Summer 8-1-2019

Optimal Design for a Causal Structure

Zaher Kmail
University of Nebraska-Lincoln, zaher.m.kmail@gmail.com

Follow this and additional works at: https://digitalcommons.unl.edu/statisticsdiss

Part of the Applied Statistics Commons, Design of Experiments and Sample Surveys Commons, Multivariate Analysis Commons, Statistical Methodology Commons, and the Statistical Theory Commons

https://digitalcommons.unl.edu/statisticsdiss/21

This Article is brought to you for free and open access by the Statistics, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Dissertations and Theses in Statistics by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Optimal Design for a Causal Structure

by

Zaher M. Kmail

A DISSERTATION

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

Major: Statistics

Under the Supervision of Professor Kent M. Eskridge

Lincoln, Nebraska
August 2019
Optimal Design for a Causal Structure

Zaher Kmail, PhD

University of Nebraska, 2019

Advisor: Kent M. Eskridge

Linear models and mixed models are important statistical tools. But in many natural phenomena, there is more than one endogenous variable involved and these variables are related in a sophisticated way. Structural Equation Modeling (SEM) is often used to model the complex relationships between the endogenous and exogenous variables. It was first implemented in research to estimate the strength and direction of direct and indirect effects among variables and to measure the relative magnitude of each causal factor.

Historically, traditional optimal design theory focuses on univariate linear, nonlinear, and mixed models. There is no current literature on the subject of optimal design for a causal structure, therefore this research is the first contribution in the field. There are five objectives for this dissertation research. For a given causal structure, the objectives of this research are to obtain an optimal design: (1) For a completely randomized experiment that produces the most precise estimates for the endogenous and exogenous parameters, (2) For an experiment with random blocks or split-plots that produces the most precise estimates for the endogenous and exogenous parameters, (3) For an experiment with fixed blocks that produces the most precise estimates for the endogenous and exogenous parameters, (4) For an experiment with random blocks or split-plots that produces the most precise estimates for the endogenous parameters, exogenous parameters, and the variance components, and (5) Using the methods above to demonstrate the improvement in efficiency for two applications published in previous research.
In each case, the causal relationship dramatically changed the optimal designs. The new optimal designs were more efficient. Even orthogonal designs, which are universally optimal in the univariate case, are not optimal when considering a causal structure.
DEDICATION

To the love of my life, my wife, Lyneea. Thank you for your love, support, and patience through all of the ups and downs. Your encouragement and support was always the biggest motivation to propel me forward.
ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Professor Kent Eskridge, for mentoring me and for guiding me. He took me under his wing and provided me a tremendous amount of feedback and direction. Without his support, encouragement, and valuable feedback, I would not have seen this work come to fruition. He has made a huge difference in my life and I aspire in my professional career to display the sincerity, availability, and dedication that he does. I hope to be a difference-maker in someone else’s life the way that he has been a difference-maker in mine.

I would also like to thank Professor Stephen Kachman, Professor Erin Blankenship, and Professor Allan Peterson for serving on my PhD supervisory committee. I recognize the time and effort that it takes to review a dissertation and provide commentary. I truly appreciated their time, patience, and insight throughout this process.

I appreciate the support and assistance that was provided to me by the staff of the Holland Computing Center. A portion of this work was completed utilizing the supercomputer, Crane, at the Holland Computing Center of the University of Nebraska, which receives support from the Nebraska Research Initiative. My thanks also goes out to the University of Nebraska for supporting and maintaining such a valuable resource.

Last but certainly not least, I would like to thank my family for all of their love, support, and encouragement. This has been a long and arduous journey, and they have been by my side every step of the way.

