Sigma_y a_k \\ \vdots \\ a_k'\rho_y^{n-1}\Sigma_y a_k \\ a_k'\Sigma_{xy} b_k \\ \vdots \\ a_k'\rho_{xy}^{n-1}\Sigma_{xy} b_k \\ a_k'\Sigma_y a_k \\ \vdots \\ a_k'\rho_y^{n-2}\Sigma_y a_k \\ a_k'\rho_{xy}\Sigma_{xy} b_k \\ \vdots \\ a_k'\rho_{xy}^{n-2}\Sigma_{xy} b_k \\ \vdots \\ b_k'\Sigma_x b_k \end{pmatrix}.$$

IV.4.1 Isolating Variables of Interest

Once the Duplication matrix is created and the asymptotic distribution is calculated, the values pertaining to the canonical correlations can be isolated. Only the values corresponding to $a_k'\Sigma_{xy}b_k$ are required to determine the asymptotic distribution of the canonical correlations. The position of these values within $vech(\Theta)$ are labeled in equation (30) below.

$$vech(\Theta) = \begin{pmatrix} v_1 \\ \vdots \\ v_{t+1} \\ \vdots \\ v_{2t+t+1} \\ \vdots \\ v_{2t+2t-1+t+1} \\ \vdots \\ v_{2t(i-1)-\frac{(i-1)(i-2)}{2}+t+1} \\ \vdots \\ v_{2t(t-1)-\frac{(t-1)(t-2)}{2}+t+1} \\ \vdots v_{(1+2t)t} \end{pmatrix} = \begin{pmatrix} a_k'\Sigma_{xy}a_k \\ \vdots \\ a_k'\Sigma_{xy}b_k \\ \vdots \\ a_k'\Sigma_{xy}b_k \\ \vdots \\ a_k'\Sigma_{xy}b_k \\ \vdots \\ a_k'\Sigma_{xy}b_k \\ \vdots \\ a_k'\Sigma_{xy}b_k \\ \vdots \\ b_k'\Sigma_{xy}b_k \end{pmatrix}. \tag{30}$$

The pattern of these values can be recognized and replicated. The first canonical correlation retained corresponds to the $t+1$ location in the $vech(\Theta)$ vector. The next first canonical correlation that follows in the $vech$ vector corresponds to the position 2t+t+1. The pattern continues with the ith canonical correlation in the $2t(t-1) - \frac{(t-1)(t-2)}{2} + t + 1$ position. Note all kth canonical correlations will be the result of the a_k and b_k eigenvalues. Hence if a_k and b_k are taken from the first eigenvectors all canonical correlations will be the kth canonical correlation.

Using matrix C_v defined as in equation (31) we can isolate values of the covariance for the canonical correlations. The C_v matrix has dimensions $t \times \frac{2t(2t+1)}{2}$. The resulting matrix is zero everywhere except at the t points where it is 1.

$$C_v = \begin{pmatrix} 0 & \dots & C_{v_{1,t+1}} = 1 & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & C_{v_{2,3t+1}} = 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & C_{v_{2t(t-1)-\frac{(t-1)(t-2)}{2}+t+1}=1} & \dots & 0 \end{pmatrix} \quad (31)$$

The asymptotic distribution of the canonical correlations are shown in equation (33). The C_v matrix helps isolate and subset the canonical correlation variances.

$$C_v vech(\Theta) \sim N\left(C_v vech(\Theta), \frac{2C_v D^{+}\Theta \otimes \Theta D^{+'} C_v'}{n-1}\right) \quad (32)$$

IV.5 HYPOTHESIS TEST GOING FROM I TO II

If the true variance covariance matrix is one of the covariance structures II through V from Chapter III, the canonical correlations within each time period will have the same value. As shown in Chapter III, attempts at using asymptotic maximum likelihood approximations as a basis for a hypothesis test to reduce the covariance from I to II showed a high bias in the test. This bias may have resulted from using the log likelihood ratio test approximation on an estimator that may not have been the maximum likelihood estimator. The distribution shown in equation (33) provides a test going from covariance matrix I to II that should bring the Type I error down closer to 0.05 when we specify the correct covariance.

Since under structure II through V all sigma matrices are equal and hence their canonical correlations are also equal, a reasonable test to determine reduction of the covariance I to covariance II would be to test if the 1st canonical correlations are equal.

Let the null hypothesis be that all the first canonical correlations are equal to each other, that is,

$$H_o : \lambda_1 = \lambda_2 = \ldots = \lambda_t.$$

Note the λ values are the first canonical correlation coefficient for each set of data at a specific time point. Hence λ_1 is the first canonical correlation for the first time period, λ_2 is the first canonical correlation for the second time period, and so on.

The variance covariance matrix was estimated at each of the time points, as

$$S_i = \begin{pmatrix} S_{y_i} & S_{yx_i} \\ S_{xy_i} & S_{x_i} \end{pmatrix}, \text{ where } i = 1, 2, \ldots, t.$$

The S_i matrices were averaged together to produce one matrix $\bar{\Sigma}$. Canonical correlation analysis was then performed using the $\bar{\Sigma}$ matrix and the first set of canonical coefficients $\bar{a}$ were produced for the Y values along with the first set of canonical coefficients $\bar{b}$ for the X values.

The X and Y variables were then transformed by $\bar{b}$ and $\bar{a}$ vectors, respectively.

$$\Gamma \begin{pmatrix} Y \\ X \end{pmatrix} = \begin{pmatrix} I_t \otimes \bar{a}' & 0 \\ 0 & I_t \otimes \bar{b}' \end{pmatrix} \begin{pmatrix} Y \\ X \end{pmatrix}.$$

If the true underlying variance covariance matrix is of the form

$$\begin{bmatrix} \Psi \otimes \Sigma_y & \Psi \otimes \Sigma_{yx} \\ \Psi \otimes \Sigma_{xy} & \Psi \otimes \Sigma_x \end{bmatrix}$$

the canonical correlations should all be equal. Under the matrix II the variance covariance matrix of $\Gamma \begin{pmatrix} Y \\ X \end{pmatrix}$ is

$$\bar{\Theta} = \Gamma \begin{bmatrix} \Psi \otimes \Sigma_y & \Psi \otimes \Sigma_{yx} \\ \Psi \otimes \Sigma_{xy} & \Psi \otimes \Sigma_x \end{bmatrix} \Gamma' =$$

$$\left(\begin{array}{ccc|ccc} \bar{a}'\hat{\Sigma}_y\bar{a} & \ldots & \bar{a}'\rho_y^{n-1}\hat{\Sigma}_y\bar{a} & \bar{a}'\hat{\Sigma_{xy}}\bar{b} & \ldots & \bar{a}'\rho_{xy}^{n-1}\Sigma_{xy}\bar{b} \\ & \ddots & & & \ddots & \\ \bar{a}'\rho_y^{n-1}\Sigma_y\bar{a} & \ldots & \bar{a}'\Sigma_y\bar{a} & \bar{a}'\rho_{xy}^{n-1}\Sigma_{xy}\bar{b} & \ldots & \bar{a}'\Sigma_{xy}\bar{b} \\ \hline \bar{b}'\Sigma_{yx}\bar{a} & \ldots & \bar{b}'\rho_{xy}^{n-1}\Sigma_{xy}\bar{a} & \bar{b}'\Sigma_x\bar{b} & \ldots & \bar{b}'\rho_x^{n-1}\Sigma_x\bar{b} \\ & \ddots & & & \ddots & \\ \bar{b}'\rho_{xy}^{n-1}\Sigma_{xy}\bar{a} & \ldots & \bar{b}'\Sigma_{yx}\bar{a} & \bar{b}'\rho_x^{n-1}\Sigma_x\bar{b} & \ldots & \bar{b}'\Sigma_x\bar{b} \end{array}\right).$$

This will be asymptotically distributed as

$$C_v vech(\bar{\Theta}) \sim N\left(C_v vech(\Theta), \frac{2C_v D_v^+ \Theta \otimes \Theta D_v^{+'} C_v'}{n-1}\right).$$

To test all the $\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_t \end{pmatrix}$ values simultaneously,

$$\lambda_1 = \lambda_2$$
$$\lambda_2 = \lambda_3$$
$$\vdots$$
$$\lambda_{t-1} = \lambda_t$$

the Bonferroni correction was used to account for inflation in the Type I error rate from testing multiple hypothesis. The matrix Δ was used to construct a test for the λ values (Ravishanker and Dey, 2002, p. 217).

$$\Delta = \begin{bmatrix} 1 & -1 & 0 & 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & 1 & -1 & 0 & \dots & \dots & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 & -1 \end{bmatrix}$$

$$\Delta\hat{\lambda} = \Delta C_v vech(\tilde{\Theta}) \sim N\left(\Delta C_v vech(\Theta), \frac{2\Delta C_v D_v^+ (\Theta \otimes \Theta) D_v^{+'} C_v' \Delta'}{n-1}\right).$$

and the decision rule is to reject H_o if the absolute value of the test statistic is greater than $Z_{1-\frac{\alpha}{2(t-1)}}$

IV.5.1 Canonical Correlation Test Result

Simultaneous tests were conducted to determine if $\Delta\lambda = 0$ for all λ. A simulation was run to test the asymptotic distribution of the λ values. The distribution had samples sizes of 50 to 500. The number of X variables were fixed at 2. The number of Y variables were fixed at 4. The ρ_x, ρ_y and ρ_{xy} fluctuated from 0.2 and 0.8. The time value fluctuated between 3, 6, and 9 units. The overall rejection rate was set at $\alpha = .05$ and the number of rejections were recorded for each run of 100 simulations. The Tables (44)-(48) illustrates the rejection rates for sample sizes 50 to 500. The asymptotic distribution seems to hold with all tests showing a rejection rates of around the expected 5 out of 100 or 0.05.