Zaher Kmail
TABLE OF CONTENTS

List of Tables vii

List of Figures x

1 Introduction 1

1.1 Background 1

1.2 Problem Statement 6

1.3 Research Objectives 8

1.4 References 9

2 Literature Review 12

2.1 Optimal Design for Univariate Linear Models 12

2.1.1 Introduction 12

2.1.2 Optimality Criteria 12

2.1.3 Search Algorithms for Optimal Design 14

2.2 Optimal Design for Univariate Mixed Models 17

2.2.1 Introduction 17

2.2.2 Optimal Design for Mixed Models 18

2.2.2.1 Optimal Design for Random Mixed Models or Split-Plot Models 18

2.2.2.2 Optimal Design for Mixed Models with Fixed Blocks 24
2.3 Structural Equations

2.3.1 Introduction

2.3.2 The Reduced Form of a Causal Structure

2.3.3 Identification of the Parameters of a Causal Structure

2.3.4 Causal Structure Estimating Methods

2.4 Summary

2.5 References

Appendix 2A Derivation of the asymptotic 3SLS information matrix

Appendix 2B Derivation of the FIML information matrix

3 Optimal Design for Endogenous and Exogenous Parameters

3.1 Introduction

3.2 Optimal Design for a Causal Structure Model

3.2.1 3SLS Optimal Design for a Causal Structure

3.2.2 FIML Optimal Design for a Causal Structure

3.3 Optimal Design for a Causal Structure Model with Blocks

3.3.1 3SLS Optimal Design for a Causal Structure with Random Blocks

3.3.2 FIML Optimal Design for a Causal Structure with Random Blocks

3.4 Optimal Design for a Causal Structure with Fixed Blocks

3.4.1 3SLS Optimal Design for a Causal Structure with Fixed Blocks
3.4.2 FIML Optimal Design for a Causal Structure with Fixed Blocks

3.5 Summary

3.6 References

Appendix 3A Simulation Data for Example #1

4 Optimal Design for a Causal Structure for Endogenous Parameters, Exogenous Parameters, and Variance Components

4.1 Introduction to Variance Components in Optimal Design for a Causal Structure with Mixed Effects

4.2 3SLS Simultaneous Optimal Design for a Causal Structure for the Endogenous Parameters, Exogenous Parameters, and Variance Components Using a Composite Criterion

4.3 FIML Simultaneous Optimal Design for a Causal Structure for the Endogenous Parameters, Exogenous Parameters, and Variance Components

4.4 Residual Maximum Likelihood (REML) Simultaneous Optimal Design for a Causal Structure for the Endogenous Parameters, Exogenous Parameters, and Variance Components

4.5 Summary

4.6 References

Appendix 4A ML Information Matrix for Endogenous, Exogenous, and Covariance Estimate

Appendix 4B REML Covariance Structure
5 Applications of Optimal Design for a Causal Structure

5.1 Introduction

5.2 Optimal Design Application in Relation to Analysis of Genotype-by-Environment Interaction in Wheat Using a Structural Equation Model and Chromosome Substitution Lines

5.2.1 Applying 3SLS Methodology

5.2.2 Applying FIML Methodology

5.3 Optimal Design Application in Relation to Some Methods for Analyzing Multi-trait Genotype-by-Environment Interactions

5.3.1 Applying 3SLS Methodology

5.3.2 Applying FIML Methodology

5.4 Summary

5.5 References

6 Conclusion

6.1 Summary

6.2 Future Work

6.3 References

SAS Codes
LIST OF TABLES

Table 3.1 3SLS and FIML optimal design for Example #1 with a determinant of 134,784 56

Table 3.2 Classical optimal design for the univariate model

\[ y = \gamma_1 x_1 + \gamma_2 x_2 + \epsilon \] 58

Table 3.3 3SLS results for three simulations where Column #2 represents the parameters and Columns #3 – 5 represent the estimates of those parameters based on the simulation 58

Table 3.4 Determinants of the covariance matrices for the 3SLS causal structure optimal design and the 3SLS univariate optimal design based on three simulations each with 100,000 data sets 59

Table 3.5 FIML results for three simulations where Column #2 represents the parameters and Columns #3 – 5 represent the estimates of those parameters based on the simulation 61

Table 3.6 A 3SLS causal structure optimal design with random blocks where each block size equal to four 67

Table 3.7 An orthogonal design for four blocks where each block size is equal to four 67

Table 3.8 Asymptotic information matrix for Fig 3.2 based on the orthogonal design 68

Table 3.9 Asymptotic information matrix for Fig. 3.2 based on the 3SLS optimal design 69
| Table 4.1 | A 3SLS optimal design for four blocks with block size equal to two with $\alpha = 0.5$ with no contemporaneous correlation among the endogenous parameters or the observation within block |
| Table 4.2 | A 3SLS optimal design for four blocks with block size equal to two with $\alpha = 0.75$ with no contemporaneous correlation among the endogenous parameters or the observation within block |
| Table 4.3 | A 3SLS optimal design for four blocks with block size equal to two with $\alpha = 0.5$ with contemporaneous correlation among the endogenous parameters and the observation within block |
| Table 4.4 | An ML optimal design for four blocks with block size equal to two |
| Table 4.5 | A REML optimal design for four blocks with block size equal to two |
| Table 4.6 | The univariate optimal design for four blocks with block size equal to two using composite criteria |
| Table 4.7 | Comparison of the determinant of the REML information matrix using different designs |
| Table 5.1 | Genotype combinations for three replicates that were used in the study |
| Table 5.2 | Optimal genotype combinations for three replicates using 3SLS methodology |
| Table 5.3 | Optimal genotype combinations for three replicates using FIML methodology |
Table 5.4  Values of the exogenous variables for each location for mean daily maximum temperature (MXTD), mean daily minimum temperature (MNTD), mean sun hours per day (SHM), and monthly precipitation fall (PRF) (Vargas et al. 1998 & 1999)

Table 5.5  Optimal number of replicates for each location using 3SLS methodology

Table 5.6  Optimal number of replicates for each location using FIML methodology
## LIST OF FIGURES

| Figure 1.1 | The proposed causal model for a maize yield component study published in *Maydica* by Milander et al. (2015) | 7 |
| Figure 1.2 | The data-supported causal model for the maize yield component study published in *Maydica* by Milander et al. (2015) | 7 |
| Figure 2.1 | Causal structure where each exogenous variable directly affects only one endogenous variable | 30 |
| Figure 2.2 | Causal structure where the second exogenous variable directly affects both endogenous variables | 31 |
| Figure 2.3 | Causal structure where the first exogenous variable affects both endogenous variables | 32 |
| Figure 3.1 | The path model for Example #1 | 55 |
| Figure 3.2 | The path model for Example #2 | 67 |
| Figure 4.1 | The path model for Example #3 | 88 |
| Figure 5.1 | The selected path model for the yield genotype-by-environment interaction in the Dhungana et al. study from 2007 | 117 |
| Figure 5.2 | The path model developed for the yield genotype-by-environment interactions in the Yaseen dissertation from 2012 | 131 |
| Figure 6.1 | Path estimates of SEM of gene-by-environment interaction in the development of CHD as published in *Annals of Human Genetics* by Mi et al. (2011) | 144 |
CHAPTER 1: INTRODUCTION

1.1 Background

The definition of an experiment is the process of treating or manipulating experimental material to answer a specific question. However, in many experiments, statisticians are consulted only at the data analysis stage where problems typically present themselves. From data collection to data cleaning, it is assumed that the experiment is carefully planned in advance. However, this is often not the case. Often it is impossible to answer the most important question of the research because of the failure to properly plan and execute the experiment. Montgomery (2001) notes that all experiments are designed experiments, it is just that some are poorly-designed and some are well-designed. Many of the problems that arise in the data analysis stage can be avoided in the planning stage by carefully designing the experiment to answer the question at hand before data are ever even collected. Sir Ronald Fisher (1935, 8) observed that “experimental observations are only experience carefully planned in advance, and designed to form a secure basis of new knowledge.”

The design of experiments may very well be one of the oldest branches of Statistics. In fact, the properties of designing experiments dates back to biblical times. Stigler (1974) in his paper on Gergonne cites the book of Daniel in the Old Testament in which Daniel proved that his own diet was superior to King Nebuchadnezzar's using the king's servants as controls. The 11th-century physician, Avicenna, is also cited for laying out many of the modern principles of experimental design in his second volume of Canon of Medicine, which was the leading medical text for the next eight centuries (Stigler 1974). Avicenna outlined seven principles for medical experimentation, including things such as the need for replications, controls, the need to vary factors one at a time, the importance of avoiding confounding effects, and the importance of observation of effects at different factor levels.
There are many desirable properties by which to evaluate an experimental design. But Montgomery (2001) would suggest that the four most important are:

1. Simplicity
2. Cost effectiveness
3. Unbiasedness
4. Precision

First, the experiment should be simple. Simplicity is an important characteristic of a good experiment because in most research projects, there are more people than just the statistician involved, some of whom may not have a background in Statistics. The more complicated a design is, the more likely it might be that mistakes can be made. If there are multiple designs that give the same information, the simpler design is the preferred one.