IV.5.2 Example Variance Covariance Matrix for $t = 3$

To gain insight into the problem, the covariance matrix was calculated for the case of $t = 1, t = 2$ and $t = 3$. The AR(1) correlation matrix was used in the calculation. The $\Theta \otimes \Theta$ matrix is $(2t)^2 \times (2t)^2$. For $t = 1$ this is 4×4 for 16 elements. For $t = 2$

the matrix has $16 \times 16 = 256$ elements. For $t = 3$ the matrix results in $36 \times 36 = 1296$ elements. An algorithm was developed to multiply the matrix $C_v D_v^+$ by $\Theta \otimes \Theta$ and $C_v D_v^+ \Theta \otimes \Theta$ by $C_v D_v^{\prime +}$. The resulting method allowed the individual values of the $C_v D_v^+ (\Theta \otimes \Theta) D_v^{+\prime} C_v^\prime$ matrix to be derived.

The objective was to show how the ρ_x, ρ_y, and ρ_{xy} values propagate through the equations and influence the resulting variance covariance matrix

$$\left(2C_v D^+ (\Theta \otimes \Theta) D_v^{+\prime} C_v^\prime\right).$$

For $t = 1$ the matrix $C_v D_v^+$ was shown to be

$$C_v D_v^+ =$$

$$\begin{pmatrix} 0 & 0.5 & 0.5 & 0 \end{pmatrix}.$$

The full variance covariance matrix

$$\left(2C_v D_v^+ (\Theta \otimes \Theta) D_v^{+\prime} C_v^\prime\right)$$

was equal to

$$\left(1 + \lambda^2\right).$$

See Anderson,(1984, p. 120-121) for a comparison.
For $t = 2$ the matrix $C_v D_v^+$ was shown to be

$$C_v D_v^+ =$$

$$\begin{pmatrix} 0 & C_v D_{v_{1,3}}^+ = .5 & 0 & \dots & C_v D_{v_{1,9}}^+ = .5 & 0 & \dots \\ \dots & 0 & C_v D_{v2,8}{}^+ = .5 & 0 & \dots & 0 & C_v D_{v_{2,14}}^+ = .5 & 0 \end{pmatrix}.$$

The full variance covariance matrix

$$\left(2C_v D_v^+ (\Theta \otimes \Theta) D_v^{+\prime} C_v^\prime\right) =$$

was equal to

$$\begin{pmatrix} 1 + \lambda^2 & \lambda^2 \rho_{xy}^2 + \rho_x \rho_y \\ \lambda^2 \rho_{xy}^2 + \rho_x \rho_y & 1 + \lambda^2 \end{pmatrix}.$$

For $t = 3$ the matrix $C_v D_v^+$ was shown to be

$$C_v D_v^+ =$$

$$\begin{pmatrix} 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ C_v D_{v_{1,4}}^+ = .5 & 0 & 0 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & C_v D_{v_{2,11}}^+ = .5 & 0 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & C_v D_{v_{3,18}}^+ = .5 \\ C_v D_{v_{1,19}}^+ = .5 & 0 & 0 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & C_v D_{v_{2,26}}^+ = .5 & 0 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & C_v D_{v_{3,33}}^+ = .5 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{pmatrix}'.$$

The full variance covariance matrix

$$\left(2C_v D_v^+ (\Theta \otimes \Theta) D_v^{+'} C_v'\right) =$$

was equal to

$$\begin{pmatrix} 1+\lambda^2 & \lambda^2\rho_{xy}^2 + \rho_x\rho_y & \lambda^2\rho_{xy}^4 + \rho_x^2\rho_y^2 \\ \lambda^2\rho_{xy}^2 + \rho_x\rho_y & 1+\lambda^2 & \lambda^2\rho_{xy}^2 + \rho_x\rho_y \\ \lambda^2\rho_{xy}^4 + \rho_x^2\rho_y^2 & \lambda^2\rho_{xy}^2 + \rho_x\rho_y & 1+\lambda^2 \end{pmatrix}.$$

Matrix structure of the covariance shows that λ plays a role in all variance covariance approximations. The variance for each λ is $1+\lambda^2$. For $t = 2$ the variance of each

λ remains $1+\lambda^2$ but the covariance between λ's is $\lambda^2\rho_{xy}^2+\rho_x\rho_y$. This shows that the covariance is a function of all three, ρ_x, ρ_y, and ρ_{xy}. For $t=3$ the variance of λ remains the same and the covariance for λ between sequential points remains constant. The new term in the model is $\lambda^2\rho_{xy}^4+\rho_x^2\rho_y^2$ for the covariance between λ_1 and λ_3. The sequence repeats and hence covariance between λ_i and $\lambda_j = \lambda^2\rho_{xy}^{2|i-j|}+\rho_x^{|i-j|}\rho_y^{|i-j|}$. If $\rho_x=\rho_y=\rho_{xy}=\rho$ then the covariance between the terms λ_i and λ_j decays at a rate $(1+\lambda^2)\rho^{2|i-j|}$.However, ρ_x, ρ_y, and ρ_{xy} do not appear in the λ estimate. This makes sense since the λ is the canonical correlation of the Σ matrix.

IV.5.3 Testing Specific Structures

Using the distribution defined above, it is possible to test more specific forms of the variance covariance matrix. The objective of the following test is to rule out the covariance structures shown in Chapter II and Chapter III. After the transformation using the canonical coefficients, the distribution is

$$vech(\bar{\Theta}) \sim N\left(vech(\Theta), \frac{2D_v^+(\Theta\otimes\Theta)D_v^{+'}}{n-1}\right).$$

The delta theorem was used to calculate the asymptotic distribution for the ratio of multiple variables. The idea behind this test is to create ratios from $\bar{\Theta}$ that will equal ρ_y, ρ_{xy}, and ρ_x under the null hypothesis. If the ratios are rejected, then variance covariance structure II through V should be rejected. If the ratios are not rejected, the researcher should then estimate and test variance covariance matrices II through V using the methods discussed in Chapter II and Chapter III.

Under the null hypotheses II through V, it is possible to create ratios of random variables that can be tested to disprove the null hypothesis. For the case of $t=3$ the matrix is

$$vech(\bar{\Theta}) = vech\begin{pmatrix} 1 & \rho_y & \rho_y^2 & \lambda & \lambda\rho_{xy} & \lambda\rho_{xy}^2 \\ \rho_y & 1 & \rho_y & \lambda\rho_{xy} & \lambda & \lambda\rho_{xy} \\ \rho_y^2 & \rho_y & 1 & \lambda\rho_{xy}^2 & \lambda\rho_{xy} & \lambda \\ \lambda & \lambda\rho_{xy} & \lambda\rho_{xy}^2 & 1 & \rho_x & \rho_x^2 \\ \lambda\rho_{xy} & \lambda & \lambda\rho_{xy} & \rho_x & 1 & \rho_x \\ \lambda\rho_{xy}^2 & \lambda\rho_{xy} & \lambda & \rho_x^2 & \rho_x & 1 \end{pmatrix}.$$

Let the $\hat{\bar{\Theta}}$ matrix represent the estimates of the $\bar{\Theta}$ matrix

$$vech(\hat{\bar{\Theta}}) = \begin{pmatrix} 1 \\ \widehat{\rho_y} \\ \widehat{\rho_y^2} \\ \widehat{\lambda} \\ \widehat{\lambda\rho_{xy}} \\ \widehat{\lambda\rho_{xy}^2} \\ 1 \\ \widehat{\rho_y} \\ \widehat{\lambda\rho_{xy}} \\ \widehat{\lambda} \\ \widehat{\lambda\rho_{xy}} \\ 1 \\ \widehat{\lambda\rho_{xy}^2} \\ \widehat{\lambda\rho_{xy}} \\ \widehat{\lambda} \\ 1 \\ \widehat{\rho_x} \\ \widehat{\rho_x^2} \\ 1 \\ \widehat{\rho_x} \\ 1 \end{pmatrix} =$$

If variance covariance matrices II through V are true, specific ratios in the $vech(\hat{\bar{\Theta}})$ vector can be created to test the basic matrix structure. Hypothesis tests can be formulated from ratios of elements of the $vech(\hat{\bar{\Theta}})$ vector and the asymptotic distribution can be used to approximate the variance.

First, all similar terms must be averaged together. Use a transformation of the $vech(\hat{\bar{\Theta}})$ vector to average like terms. For the case of $t = 3$ multiply $vech(\hat{\bar{\Theta}})$ by the matrix

$$\Upsilon vech(\hat{\bar{\Theta}}) =$$
$$(\Gamma_1\Gamma_2)\, vech(\hat{\bar{\Theta}}) =$$
$$\begin{pmatrix} \bar{\widehat{\rho_y}} & \bar{\widehat{\rho_y^2}} & \bar{\widehat{\lambda}} & \bar{\widehat{\lambda\rho_{xy}}} & \bar{\widehat{\lambda\rho_{xy}^2}} & \bar{\widehat{\rho_x}} & \bar{\widehat{\rho_x^2}} \end{pmatrix}'.$$

Where

$$\Gamma_1 = \begin{pmatrix} 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0.25 & 0 & 0 & 0 & 0.25 & 0 & 0.25 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and

$$\Gamma_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

The resulting vector is distributed as

$$\Upsilon vech(\hat{\bar{\Theta}}) \sim N\left(\Upsilon vech(\Theta), \frac{2\Upsilon D_v^+(\Theta \otimes \Theta)D_v^{+'}\Upsilon'}{n-1}\right). \tag{33}$$

If hypothesis II is true, the equations in (34) should hold.

$$\frac{\widehat{\bar{\rho}_{xy}^2}}{\widehat{\bar{\rho}_y}} = \widehat{\bar{\rho}_y} \qquad \frac{\widehat{\bar{\lambda}\rho_{xy}^2}}{\widehat{\bar{\lambda}\rho_{xy}}} = \frac{\widehat{\bar{\lambda}\rho_{xy}}}{\widehat{\bar{\lambda}}} \qquad \frac{\widehat{\bar{\lambda}\rho_x^2}}{\widehat{\bar{\rho}_x}} = \widehat{\bar{\rho}_x} \tag{34}$$