Next, the experiment should be cost effective. Experimental design would not necessarily be as important if there were not any budgetary limitations in an experiment. The job of a statistician is to design an experiment that is cost effective or to try to get the maximum amount of information to answer a question within the limitations of a budget. In most cases, the sample size of an experiment is limited by the budget of the experiment. In this case, the objective would be to determine the design that will give the maximum information for the given sample size.

The experiment should also be unbiased. Bias in this sense refers to systematic bias and so the goal is ensure that the data are representative of the population in a given experiment. For example, an experiment that includes only males cannot be representative of the entire population. Similarly, an experiment that includes subjects from only one socio-economic class or only one ethnicity would be inherently biased. Careful planning of the design and randomization will ensure unbiasedness.
Finally, and arguably most importantly, is the characteristic of precision in an experiment. The effect of random errors should be minimized as feasibly as possible (Cox and Reid 2000). The objective is to obtain the design that will produce the most precise estimates for a given budget and to make sure that the planning and randomization are put in place to avoid any systematic error or bias.

The first well-documented contribution to optimal design theory was made by Kirstine Smith in 1918. She explored the regression problem for univariate polynomials of order up to six, with control variable varying between -1 and 1 (Smith 1918). Since then, optimal design theory and its applications have been a big part of Statistics literature.

In the 1920's, the applications of optimal design theory were mostly in agriculture and the biological sciences. In that period, the work of Ronald Fisher (1925), Frank Yates (1935), and David Finney (1955) led the way and their contributions had a fundamental impact on the growth of the theory and the application of experimental design. One of the fundamental contributions by Fisher and Yates (1943) is the use of projective geometry in the construction of designs.

The growth of experimental design theory continued, which led to the development of the area of factorial experiments. Originally, most of the designs that were available were orthogonal designs and very little work had been done on non-orthogonal designs (Mitchell and Bayne 1978). The disadvantage of orthogonal designs is that they require a large sample size. This was not a major concern in the 1920's and 1930's because most of the applications were in agriculture, where large sample sizes were not an issue. However, after World War II, development of new industrial applications required new statistical techniques due to high costs and the complexity of experiments in industrial processes. It is not uncommon in industrial experimentation to have the number of runs (sample size) only just larger than or even equal to the number of model parameters. (Mylona, Goos, and Jones 2014).
Technological advancement has influenced optimal design theory and made it possible to obtain optimal designs under new challenges and constraints. This marked a new research era that was stimulated by the interest in designs that are optimal with respect to some design criteria, usually a function of the variance-covariance matrix of the estimated parameters (Kiefer and Wolfowitz 1959; Altan and Singh 2001). \(A\)-, \(D\)-, and \(E\)-optimality criteria are the most widely used and need search algorithms to obtain an optimal design. Technological advancements made it possible to develop the search algorithms that are needed to obtain the optimal designs of the Sequential, or Dykstra, Method (Dykstra 1971), the Simple Exchange, or Wynn-Mitchell, Method (Mitchell 1974a), the DETMAX, or Exchange with Excursions Algorithm (Mitchell 1974b), the Fedorov Method (Fedorov 1972), and the Modified Fedorov, or Simultaneous Switching, Method (Cook and Nachtsheim 1980).

The new growth in the application of experimental design came about because of the lack of theoretical results in the optimal design of blocked and split-plot experiments, which made the computerized design algorithms popular and led to three proposals for point-exchange algorithms for constructing \(D\)-optimal blocked designs (Cheng 1995; Atkins and Cheng 1999). The first algorithm was proposed by Atkinson and Donev (1989), the second algorithm was proposed by Cook and Nachtsheim (1989), and the third algorithm was proposed by Goos and Vandebroek (2001). For split-plot optimal design, Goos and Vandebroek (2003) presented a point-exchange algorithm to construct a \(D\)-optimal design for a split-plot response surface. Goos and Jones (2012) discussed a coordinate-exchange algorithm for constructing \(I\)-optimal split-plot designs.

The major weakness and criticism for the previous work, especially for the optimal designs for split-plot and blocked designs, is that these designs focused entirely on the fixed effects precision estimates and did not take into account the estimates of the covariance parameters, which may result in poor estimates for the covariance parameters or, in the worst case scenario, render them inestimable (Mylona, Goos, and Jones 2014). This can lead to poor statistical inference or
impossible statistical inference in the worst case scenario. For that reason, there was more emphasis on simultaneous optimal designs that take into account both the fixed effects and the random effects parameters.

Historically, what drove the growth of optimal design theory was the need to have a design that is optimal for all of the parameters that need to be estimated. Most statistical analyses focus on univariate linear models, which have limitations because most random processes involve multiple dependent variables. This, in turn, led to the development of the multivariate linear model.

The multivariate linear model gives insights that are not possible with the univariate linear model by allowing for the analysis of many dependent variables simultaneously which provides a better understanding of the underlying processes as compared to univariate methods. The most common types of multivariate analysis such as cluster analysis, principal component analysis, factor analysis, discriminate analysis, MANOVA, etc. are based on the association among the dependent variables. Although multivariate methods are a step toward a more complete understanding of complex systems, in many natural systems the variables are related in a much more sophisticated way, which led to the development of structural equation models (Jöreskog 1970).

Structural equation modeling (SEM), also called causal structure modeling, has many applications in the fields of agriculture, epidemiology, sociology, and many others. Path analysis, a special case of SEM, was developed by Wright (1921) and applied to crested wheatgrass seed production (Dewey and Lu 1959). Structural equation modeling requires known causal relationships among variables and uses linear regression analysis that allows the causal relationships to be decomposed into direct and indirect effects (Li 1975). In addition, it permits the factors that produce a particular correlation to be examined more closely and measures the relative magnitude of each causal factors (Dewey and Lu 1959).
There are many estimating methods for structural equation modeling, but in this dissertation the focus will be on two methods, the Three-Stage Least Squares (3SLS) and the Full Information Maximum Likelihood (FIML). Both estimation approaches are full-system methods that estimate all parameters simultaneously as opposed to single equation approaches, which can be burdened by complexity and problems associated with overestimation (Schmidt 1976).

1.2 Problem Statement

This work was motivated by a project with Professor Stephen Mason where one of the objectives of the study was to use path analysis to determine the influence of seeding rate on the yield components of waxy maize, and to better determine the interrelationship of waxy maize yield and yield components. Path analysis was used to estimate the direct effect of a path model which was proposed based on the biological understanding of the intrarelationship among the endogenous and exogenous variables. Endogenous variables are those variables which are determined within the system (Koopmans, Rubin, and Leipnik 1950). In the study, the endogenous variables were the yield components rows ear\(^{-1}\), ear length, kernels row\(^{-1}\), kernels ear\(^{-1}\), grain yield, ear circumference, and kernel weight. Exogenous variables are those which represent forces outside the confines of the system (Koopmans, Rubin, and Leipnik 1950). In Professor Mason’s study, seeding rate was the exogenous variable. Figure 1.1 represents the path analysis model that was proposed for the study based on the biological evidence (Milander et al. 2015).
Figure 1.1: The proposed causal model for a maize yield component study published in *Maydica* by Milander et al. (2015).

The goodness-of-fit parameters for the proposed model in Figure 1.1 were extremely poor. The data were reanalyzed to find a model that better fit the data. After the data were collected and the path analysis was performed, the significant direct effects were shown as in Figure 1.2.

Figure 1.2: The data-supported causal model for the maize yield component study published in *Maydica* by Milander et al. (2015).

The model in Figure 1.2 was disappointing since it did not give any insight and understanding into the complicated intrarelationship among many of the variables. These results raised the question of whether or not the design was adequate for a causal structure model. This motivated the need to develop search algorithms that would produce an optimal design for a given causal structure and is the subject matter of this research.
The original model was known before the data were collected but the researchers did not take advantage of that information to design an experiment that best fit the proposed model and gave the best estimates. Another important step that could have been taken would have been to conduct a power analysis to determine the minimum number of observations needed to support the proposed model. Even if the researchers would have proceeded to find the optimal design for the causal structure, there is no previous or current literature to support such an objective. Therefore, the work in this dissertation is unique because these are the first applications of optimal design for a causal structure.