If hypothesis III is true, the equations in (35) or (36) should hold.

$$\frac{\widehat{\bar{\rho}_{xy}^2}}{\widehat{\bar{\rho}_y}} = \widehat{\bar{\rho}_y} = 1 \qquad \frac{\widehat{\bar{\lambda}\rho_{xy}^2}}{\widehat{\bar{\lambda}\rho_{xy}}} = \frac{\widehat{\bar{\lambda}\rho_{xy}}}{\widehat{\bar{\lambda}}} \qquad \frac{\widehat{\bar{\lambda}\rho_x^2}}{\widehat{\bar{\rho}_x}} = \widehat{\bar{\rho}_x} \tag{35}$$

or

$$\frac{\widehat{\bar{\rho}_{xy}^2}}{\widehat{\bar{\rho}_y}} = \widehat{\bar{\rho}_y} \qquad \frac{\widehat{\bar{\lambda}\rho_{xy}^2}}{\widehat{\bar{\lambda}\rho_{xy}}} = \frac{\widehat{\bar{\lambda}\rho_{xy}}}{\widehat{\bar{\lambda}}} \qquad \frac{\widehat{\bar{\lambda}\rho_x^2}}{\widehat{\bar{\rho}_x}} = \widehat{\bar{\rho}_x} = 1 \tag{36}$$

If hypothesis IV is true, equation (37) should hold.

$$\frac{\widehat{\bar{\rho}_{xy}^2}}{\widehat{\bar{\rho}_y}} = \widehat{\bar{\rho}_y} = \frac{\widehat{\bar{\lambda}\rho_{xy}^2}}{\widehat{\bar{\lambda}\rho_{xy}}} = \frac{\widehat{\bar{\lambda}\rho_{xy}}}{\widehat{\bar{\lambda}}} = \frac{\widehat{\bar{\lambda}\rho_x^2}}{\widehat{\bar{\rho}_x}} = \widehat{\bar{\rho}_x} \tag{37}$$

If hypothesis V is true, equation (38) should hold.

$$\bar{\widehat{\rho_y^2}} = \bar{\widehat{\rho_y}} = \bar{\widehat{\lambda\rho_{xy}^2}} = \bar{\widehat{\lambda\rho_{xy}}} = \bar{\widehat{\lambda\rho_x^2}} = \bar{\widehat{\rho_x}} = 0 \tag{38}$$

IV.5.4 Multivariate Delta Theorem Application

The Multivariate Delta Theorem can be used to estimate the asymptotic distribution of the ratios shown in equations (34) through (38). It is important to remember in the following derivations that $\bar{\widehat{\rho_y^2}}$ and $\bar{\widehat{\rho_y}}$, $\bar{\widehat{\lambda\rho_x^2}}$ and $\bar{\widehat{\lambda\rho_x^2}}$, and $\bar{\widehat{\lambda\rho_{xy}^2}}$, $\bar{\widehat{\lambda\rho_{xy}}}$, and $\bar{\widehat{\lambda}}$ are calculated as separate random variables and are not functions of one another.

Let

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}, \begin{pmatrix} \sigma_{1,1} & \cdots & \sigma_{1,n} \\ \vdots & \ddots & \vdots \\ \sigma_{n,1} & \cdots & \sigma_{n,n} \end{pmatrix}\right)$$

The multivariate delta theorem states the asymptotic distribution of

$$\begin{pmatrix} f_1(x_1,\ldots,x_n) \\ \vdots \\ f_m(x_1,\ldots,x_n) \end{pmatrix} \sim N\left(\begin{pmatrix} f_1(\mu_1,\ldots,\mu_n) \\ \vdots \\ f_m(\mu_1,\ldots,\mu_n) \end{pmatrix}, \zeta \begin{pmatrix} \sigma_{1,1} & \cdots & \sigma_{1,n} \\ \vdots & \ddots & \vdots \\ \sigma_{n,1} & \cdots & \sigma_{n,n} \end{pmatrix} \zeta'\right),$$

where ζ is defined as

$$\zeta = \begin{pmatrix} \frac{\delta f_1(x_1,\ldots,x_n)}{\delta x_1} & \cdots & \frac{\delta f_1(x_1,\ldots,x_n)}{\delta x_n} \\ \vdots & \ddots & \vdots \\ \frac{\delta f_m(x_1,\ldots,x_n)}{\delta x_1} & \cdots & \frac{\delta f_m(x_1,\ldots,x_n)}{\delta x_n} \end{pmatrix}.$$

In the case of equation (34), ζ evaluated at the estimated value yields

$$\zeta_{II} = \begin{pmatrix} \frac{-\bar{\widehat{\rho_y^2}}}{(\bar{\widehat{\rho_y}})^2} & \frac{1}{(\bar{\widehat{\rho_y}})} & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{-\bar{\widehat{\lambda\rho_{xy}^2}}}{(\bar{\widehat{\lambda\rho_{xy}}})^2} & \frac{1}{(\bar{\widehat{\lambda\rho_{xy}}})} & 0 & 0 \\ 0 & 0 & \frac{-\bar{\widehat{\lambda\rho_{xy}}}}{(\bar{\widehat{\lambda}})^2} & \frac{(\bar{\widehat{\lambda\rho_{xy}}})}{\bar{\widehat{\lambda}}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{-\bar{\widehat{\rho_x^2}}}{(\bar{\widehat{\rho_x}})^2} & \frac{1}{(\bar{\widehat{\rho_x}})} \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Under the covariance function II, the distribution of the ratios in 34 are

$$\begin{pmatrix} \frac{\bar{\widehat{\rho_y^2}}}{\widehat{\rho_y}} \\ \bar{\widehat{\rho_y}} \\ \frac{\bar{\widehat{\lambda\rho_{xy}^2}}}{\widehat{\lambda\rho_{xy}}} \\ \frac{\bar{\widehat{\lambda\rho_{xy}}}}{\hat{\lambda}} \\ \frac{\bar{\widehat{\lambda\rho_x^2}}}{\widehat{\rho_x}} \\ \bar{\widehat{\rho_x}} \end{pmatrix} \sim N \left(\begin{pmatrix} \rho_y \\ \rho_y \\ \rho_{xy} \\ \rho_{xy} \\ \rho_x \\ \rho_x \end{pmatrix}, \frac{2\zeta_{II}\Upsilon D^{+}(\Theta \otimes \Theta)D^{+'}\Upsilon'\zeta_{II}'}{n-1} \right).$$

The methods defined previously can be used to test the covariance II structure. Let Δ_{II} be defined as

$$\Delta_{II} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

The asymptotic distribution of the test for equation (34) is

$$\Delta_{II} \begin{pmatrix} \frac{\bar{\widehat{\rho_y^2}}}{\widehat{\rho_y}} \\ \bar{\widehat{\rho_y}} \\ \frac{\bar{\widehat{\lambda\rho_{xy}^2}}}{\widehat{\lambda\rho_{xy}}} \\ \frac{\bar{\widehat{\lambda\rho_{xy}}}}{\hat{\lambda}} \\ \frac{\bar{\widehat{\lambda\rho_x^2}}}{\widehat{\rho_x}} \\ \bar{\widehat{\rho_x}} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \frac{2\Delta_{II}\zeta_{II}\Upsilon D^{+}(\Theta \otimes \Theta)D^{+'}\Upsilon'\zeta_{II}'\Delta_{II}'}{n-1} \right).$$

Non rejection of these tests do not necessarily mean the type II hypothesis is true. There are many forms of the matrix that should not reject under the null

hypothesis. However, the objective of this chapter was to derive simple tests to alert the researcher that more complicated models may be able to aid in modeling the data. Tests listed here are designed to rule out hypothesis II through V. If these simple tests do not reject hypothesis II through V, it is an indication that more time should be spent to investigate the more advance models.

Sample Size	Number of X values	Number of Y value	Time Units	ρ_x	ρ_{xy}	ρ_y	Number Rejected
50	2	4	3	0.2	0.2	0.2	1
50	2	4	3	0.2	0.2	0.8	3
50	2	4	3	0.2	0.8	0.8	2
50	2	4	3	0.8	0.2	0.2	7
50	2	4	3	0.8	0.8	0.2	1
50	2	4	3	0.8	0.8	0.8	5
50	2	4	6	0.2	0.2	0.2	0
50	2	4	6	0.2	0.2	0.8	4
50	2	4	6	0.2	0.8	0.8	0
50	2	4	6	0.8	0.2	0.2	0
50	2	4	6	0.8	0.8	0.2	1
50	2	4	6	0.8	0.8	0.8	3
50	2	4	9	0.2	0.2	0.2	3
50	2	4	9	0.2	0.2	0.8	4
50	2	4	9	0.2	0.8	0.8	2
50	2	4	9	0.8	0.2	0.2	3
50	2	4	9	0.8	0.8	0.2	3
50	2	4	9	0.8	0.8	0.8	3
100	2	4	3	0.2	0.2	0.2	4
100	2	4	3	0.2	0.2	0.8	3
100	2	4	3	0.2	0.8	0.8	4
100	2	4	3	0.8	0.2	0.2	4
100	2	4	3	0.8	0.8	0.2	2
100	2	4	3	0.8	0.8	0.8	4
100	2	4	6	0.2	0.2	0.2	3
100	2	4	6	0.2	0.2	0.8	5
100	2	4	6	0.2	0.8	0.8	2
100	2	4	6	0.8	0.2	0.2	3
100	2	4	6	0.8	0.8	0.2	5
100	2	4	6	0.8	0.8	0.8	4
100	2	4	9	0.2	0.2	0.2	2
100	2	4	9	0.2	0.2	0.8	4
100	2	4	9	0.2	0.8	0.8	4
100	2	4	9	0.8	0.2	0.2	4
100	2	4	9	0.8	0.8	0.2	3
100	2	4	9	0.8	0.8	0.8	1

TABLE 44: Rejection Rates for Test $\lambda_1 = \lambda_2 = \ldots = \lambda_t$, Sample sizes 50 to 100

Sample Size	Number of X values	Number of Y value	Time Units	ρ_x	ρ_{xy}	ρ_y	Rejection Rate
150	2	4	3	0.2	0.2	0.2	8
150	2	4	3	0.2	0.2	0.8	5
150	2	4	3	0.2	0.8	0.8	3
150	2	4	3	0.8	0.2	0.2	2
150	2	4	3	0.8	0.8	0.2	3
150	2	4	3	0.8	0.8	0.8	7
150	2	4	6	0.2	0.2	0.2	4
150	2	4	6	0.2	0.2	0.8	5
150	2	4	6	0.2	0.8	0.8	4
150	2	4	6	0.8	0.2	0.2	5
150	2	4	6	0.8	0.8	0.2	4
150	2	4	6	0.8	0.8	0.8	4
150	2	4	9	0.2	0.2	0.2	1
150	2	4	9	0.2	0.2	0.8	7
150	2	4	9	0.2	0.8	0.8	1
150	2	4	9	0.8	0.2	0.2	7
150	2	4	9	0.8	0.8	0.2	2
150	2	4	9	0.8	0.8	0.8	2
200	2	4	3	0.2	0.2	0.2	3
200	2	4	3	0.2	0.2	0.8	7
200	2	4	3	0.2	0.8	0.8	5
200	2	4	3	0.8	0.2	0.2	8
200	2	4	3	0.8	0.8	0.2	9
200	2	4	3	0.8	0.8	0.8	5
200	2	4	6	0.2	0.2	0.2	3
200	2	4	6	0.2	0.2	0.8	5
200	2	4	6	0.2	0.8	0.8	5
200	2	4	6	0.8	0.2	0.2	5
200	2	4	6	0.8	0.8	0.2	2
200	2	4	6	0.8	0.8	0.8	0
200	2	4	9	0.2	0.2	0.2	6
200	2	4	9	0.2	0.2	0.8	1
200	2	4	9	0.2	0.8	0.8	4
200	2	4	9	0.8	0.2	0.2	2
200	2	4	9	0.8	0.8	0.2	4
200	2	4	9	0.8	0.8	0.8	3

TABLE 45: Rejection Rates for Test $\lambda_1 = \lambda_2 = \ldots = \lambda_t$, Sample sizes 150 to 200

Sample Size	Number of X values	Number of Y value	Time Units	ρ_x	ρ_{xy}	ρ_y	Rejection Rate
250	2	4	3	0.2	0.2	0.2	8
250	2	4	3	0.2	0.2	0.8	1
250	2	4	3	0.2	0.8	0.8	2
250	2	4	3	0.8	0.2	0.2	4
250	2	4	3	0.8	0.8	0.2	4
250	2	4	3	0.8	0.8	0.8	2
250	2	4	6	0.2	0.2	0.2	7
250	2	4	6	0.2	0.2	0.8	7
250	2	4	6	0.2	0.8	0.8	2
250	2	4	6	0.8	0.2	0.2	10
250	2	4	6	0.8	0.8	0.2	5
250	2	4	6	0.8	0.8	0.8	3
250	2	4	9	0.2	0.2	0.2	4
250	2	4	9	0.2	0.2	0.8	9
250	2	4	9	0.2	0.8	0.8	3
250	2	4	9	0.8	0.2	0.2	3
250	2	4	9	0.8	0.8	0.2	3
250	2	4	9	0.8	0.8	0.8	4
300	2	4	3	0.2	0.2	0.2	4
300	2	4	3	0.2	0.2	0.8	4
300	2	4	3	0.2	0.8	0.8	2
300	2	4	3	0.8	0.2	0.2	7
300	2	4	3	0.8	0.8	0.2	3
300	2	4	3	0.8	0.8	0.8	5
300	2	4	6	0.2	0.2	0.2	2
300	2	4	6	0.2	0.2	0.8	3
300	2	4	6	0.2	0.8	0.8	6
300	2	4	6	0.8	0.2	0.2	6
300	2	4	6	0.8	0.8	0.2	3
300	2	4	6	0.8	0.8	0.8	4
300	2	4	9	0.2	0.2	0.2	7
300	2	4	9	0.2	0.2	0.8	7
300	2	4	9	0.2	0.8	0.8	6
300	2	4	9	0.8	0.2	0.2	5
300	2	4	9	0.8	0.8	0.2	5
300	2	4	9	0.8	0.8	0.8	3

TABLE 46: Rejection Rates for Test $\lambda_1 = \lambda_2 = \ldots = \lambda_t$, Sample sizes 250 to 300

Sample Size	Number of X values	Number of Y value	Time Units	ρ_x	ρ_{xy}	ρ_y	Rejection Rate
350	2	4	3	0.2	0.2	0.2	0
350	2	4	3	0.2	0.2	0.8	4
350	2	4	3	0.2	0.8	0.8	5
350	2	4	3	0.8	0.2	0.2	5
350	2	4	3	0.8	0.8	0.2	2
350	2	4	3	0.8	0.8	0.8	5
350	2	4	6	0.2	0.2	0.2	5
350	2	4	6	0.2	0.2	0.8	7
350	2	4	6	0.2	0.8	0.8	5
350	2	4	6	0.8	0.2	0.2	6
350	2	4	6	0.8	0.8	0.2	5
350	2	4	6	0.8	0.8	0.8	5
350	2	4	9	0.2	0.2	0.2	2
350	2	4	9	0.2	0.2	0.8	5
350	2	4	9	0.2	0.8	0.8	6
350	2	4	9	0.8	0.2	0.2	5
350	2	4	9	0.8	0.8	0.2	7
350	2	4	9	0.8	0.8	0.8	5
400	2	4	3	0.2	0.2	0.2	5
400	2	4	3	0.2	0.2	0.8	3
400	2	4	3	0.2	0.8	0.8	5
400	2	4	3	0.8	0.2	0.2	6
400	2	4	3	0.8	0.8	0.2	6
400	2	4	3	0.8	0.8	0.8	5
400	2	4	6	0.2	0.2	0.2	3
400	2	4	6	0.2	0.2	0.8	5
400	2	4	6	0.2	0.8	0.8	5
400	2	4	6	0.8	0.2	0.2	7
400	2	4	6	0.8	0.8	0.2	3
400	2	4	6	0.8	0.8	0.8	2
400	2	4	9	0.2	0.2	0.2	3
400	2	4	9	0.2	0.2	0.8	3
400	2	4	9	0.2	0.8	0.8	1
400	2	4	9	0.8	0.2	0.2	1
400	2	4	9	0.8	0.8	0.2	1
400	2	4	9	0.8	0.8	0.8	6

TABLE 47: Rejection Rates for Test $\lambda_1 = \lambda_2 = \ldots = \lambda_t$, Sample sizes 350 to 400

Sample Size	Number of X values	Number of Y value	Time Units	ρ_x	ρ_{xy}	ρ_y	Rejection Rate
450	2	4	3	0.2	0.2	0.2	3
450	2	4	3	0.2	0.2	0.8	5
450	2	4	3	0.2	0.8	0.8	4
450	2	4	3	0.8	0.2	0.2	7
450	2	4	3	0.8	0.8	0.2	5
450	2	4	3	0.8	0.8	0.8	3
450	2	4	6	0.2	0.2	0.2	5
450	2	4	6	0.2	0.2	0.8	5
450	2	4	6	0.2	0.8	0.8	3
450	2	4	6	0.8	0.2	0.2	5
450	2	4	6	0.8	0.8	0.2	3
450	2	4	6	0.8	0.8	0.8	5
450	2	4	9	0.2	0.2	0.2	4
450	2	4	9	0.2	0.2	0.8	3
450	2	4	9	0.2	0.8	0.8	3
450	2	4	9	0.8	0.2	0.2	2
450	2	4	9	0.8	0.8	0.2	4
450	2	4	9	0.8	0.8	0.8	4
500	2	4	3	0.2	0.2	0.2	2
500	2	4	3	0.2	0.2	0.8	5
500	2	4	3	0.2	0.8	0.8	5
500	2	4	3	0.8	0.2	0.2	5
500	2	4	3	0.8	0.8	0.2	8
500	2	4	3	0.8	0.8	0.8	5
500	2	4	6	0.2	0.2	0.2	3
500	2	4	6	0.2	0.2	0.8	5
500	2	4	6	0.2	0.8	0.8	5
500	2	4	6	0.8	0.2	0.2	4
500	2	4	6	0.8	0.8	0.2	6
500	2	4	6	0.8	0.8	0.8	0
500	2	4	9	0.2	0.2	0.2	2
500	2	4	9	0.2	0.2	0.8	4
500	2	4	9	0.2	0.8	0.8	5
500	2	4	9	0.8	0.2	0.2	3
500	2	4	9	0.8	0.8	0.2	1
500	2	4	9	0.8	0.8	0.8	7

TABLE 48: Rejection Rates for Test $\lambda_1 = \lambda_2 = \ldots = \lambda_t$, Sample sizes 450 to 500

BIBLIOGRAPHY

[1] Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis, 2nd Edition*, New York: John Wiley & Sons.

[2] Bakewell, D., and Wit, E. (2007), "Bartlett Correction for Likelihood Ratio Test", *Unpublished report*, 3 pages.

[3] Chaganty, N. R., and Naik, D. N. (2002), "Analysis of Multivariate Longitudinal Data using Quasi-Least Squares", *Journal of Statistical Planning and Inference*, 103, 421-436.