Two of the most important aspects of path analysis are estimation of model parameters and model selection. To obtain an accurate inference, we need precise estimates for the parameters. In order to select a model, the model must make sense biologically and fit the data as demonstrated by “good” goodness-of-fit criteria, which could be problematic. There are many models and some could be considered to be “good” based on some goodness-of-fit criteria, but not others. If the model is known and the researcher is not interested in model selection, then it would be in the researchers’ best interests to have the most precise estimates to allow for an accurate inference. Ideally, the researchers would design an experiment that allows for the most precise estimates for the designated parameters, which in turn would lead to more precise inferential statistics.

1.3 Research Objectives

For a given causal structure, the objectives of this research are to obtain an optimal design: (1) For a completely randomized experiment that produces the most precise estimates for the endogenous and exogenous parameters as discussed in Section 3.1, (2) For an experiment with random blocks or split-plots that produces the most precise estimates for the endogenous and exogenous parameters as discussed in Section 3.2, (3) For an experiment with fixed blocks that produces the most precise estimates for the endogenous and exogenous parameters as discussed in Section 3.3, (4) For an experiment with random blocks or split-plots that produces the most
precise estimates for the endogenous and exogenous parameters and the variance components as discussed in Chapter 4, and (5) Using the methods above to demonstrate the improvement in efficiency for two applications published in previous research as discussed in Chapter 5.

1.4 References


   *Technometrics* 16 (2): 211–220.


   and Sons, Inc.

   experiments for fixed effects and variance component estimation. *Technometrics* 56 (2):132–  
   143.


   Observed Polynomial Function and Its Constants and the Guidance They Give Towards a  


    585.

CHAPTER 2: LITERATURE REVIEW

2.1 Optimal Design for Univariate Linear Models

2.1.1 Introduction

A linear model can be defined as \( y = X\beta + \epsilon \), where \( y \) is an \( n \times 1 \) vector of observations, \( X \) is an \( n \times q \) design matrix of rank \( q \), \( \beta \) is a \( q \times 1 \) vector of unknown coefficients that are estimable, and \( \epsilon \) is an \( n \times 1 \) identically independently normally distributed vector with \( E(\epsilon) = 0 \) and \( Var(\epsilon) = \sigma^2 I \).

The least-squares estimate of \( \beta \) is \( \hat{\beta} = (X'X)^{-1}X'y \), which is an unbiased estimator of \( \beta \) and the \( q \times q \) covariance matrix \( Var(\hat{\beta}) = (X'X)^{-1}\sigma^2 \). The most precise estimate is equivalent to minimizing \( Var(\hat{\beta}) \propto (X'X)^{-1} \) (Mitchell 1974a). However, one issue that arises is that \( Var(\hat{\beta}) = (X'X)^{-1}\sigma^2 \) is a matrix which cannot be minimized since a matrix has multiple entries and minimizing some entries may lead to the inflation of other entries. For example, minimizing the diagonal entries may lead to the inflation of the off-diagonal entries. This issue led to the development of optimality criteria.

2.1.2 Optimality Criteria

Optimal designs can be based on different criteria that minimize a function of the variance-covariance matrix of the least-squares estimator (Dette 1997). This section discusses some of the most well-known optimality criteria.

\( D \)-optimality. A \( D \)-optimal design \( X \) minimizes \( |(X'X)^{-1}| \), or \( \min_X |(X'X)^{-1}| \), which is equivalent to \( \max_X |X'X| \). Note that \( |X'X| = \prod_{i=1}^q \frac{1}{\lambda_i} \), where \( \lambda_i \) is the eigenvalue of \( |(X'X)^{-1}| \). Graphically, \( \lambda_i \) is proportional to half of the length of the interval ellipsoid for \( \beta \), which means that \( \max_X |X'X| \) is
equivalent to the minimum volume of the interval ellipsoid (Wald 1943; Kiefer and Wolfowitz 1959).

A-optimality. An A-optimal design matrix $X$ minimizes the average variance of the parameter estimates $\hat{\beta}$ which is equivalent to $\min_X \text{tr}(X'X)^{-1}$ (Chernoff 1953).

$E$-optimality. An $E$-optimal design matrix $X$ is found by choosing the design that minimizes the worst possible variance of any contrast $\min_X \max_\lambda (\lambda_i)$. This criterion can be thought of as the largest possible variance of any contrast based on the least-squares estimator (Ehrenfeld 1955).