[4] Efron, B. and Tibshirani, R. (1993), *An Introduction to the Bootstrap*, New York: Chapman & Hall.

[5] Harville, D. A. (1997), *Matrix Alegebra From a Statistician's Perspective*, New York: Springer.

[6] Hotelling, H. (1936), "Relations Between Two Sets of Variates", *Biometrika*, 28, 321-377.

[7] Johnson, R. A., and Wichern, D. W. (2002), *Applied Multivariate Statistical Analysis*, New Jersey: Prentice-Hall.

[8] Kettenring, J. R. (1971), "Canonical Analysis of Several Sets of Variables", *Biometrika*, 58, 433-451.

[9] Khattree, R., and Naik, D. N. (2000), *Multivariate Data Reduction and Discrimination with SAS Software.*, North Carolina: Wiley-SAS.

[10] Mardia, K. V., Kent, J. J., and Bibby, J. M. (1979), *Multivariate Analysis*, New York: Academic Press.

[11] Naik, D. N., and Rao, S. (2001), "Analysis of Multivariate Repeated Measures Data with a Kronecker Product Structured Covariance Matrix", *Journal of Applied Statistics*, 28, 91-105.

[12] Ravishanker, N. and Dey, D. (2002), *A First Course in Linear Models Theory*, New York: Chapman & Hall/CRC.

[13] Roy, A., and Khattree, R. (2005), "On implementation of a test for Kronecker Product Covariance Structure for Multivariate Repeated Measures Data", *Statistical Methodology*, 2, 297-306.

[14] Schott, J. R. (1997), *Matrix Analysis for Statistics*, New York: John Wiley & Sons, Inc.

[15] Srivastava, J. (2007), "Canonical Correlation and Correspondence Analysis of Longitudinal Data", *Ph.D. thesis, Department of Mathematics and Statistics, Old Dominion University.*

[16] Srivastava, M. S., Nahtman, T., and von Rosen, D. (2008), "Estimation in general multivariate linear models with Kronecker product covariance structure", *Research Report, Swedish University of Agricultural Sciences*, 21 pages.

[17] Srivastava, M. S., Nahtman, T., and von Rosen, D. (2008), "Models with a Kronecker Product Covariance Structure: Estimation and Testing", *Mathematical Methods of Statistics*, 17(4), 357-370.

[18] Srivastava, J, and Naik, D. N. (2008), "Canonical Correlation Analysis of Longitudinal Data", *Denver JSM 2008 Proceedings, Biometrics Section*, 563-568.

[19] Tan, W. Y. (1973), "Multivariate Studentization and its Applications", *The Canadian Journal of Statistics*, 1, 181-199.

[20] Tim, N. H.(2002) , *Applied Multivariate Analysis*, New York: Springer.

APPENDIX A

THE PROGRAM

```
/*******************************************************
This program performs Bootstrap testing on real data or
 on simulated data. The program provides the
 transformation estimates shown in Chapter II and III.
*******************************************************/

dm'log;clear;output;clear;';
options nodate pageno=1;

*Set up the library directory;
libname NASA "C:\Documents and Settings\518383\My
    Documents\Ray\Backup090110\Ph.D111810\NASA data";

*Input the data for analysis. See below for the required
    analysis format.;
data nasa;
set  nasa.nasacxtrap; if Altitude_m_>400 then delete;
run;

PROC IML;

*AR1 program creates an AR1 matrix given rho and t;
start ar1(t,rho,phi);
phi=shape(0,t,t);
do r=1 to t; do s=1 to t;
if (rho=0 )then  phi[R,S]=0;
if (rho=0 & r=s )then  phi[R,S]=1;
if abs(rho)>0 then phi[R,S]=rho**abs(r-s);
end; end;
finish ar1;
```

```
*Take the square root of matrix a1 and return as matrix
     b1;
start sqrtmat1(a1,b1);
run eigen(value,vector,a1);
b1=shape(0,nrow(a1),nrow(a1));
do d1=1 to nrow(a1);
if value[d1,1]<0 then con=0;
if value[d1,1]>0 then con=sqrt(value[d1,1]);
b1=b1+con*vector[,d1]*t(vector[,d1]);
end;*b is the square root of a;
finish sqrtmat1;

*Nasadata takes the data and differences it, fits it to 4th
degree polynomials then, groups it into groups of 23, and
thins it;
start nasadata(ndtemp2,samplesize);
*Program to Process NASA data;
use Nasa;read all var _all_ into nd1;close Nasa;
*Calculate difference in the data;difference=1;
if difference >0 then nd1=nd1[1+difference:nrow(nd1),]
    -nd1[1:nrow(nd1)-difference,];
ndtemp=nd1;
*Add a time element to the data;
nd3=shape(0,nrow(nd1),1);

*Fit fourth degree polynomial to the data and
    use the residuals;
a=(ndtemp[,1]-j(nrow(ndtemp),1,1)*ndtemp[:,1])/1139.335;
int=j(nrow(a),1,1);a2=a#a;a3=a2#a;a4=a2#a2;
x=int||a||a2||a3||a4;
residuals=ndtemp[,2:7]-x*inv(t(x)*x)*t(x)*ndtemp[,2:7];
nd=ndtemp[,1]||residuals;
```

```
*Group data and thin the group to use only 3 values;
group=23; samplesize=int(nrow(nd1)/group);
do rd=1 to samplesize;

*Thin  the data;
ndy1=nd[1+(rd-1)*group,2:4];ndx1=nd[1+(rd-1)*group,5:7];
ndy2=nd[6+(rd-1)*group,2:4];ndx2=nd[6+(rd-1)*group,5:7];
ndy3=nd[11+(rd-1)*group,2:4];ndx3=nd[11+(rd-1)*group,5:7];
ndtemp2=ndtemp2//(ndy1||ndy2||ndy3||ndx1||ndx2||ndx3);
end;
finish nasadata;

*Create Helmert matrix;
start helmertsub12(dim,helmert12);
helmert12=shape(1/sqrt(dim),dim,dim);
do h=2 to dim;do c=1 to dim;
if c<h then helmert12[h,c]=1/sqrt(h*(h-1));
if c=h then helmert12[h,c]=(1-h)/sqrt(h*(h-1));
if c>h then helmert12[h,c]=0;
end;end;
finish helmertsub12;

*Use the Helert matrix to create covariance matrix.;
start covariance34(vals34,helmert34,cov34);
*vals34 are eigan values, Helmert34 is helmert matrix,
cov34 is output result;
cov34=shape(0,nrow(vals34),nrow(vals34));
do      h34=1 to nrow(vals34);
matrix34=helmert34[,h34]*t(helmert34[,h34]);
constant34=vals34[h34,1];cov34=constant34*matrix34+cov34;
end;
finish covariancc34;

start rannorm45(seed45,rowkron45,mean45,
```

```
    kroncov45,iter45,obs45);
*generate normal random numbers;
*initialize obs45 matrix to 0s and 1 row;
*obs45=shape(0,1,nrow(kroncov45));
c = j(rowkron45,1,seed45);
run sqrtmat1(kroncov45,sqrts45);
do i45=1 to iter45;
*create standard normal zero one data;
norms45 = normal(c);
*multiply by square root of the kronecker product;
norm45=sqrts45*norms45;
*append data so rows are observations;
if obs45[1,1]=0 then obs45=     t(norm45);
else obs45=obs45//t(norm45);
*obs45 is the vector of randomnormal variables with
    kroncov covariance matrix;
end;
finish rannorm45;

*Heristic estimate p is the time and the number of rows
of phi q is the total of multivariate variables in the
 matrix;
*This subroutine requires Obs56 p56 q56 returns diff56 s56;
start estimation56(obs56,p56,q56,diff56,s56);
*average of all matrices;
xbar56=obs56[:,];S56=shape(0,p56,p56);
n56=nrow(obs56);*number of rows of data;
diff56=shape(0,nrow(obs56),ncol(obs56));
do i56 = 1 to n56;
diff56[i56,]=obs56[i56,]-xbar56;
end;
do k56=1 to n56;do overq=1 to q56;
a56=(overq-1)*p56+1;a562=overq*p56;
b56=k56;m56=diff56[b56,a56:a562];
```

```
s56=t(m56)*m56/n56/q56+s56;
end;end;
finish estimation56;

*This subroutine puts the phixy estimate into an ar1
structure;
start ar1assumption(phixy,phixyout,t,n);
b1=shape(0,t,t);b2=b1;b3=b1;
do i=1 to t;
if i>1 then b2[i,i-1]=1;
if i>1 then b2[i-1,i]=1;
if (i>1 & i<t) then b3[i,i]=1;
end;
Sn=phixy;BB1=trace(Sn); BB2=trace(b2*Sn);BB3=trace(b3*Sn);
p=shape(0,1,4); p[1,1]= -.5*(2*n*t-2*n);p[1,2]= bb2/2;
p[1,3]= -.5*(2*bb1+2*BB3+2*n-2*n*t);p[1,4]= bb2/2;
r=polyroot(p);
do ch=1 to 3;if r[ch,2]=0 then r1=r[ch,1];      end;
if r1>1 then r1=.999999999;      if r1<-1 then r1=-.99999999;
call ar1(t,r1,phixyout);
finish ar1assumption;

*The program creates Phi from Sigma. q is x
    while p is the time.;
start mlephi67(s67,phi67,diff67,q67);
t=q67;  phi67=shape(0,q67,q67);
p67=ncol(diff67)/q67; x67=shape(0,p67,q67);
n=nrow(diff67);
*Step through all observations k67 is the
    observation number;
do k67=1 to nrow(diff67);
*construct each x (p x q) matrix;
do overq=1 to q67;
*m67 consists of qth column of vecoter;
```

```
x67[,overq]=t(diff67[k67,(overq-1)*p67+1:overq*p67]);
end;
phi67=t(x67)*inv(s67)*x67/p67/nrow(diff67)+phi67;
end;
finish  mlephi67;

*Mle of sigma. It takes data and puts it into the correct
     form. for the MLE estimate;
start mlesig7(s7,phi7,diff7,q7);
p7=ncol(diff7)/q7;
s7=shape(0,p7,p7);
x7=shape(0,p7,q7);
do k7=1 to nrow(diff7);
*construct each x (p x q) matrix;
do overq=1 to q7;
*m67 consists of qth column of vecoter;
x7[,overq]=t(diff7[k7,(overq-1)*p7+1:overq*p7]);
end;
s7=x7*t(inv(phi7))*t(x7)/nrow(diff7)/q7+s7;
end;
finish  mlesig7;

start mlesings(diff,nrowy,nrowx,t,phixy,sxy);
yic=shape(0,nrowy,t);xic=shape(0,nrowx,t);
n=nrow(diff);sxy=shape(0,nrowy,nrowx);
sxyt=shape(0,nrowy,nrowx);
do i=1 to n;do j=1 to t;
yic[,j]=t(diff[i,nrowy*(j-1)+1:nrowy*j]);
end;
do j=1 to t;
xic[,j]=t(diff[i,nrowy*t+nrowx*(j-1)+1:nrowy*t+nrowx*j]);
end;
sxyt=(yic)*t(inv(phixy))*t(xic)/n/t+sxyt;
sxy=(yic)*(inv(phixy))*t(xic)/n/t+sxy;
```

```
end;
finish mlesings;

**Subroutine final transformation**;
*This subroutine transforms the data and uses the SVN2008
solution to gain an estimate of the Psixy matrix.;

start finaltrans(phyy,syy,phxx,sxx,diff,kromodelxy,
    yn,xn,t,phixyest,sigmaxyest);
k=1;conv=10;sigmaxyest=shape(0,nrow(syy),nrow(sxx));
do while (conv>.0001);
sigmaxyestold=sigmaxyest;
*create total=[A 0,0 B] matrix;
naiky=phyy@syy;naikx=phxx@sxx;
zero=shape(0,nrow(naiky),nrow(naikx));
total=(naiky||zero)//(t(zero)||naikx);
*transform diff using isqtot=[A 0, 0 B]^-1/2;
call sqrtmat1(inv(total),isqtot);
*transformed data to give I D, D I covariance matrix;
IDDIdata=t(isqtot*t(diff));
*estimate actual matrix [I D, D I];
IDDI=t(iddidata)*iddidata/(nrow(diff));
*create sqruare root of iddi isiddi=hat[I D,D I]^-1/2;
call sqrtmat1(inv(iddi),isqiddi);
*This creates Idata has hat[I 0, 0 I] covariance matrix;
Idata=t(isqiddi*t(iddidata));
*Dmatrix will have hat[0 D, D 0]. Procedure below makes
    this happen. This create matrix from results 0 phi
    sigma phi sigma 0;
Dmatrix=iddi;
*replace the yy values which reside in the upper right
    hand corner.;
do aa=1 to yn*t;do bb=1 to yn*t;Dmatrix[aa,bb]=0;end;end;
*replace the xx values which reside in the bottom right
```

```
    hand corner.;
do aa=1 to xn*t;do bb=1 to xn*t;
Dmatrix[aa+yn*t,bb+yn*t]=0;      end;end;
*datac has matrix [DDp 0, 0 DpD];
Ddata=t(Dmatrix*t(idata));yc=Ddata[,1:yn*t];
xc=Ddata[,yn*t+1:yn*t+xn*t];
*use estimates from above for phyy syy and phxx sxx;
esxd=t(xc)*xc/nrow(xc);
*Get initial estimate;dpdS=esxd[1:xn,1:xn];
*bottom right corner is 1 on phi;
dpdS=esxd[(t-1)*xn+1:(t-1)*xn+xn,(t-1)*xn+1:(t-1)*xn+xn];
call sqrtmat1(phyy,sqphyy);
call sqrtmat1(inv(phyy),isqphyy);
call sqrtmat1(phxx,sqphxx);
call sqrtmat1(inv(phxx),isqphxx);
md=1; i=0;
do while (md>.001);
i=i+1;dpdSold=dpdS;
*use dpdP and dpdS as initial estimate for sxy and phixy;
run mlephi67(dpdS,dpdP,xc,t);
*dP is an estimate of phi piece of dp=b^-1/2 c a^-1/2;
call sqrtmat1(dpdP,dP);
*remove b and a elements to get C estimate;
cp=sqphyy*dP*sqphxx;cP=sqphxx*dP*sqphyy;
*print cP;
*The model estimates the AR(1) structure;
cPin=cP*nrow(xc);call ar1assumption(cPin,cPout,t,nrow(xc));
cP=cPout;
*Insures the resulting matrix is correlation if AR1
    is present this does nothing;
cp=sqrt(inv(diag(cp)))*cp*sqrt(inv(diag(cp)));  cP[t,t]=1;
dp=isqphxx*cp*isqphyy;  dpdP=dP*t(dP);
    run mlesig7(dpdS,dpdP,xc,t);
mdm=abs(dpdSold-dpdS);  md=mdm[+,+];if i>2000 then md=0;
```

```
end;
*this line gives estimate of CP;
phixyest=cp;
*generate estimate of;
call mlesings(diff,yn,xn,t,phixyest,sigmaxyest);
call sqrtmat1(inv(sxx),isqsxx);
call sqrtmat1(inv(syy),isqsyy);
testsigma1=(isqsxx)*t(sigmaxyest)*t(isqsyy)
    *t(isqsyy)*(sigmaxyest)*t(isqsxx);
k=k+1;  converge=abs(sigmaxyestold-sigmaxyest);
    conv=converge[+,+];
kromodelxy=phixyest@sigmaxyest; if k>2000 then conv=0;
if k>2000 then print "final method failed";
end;
finish finaltrans;

*This program provides the hypothesis tests statistics;
start test(diff,H1model, iter,H0model,xn, yn,t,statistic,
pvalue,df);
dth0=det(h1model);      dth1=det(h0model);      logexp=-10000;
if (det(h1model)>0 & det(h0model)>0) then
logexp=-iter*log(det(H0model))/2-trace(diff*inv(H0model)
    *t(diff))/2+iter*log(det(H1model))/2+trace(diff*
    inv(H1model)*t(diff))/2;
ho=     iter*log(det(H0model))/2-trace(diff*inv(H0model)
    *t(diff))/2;
h1=     iter*log(det(H1model))/2+trace(diff*inv(H1model)
    *t(diff))/2;
if det(h1model)<=0 then print "Determinate H1 <=0";
if det(h0model)<=0 then print "Determinate H0 <=0";
statistic=(iter-2)/iter*-2*logexp;
pvalue=1-CDF('CHISQUARE',statistic,df);
nhere=nrow(diff);
finish test;
```

```
start bootcheck(boot0a,boot1a,stats);
b=nrow(boot1a); c=ncol(boot1a); stats=shape(0,1,c);
do i=1 to b;do j=1 to c;
if boot0a[1,j]>boot1a[i,j] then
    stats[1,j]=stats[1,j]+1/b;
end;end;
finish;


\*
This is the main section of the program. It can be used
to generate the matrices for the Kronecker product.
It also allows the input of data to be analyzed using
the transformation estimate and tested using
Bootstrapping.
*\
start main(iter,sims,bkronecker,bstrap,bootout,ndtemp,
    inputdata,ho,xn,yn,t,rhoxx,rhoxy,rhoyy);
do simulations=1 to sims;
*number of iterations= sample size;samplesize=iter;
*print simulations sims;
*seed is used for the random number generator;seed=iter;
*mean is the mean of random variables;  mean=0;
*x must be less than or equal to y;
RUN ar1(t,rhoxx,phiXX); *generate phiXY AR(1) structure;
RUN ar1(t,rhoxy,phiXY); *generate phiYY AR(1) structure;
RUN ar1(t,rhoyy,phiYY);
eigenvalues1={12,1,8,6,5,4,5};eigenvalues=t(eigenvalues1);
eigenvaluest=t(eigenvalues1);q=nrow(eigenvalues1);
*generate helmert matricies;
run helmertsub12(nrow(eigenvalues1),helmert1);
*generate covariance matrix from helmert and eiganvalues;
run covariance34(eigenvalues1,helmert1,covarianceq);
```

```
*SIGMAXX;SIGMAXX=COVARIANCEQ[yn+1:xn+yn,yn+1:xn+yn];
*SIGMAYY;SIGMAYY=COVARIANCEQ[1:yn,1:yn];
*SIGMAYX;SIGMAYX=COVARIANCEQ[1:yn,yn+1:xn+yn];
call svd(uyx,eyx,vyx,sigmayx);
*sigma is the covariance matrix for multiple observations
at a single time interval.;
YY=PHIYY@SIGMAYY;YX=PHIXY@SIGMAYX;XX=PHIXX@SIGMAXX;
TOP=YY||YX;      BOTTOM=t(yx)||XX;PHISIGMA=TOP//BOTTOM;
*if bstrap=1 then PHISIGMA value comes from outside the
    program. This is the estimated matrix from a previous
    data set;
if bstrap=1 then PHISIGMA=bkronecker;
call svd(us,eigen,vs,phisigma);
rowkron=nrow(phisigma); obs=shape(0,1,nrow(phisigma));
run rannorm45(seed,rowkron,mean,phisigma,iter,obs);
*determines if the data set will be input and if so
assigns the input data to obs which was the sample data;
if inputdata=1 then obs=ndtemp; qxx=t;
*subset the data data exists as n by Yt+xt matrix,1 to yt
     are the y observations;
y=obs[,1:nrow(yy)];      *and yt+1 to yt+1+xt are the X
    observations;
x=obs[,nrow(yy)+1:nrow(phisigma)];
run estimation56(y,nrow(sigmayy),t,diffy,syy);

* For Hypothesis 3 termed 3a below assume RhoYY equals
   I_t so only one iteration is needed;
syy3a=syy;       phyy3a=I(t);
run mlesig7(syy3A,phyy3A,diffy,t);
kromodelyy3a=phyy3a@syy3a;
*gives the MLE of model 2 with phyy unrestricted;
md=1; ii=0;
do while (md>.000000001);
ii=ii+1; syyold=syy;run mlephi67(syy,phyy,diffy,t);
```

```
run mlesig7(syy,phyy,diffy,t);
*AR1 structure;
phyyin=phyy*nrow(diffy);
call ar1assumption(phyyin,phyyout,nrow(phiyy)
    ,nrow(diffy));
phyy=phyyout;
phyy=sqrt(inv(diag(phyy)))*phyy*sqrt(inv(diag(phyy)));
mdm=abs(syyold-syy);md=mdm[+,+];
if ii>2000 then md=0;
end;
kromodelyy=phyy@syy;      rhoyyest=phyy[1,2];
syys=shape(0,1,yn*(yn+1)/2);
syysactual=shape(0,1,yn*(yn+1)/2);
do r= 1 to yn;do c=1 to r;
syysactual[1,(r-1)*r/2+c]=sigmayy[r,c];
syys[1,(r-1)*r/2+c]=syy[r,c];
end;end;
ryyt=ryyt//rhoyyest;syyt=syyt//syys;
*This section estimates Psixx and Sigmaxx;
;
run estimation56(x,nrow(sigmaxx),t,diffx,sxx);
sxx4=sxx;phxx4=I(t);
*gives the MLE of model 3a model with phyy = I;
run mlesig7(sxx4,phxx4,diffx,t);kromodelxx4=phxx4@sxx4;
md=1; ii=0;
do while (md>.000000001);
ii=ii+1;sxxold=sxx;
run mlephi67(sxx,phxx,diffx,t);
run mlesig7(sxx,phxx,diffx,t);
phxxin=phxx*nrow(diffx);
call ar1assumption(phxxin,phxxout,nrow(phixx),nrow(diffx));
phxx=phxxout;
phxx=sqrt(inv(diag(phxx)))*phxx*sqrt(inv(diag(phxx)));
phxx[t,t]=1;mdm=abs(sxxold-sxx);md=mdm[+,+];
```

```
if ii>2000 then md=0;
end;
kromodelxx=phxx@sxx;rhoxxest=phxx[1,2];
sxxs=shape(0,1,xn*(xn+1)/2);
sxxsactual=shape(0,1,xn*(xn+1)/2);
do r= 1 to xn;  do c=1 to r;
sxxsactual[1,(r-1)*r/2+c]=sigmaxx[r,c];
sxxs[1,(r-1)*r/2+c]=sxx[r,c];
end;end;
*print sxxs rhoxx;*print sxx phxx t, sigmaxx phixx;
*cumulate the values for summarization statistics;
rxxt=rxxt//rhoxxest;sxxt=sxxt//sxxs;
run estimation56(obs,nrow(sigmaxx)+nrow(sigmayy),
    t,diff,s);
kroest=t(diff)*diff/iter; *estimate unrestricted Cov;
*use small upper left hand corner piece for sigma;
sigmaxyest=t(diff[,1:nrow(sigmayy)])*diff[,nrow(yy)+1
    :nrow(yy)+nrow(sigmaxx)]/iter;
run estimation56(obs,nrow(sigmaxx)+nrow(sigmayy),
    t,diff,s);
kroest=t(diff)*diff/iter;
*use small upper left hand corner piece for sigma ;

*rearrange collumns to change estimated matrix
From    y11 y21 y31 ... yp1 ... y21 ... ypt x11 ...  xqt
to              y11 .. yn1 x11 .. xq1 y12 ... yn2 x12 ... xq2
such that all time point 1 values are together, then time
    point 2 values and so on.;
b3diff=diff;
do i=1 to t;
b3diff[,(i-1)*(xn+yn)+1:(i-1)*(xn+yn)+yn]
    =diff[,(i-1)*yn+1:i*yn];
b3diff[,1+(i-1)*(xn+yn)+yn:i*(xn+yn)]
    =diff[,yn*t+(i-1)*xn+1:yn*t+i*xn];
```

```
end;
*Initial estimate ofr s3b;        s3b=s;md=1; ii=0;
do while (md>.0001);
ii=ii+1;s3bold=s3b;
run mlephi67(s3b,ph3b,b3diff,t);
run mlesig7(s3b,ph3b,b3diff,t);
ph3bin=ph3b*nrow(diff);
call arlassumption(ph3bin,ph3bout,nrow(ph3b),nrow(diff));
ph3b=ph3bout; ph3b[t,t]=1;
mdm=abs(s3bold-s3b);md=mdm[+,+];
if ii>2000 then md=0;
end;
*estimate of model kroneckerxx;
s3byy=s3b[1:yn,1:yn];s3byx=s3b[1:yn,yn+1:xn+yn];
s3bxx=s3b[yn+1:yn+xn,yn+1:yn+xn];
kromodel3b=( (ph3b@s3byy)||(ph3b@s3byx) )//
    ( (ph3b@t(s3byx))||(ph3b@s3bxx) );
rho3best=ph3b[1,2];
L3b=- iter*log(det(kromodel3b))/2
    -trace(diff*inv(kromodel3b)*t(diff))/2;
phixyest5=I(t);
run mlesig7(sigmaxyest5,phixyest5,b3diff,t);

s5yy=sigmaxyest5[1:yn,1:yn];
s5yx=sigmaxyest5[1:yn,yn+1:xn+yn];
s5xx=sigmaxyest5[yn+1:yn+xn,yn+1:yn+xn];
kromodel5=( (phixyest5@s5yy)||(phixyest5@s5yx) )
    //( (phixyest5@t(s5yx))||(phixyest5@s5xx) ); ;
call finaltrans(phyy,syy,phxx,sxx,diff,kromodelxy2,
    yn,xn,t,phixy2,sigmaxy2);
call finaltrans(phyy3a,syy3a,phxx,sxx,diff,kromodelxy3a
    ,yn,xn,t,phixy3a,sigmaxy3a);
call finaltrans(phyy3a,syy3a,phxx4,sxx4,diff,kromodelxy4
    ,yn,xn,t,phixy4,sigmaxy4);
```

```
kromodel4= (kromodelyy3a||kromodelxy4)
    //(t(kromodelxy4)||kromodelxx4)      ;
kromodel3a= (kromodelyy3a||kromodelxy3a)
    //(t(kromodelxy3a)||kromodelxx)      ;
kromodel2= (kromodelyy||kromodelxy2)
    //(t(kromodelxy2)||kromodelxx)       ;
Kromodel1= kroest;
*Calculate determinates of estimates;
det1=det(kromodel1);det2=det(kromodel2);
 det3a=det(kromodel3a);det3b=det(kromodel3b);
 det4=det(kromodel4);det5=det(kromodel5);
pvalue12=0;pvaluc23a=0;pvalue3a4=0;pvaluc45=0;
pvalue23b=0;pvaluc3b5=0;
*test model 2 vs model 3b;
*model 2 vs 3b; df23b=1+(yn+1)*yn/2+1+yn*xn+1+(xn+1)
    *xn/2-1+(yn+1)*yn/2+yn*xn+(xn+1)*xn/2;
if (det2>0 & det3b>0) then call test(diff,kromodel2,
 iter,kromodel3b , xn, yn,t,statistic,pvalue23b,df23b);
if simulations=1 then count23b=0;
if (det2>0 & det3b>0 & pvalue23b<.05) then
    count23b=count23b+1;
if (det2>0 & det3b>0) then
    pvaluet23b=pvalue23b//pvaluet23b;
*test model 3b vs model 5;*model 3b vs 5;
df3b5= ( (yn+1)*yn/2+(xn+1)*xn/2+yn*xn+1 ) -
    ( (yn+1)*yn/2+(xn+1)*xn/2+yn*xn );
if (det5>0 & det3b>0) then call test(diff,kromodel3b,
    iter,kromodel5 , xn, yn,t,statistic,pvalue3b5,df3b5);
*if det2<0 then pvalue12=0;      if simulations=1 then
    count3b5=0;
if (det5>0 & det3b>0 & pvalue3b5<.05) then
    count3b5=count3b5+1;
if (det5>0 & det3b>0) then pvaluet3b5=pvalue3b5
```

```
    //pvaluet3b5;
*test model 1 vs model 2;*model 1 vs 2;
df12=(yn+xn)*t*((yn+xn)*t+1)/2-
    ( (yn+1)*yn/2+(xn+1)*xn/2+yn*xn+3 );
if (det1>0 & det2>0) then call test(diff,kromodel1,iter,
    kromodel2 , xn, yn,t,statistic,pvalue12,df12);
if simulations=1 then count12=0;
if (det1>0 & det2>0 & pvalue12<.05) then
    count12=count12+1;
if (det1>0 & det2>0) then pvaluet12=pvalue12//pvaluet12;

*test model 2 vs model 3;*model 2 vs 3;
df23a=( (yn+1)*yn/2+(xn+1)*xn/2+yn*xn+3 )-
    ((yn+1)*yn/2+(xn+1)*xn/2+1+yn*xn+1);
if (det2>0 & det3a>0) then call test(diff,kromodel2,
    iter,kromodel3a , xn, yn,t,statistic,pvalue23a,df23a);
if simulations=1 then count23a=0;
if (det2>0 & det3a>0 & pvalue23a<.05) then
    count23a=count23a+1;
if (det2>0 & det3a>0) then pvaluet23a=
    pvalue23a//pvaluet23a;

*test model 3 vs model 4;*model 3 vs 4;
df3a4=((yn+1)*yn/2+(xn+1)*xn/2+1+yn*xn+1)-
    ((yn+1)*yn/2+(xn+1)*xn/2+yn*xn+1);
if (det3a>0 & det4>0) then call test(diff,kromodel3a,
    iter,kromodel4 , xn, yn,t,statistic,pvalue3a4,df3a4);
if simulations=1 then count3a4=0;
if (det3a>0 & det4>0 & pvalue3a4<.05) then
    count3a4=count3a4+1;
if (det3a>0 & det4>0) then pvaluet3a4=
    pvalue3a4//pvaluet3a4;

*test     model 4 vs model 5;*model 4 vs 5;
```

```
df45=((yn+1)*yn/2+(xn+1)*xn/2+yn*xn+1)-
     ((yn+1)*yn/2+(xn+1)*xn/2+yn*xn);
if (det4>0 & det5>0) then call test(diff,kromodel4,
     iter,kromodel5 , xn, yn,t,statistic,pvalue45,df45);
if simulations=1 then count45=0;
if (det4>0 & det5>0 & pvalue45<.05) then count45=count45+1;
if (det4>0 & det5>0) then pvaluet45=pvalue45//pvaluet45;


df3a5=((yn+1)*yn/2+(xn+1)*xn/2+yn*xn+2)-
     ((yn+1)*yn/2+(xn+1)*xn/2+yn*xn);
if (det3a>0 & det5>0) then call test(diff,kromodel3a,
     iter,kromodel5 , xn, yn,t,statistic,pvalue3a5,df3a5);
pvalue3a5=0;if simulations=1 then count3a5=0;
if (det3a>0 & det5>0 & pvalue3a5<.05) then
     count3a5=count3a5+1;
if (det3a>0 & det5>0) then pvaluet3a5=
     pvalue3a5//pvaluet3a5;

n12=nrow(pvaluet12);n23a=nrow(pvaluet23a);
n3a4=nrow(pvaluet3a4);n45=nrow(pvaluet45);
n23b=nrow(pvaluet23b);n3b5=nrow(pvaluet3b5);
n3a5=nrow(pvaluet3a5);


**For Boot Strapping, need to have the test statistics***;
*if bstrap=0 this is the first run of the program. The
simulation is estimating the first matrix to run the
bootstrap samples. It is important for this matrix to be
output to the simulation.
if bstrap=1 bkronecker is being input into the simulaiton
and should not change between runs.;
* this sets bkronecker to deliver it outside the model.
if bstrap=1 it does not change bkronecker becuase we
```

```
need to use the same one.;

if (bstrap=0 & ho=2) then bkronecker=kromodel2;
if (bstrap=0 & ho=3) then bkronecker=kromodel3a;
if (bstrap=0 & ho=4) then bkronecker=kromodel3b;
if (bstrap=0 & ho=5) then bkronecker=kromodel5;
boot=shape(0,1,5);
if ho=2 then call test(diff,kromodel1,iter,kromodel2 , xn,
    yn,t,statistic12 ,pvalue12 ,df12);
if ho=3 then call test(diff,kromodel2,iter,kromodel3a, xn,
    yn,t,statistic23a ,pvalue23a ,df23a);
if ho=5 then call test(diff,kromodel3a,iter,kromodel5 , xn,
    yn,t,statistic3a5 ,pvalue3a5 ,df3a5);
if ho=4 then call test(diff,kromodel2, iter,kromodel3b, xn,
    yn,t,statistic23b ,pvalue23b ,df23b);
if ho=5 then call test(diff,kromodel3b,iter,kromodel5 , xn,
    yn,t,statistic3b5 ,pvaluc3b5 ,df3b5);

if bstrap=0 then Print"Covariance Matrix II";
psiy2=phyy; sigmay2=syy; psixy2=phixy2; sigmaxy2=sigmaxy2;
psix2=phxx; sigmax2=sxx;
if bstrap=0 then print psiy2 sigmay2 psixy2 sigmaxy2 psix2
    sigmax2;


if bstrap=0 then Print"Covariance Matrix III";
psiy3=phyy3a; sigmay3=syy3a; psixy3=phixy3a;
sigma3=sigmaxy3a; psix3=phxx; sigmax3=sxx;
if bstrap=0 then print psiy3 sigmay3 psixy3 sigma3
psix3 sigmax3;

if bstrap=0 then Print"Covariance Matrix IV";
sigmay4=s3byy; sigmaxy4=s3byx; sigmax4=s3bxx; psi4=ph3b;
if bstrap=0 then print  sigmay4 sigmaxy4 sigmax4 psi4;
```

```
if bstrap=0 then Print"Covariance Matrix V";
sigmay5=s5yy; sigmaxy5=s5yx; sigmax5=s5xx;
if bstrap=0 then print sigmay5 sigmaxy5 sigmax5;

if ho=2 then boot[1,1]=statistic12;
if ho=3 then boot[1,2]=statistic23a;
if ho=5 then boot[1,3]=statistic3a5;
if ho=4 then boot[1,4]=statistic23b;
if ho=5 then boot[1,5]=statistic3b5;

if simulations=1 then bootout=boot;
if simulations>1 then bootout=bootout//boot;

**Summary Statistics***;
*record estimate of phixy phixy2,sigmaxy2 ;
rhoxyest=phixy2[1,2];sxys=shape(0,1,xn*yn);
sxysactual=shape(0,1,xn*yn);
rxyest=shape(0,1,t*t);rxysactual=shape(0,1,t*t) ;
*load sample values into matrix put sigmaxyest into row
vector;
do r=1 to nrow(phixy2);do c=1 to nrow(phixy2);
rxysactual[1,(r-1)*ncol(phixy2)+c]=phixy[r,c];
rxyest[1,(r-1)*ncol(phixy2)+c]=phixy2[r,c];
end;end;

do r= 1 to nrow(sigmaxyest);do c=1 to ncol(sigmaxyest);
sxysactual[1,(r-1)*ncol(sigmaxyest)+c]=sigmayx[r,c];
sxys[1,(r-1)*ncol(sigmaxyest)+c]=sigmaxy2[r,c];
end;end;
*vectors of the solutions 2 estimates transformation
    method rxt has all the rhoxy values sxyt has all the
    sigma sxyt values;
rxyt=rxyt//rhoxyest;sxyt=sxyt//sxys;kro=phixy@sigmayx;
```

```
end;
finish main;
start main1;
/*
*start main(samplesize,sims,bkronecker,bstrap,bootout,
    ndtemp,inputdata,ho,xn,yn,t,rhoxx,rhoxy,rhoyy);
sample size is the number of complete samples in the data.
sims is the number of simulations to run within main.
for real data make this 1 to read in the first data set
get the main statistics.
for the bootstrap samples make it b;
bkronecker is the bootstrapping kronecker product matrix.
if bstrap=0 the bkronecker value is generated by the
program either calculated from outside data or from
simulated data
if bstrap=1 then bkronecker matrix is generated outside
the program and read in.
ndtemp is the data to be input into the program.
if inputdata=1 data is input from ndtemp.
if inputdata=0 data is not input into the program.
Data must be in the form
y11y21...yq1...y1t...yqt x11...xp1...x1t...xpt;
for bootstrapping inputdata=1 for the first run and
inputdata=0 for the other runs;
ho is the null hypothesis to test. II=2 III=3 IV=4 V=5
the pvalue is set to 1 for all other tests.
The pvalue for the test of interest is the only value
correct in the output
xn yn t are the number of data points in the x y
multivariate vectors and t is the number of time periods.
These inputs mean the same thing all the time
rhoxx,rhoxy,rhoyy assign values to the respective rho
values. these are only used if bstrap=0 and inputdata=0
*/
```

```
xn=3;yn=3;T=3;RHOxx=.7;RHOxy=.8;RHOyy=.7;
samplesize=53;B=300;count=0; Hyp=4;
*simul is the number of simulations to run. For real data
this value is 1 to input the data and 1 of b simulations
within main to get bootstrap tests. To run simulations
this value should be higher.;
simul=1;
*Input data here will change the sample size;
run nasadata(ndtemp,samplesize);
statcount=shape(0,1,5);
do s=1 to simul;
*calls the first iteration to estimate the inital cov var
     matrix from the data;
call main(samplesize,1,bkron,0,boot0,ndtemp,1,Hyp,xn,yn,
    t,rhoxx,rhoxy,rhoyy);
*perform the bootstrap calculations;
call main(samplesize,b,bkron,1,boot1,ndtemp,0,Hyp,xn,yn,
    t,rhoxx,rhoxy,rhoyy);
call bootcheck(boot0,boot1,stat);*print stat;
do i=1 to 5;
*This looks at all the pvalues and count the one below
.05. if simul =1 as in real data this will be 1 or 0;
if stat[1,i]>.95 then statcount[1,i]=statcount[1,i]+1/simul;
end;
end;
Bootpvalue=1-stat;
print 'IvsII IIvsIII IIIvsV IIsIV IVvsV';
*only the bootstrap value for the test specified in hyp
is correct. The  others are set to 1;
print Bootpvalue;
if simul>1 then print stat;
finish main1;
run main1;quit;
```

VITA

Raymond McCollum
Department of Mathematics and Statistics
Old Dominion University
Norfolk, VA 23529

Education

Ph.D. Old Dominion University, Norfolk VA (December 2010)
Major: Computational and Applied Mathematics [Statistics]

M.S. Kansas State, Manhattan KS (May 2000)
Major: Statistics

B.S. Old Dominion University, Norfolk VA (August 1996)
Major: Computational and Applied Mathematics [Statistics]

Experience

Senior Statistician (March 2009 - December 2010)
Booz Allen Hamilton, Norfolk, VA

Operations Research Analyst (May 2002 - February 2009)
Surface Warfare Development Group U.S., Navy, Norfolk, VA

Statistician (August 2000 - May 2002)
United Technologies Pratt and Whitney, East Hartford, CT

Typeset using LaTeX.