Other optimality criteria that are based on the covariance matrix of the estimates were also developed such as $G$-optimality (Kiefer and Wolfowitz 1959) and $I$-optimality (Fedorov 1972). The relationships among other optimality criteria such as $T$-optimality, $L$-optimality, and $C$-optimality are discussed by Rady, Abd El-Monsef, and Seyam (2009). The most widely used criteria are the $D$-optimal and the $A$-optimal criteria (Mitchell 1974b). A common practice is to use $D$-optimality to identify multiple optimal designs and then to rank the designs based on $A$-optimality. If there is a tie among $A$-optimal designs, then a third criterion will be used to break the tie between these competing designs.

All of the previous optimal criteria are based on the information matrix. However, there are other methods that are based on distance criteria (Johnson, Moore, and Ylvisaker 1990). These designs are based on the distance $d(X, A)$ from a point $X$ in the m-dimensional Euclidean space $R^m$ to a set $A \subseteq R^m$ (Rady, Abd El-Monsef, and Seyam 2009). Some examples of distance-based methods, sometimes termed Cloud Methods, are $U$-optimality and $S$-optimality. The disadvantage of distance-based criteria is that they are difficult to optimize computationally.
2.1.3 Search Algorithms for Optimal Design

Once a criterion has been identified, the next challenge is obtaining an optimal design. To demonstrate the challenge, assume that there are five treatments with no interaction that affect response for 20 observations. In this case, the model would be \( y_i = \sum_{j=1}^{5} \beta_j x_{ij} + \epsilon_i \) where \( x_{ij} = 0 \) or 1, \( \epsilon_i \sim N(0, \sigma^2) \), and where \( \beta_j \) are the treatment effects. In this scenario, \( X \) is a \( 20 \times 5 \) design matrix. The objective now is to obtain the combination of the five treatments for each observation that will produce the most precise estimate of the treatment effects on the response. Mathematically, the objective is to select \( X_{20\times5} \), which will maximize \( |X'X| \) or any of the other criteria that were discussed above. In this case, \( X_{20\times5} \) has \( 20 \times 5 = 100 \) entries and each entry has two possibilities, which leads to a total of \( 2^{100} \)

\[ 1,267,650,600,228,229,401,496,703,205,376. \]

An exhaustive search to find an optimal design would compute the \( |X'X| \) for each one of these matrices. However, it is almost impossible because of the large number of possibilities of the design matrix to select a design with the largest \( |X'X| \).

This challenge motivated the development of at least seven algorithms to obtain the optimal design without the need to exhaust all possibilities of the design matrix. The seven algorithms are the Kiefer round-off procedure (Kiefer 1970), the Fedorov algorithm (Fedorov 1972), the Wynn-Mitchell algorithm (Mitchell and Miller 1970; Wynn 1971), the Van Schalkwyk algorithm (Van Schalkwyk 1971), the Mitchell algorithm (Mitchell 1974a), the modified Fedorov algorithm (Cook and Nachtsheim 1980), and the combined Fedorov-Wynn-Mitchell algorithm (Cook and Nachtsheim 1980). A study by Cook and Nachtsheim (1980) compared these seven algorithms and made two recommendations. First, in order to construct efficient designs inexpensively, either the Wynn-Mitchell or Van Schalkwyk algorithms is satisfactory. The second recommendation was to use the modified Fedorov algorithm when highly efficient designs are required.
Based on these recommendations, the modified Fedorov algorithm will be used in this research. However, the modified Fedorov algorithm requires an initial design. The initial design could be random, but the closer the initial design is to the optimal design, the faster the modified Fedorov algorithm will converge and obtain an optimal design. For this reason, the Dykstra Method (Dykstra 1971) is often used, which is not a search algorithm but rather a sequential method.

The Dykstra Method, or Sequential Method (Dykstra 1971), starts with an empty design and will search through a list of given points called candidate points, or design points. Through each step, the algorithm will choose the candidate point that will maximize $|X'X|$. In this method, points are added sequentially until the design with the required runs, or the sample size $n$, is obtained. This method is the fastest, but it is not efficient. Therefore, it may be used to construct the initial design for all other methods.

The Wynn-Mitchell Method, or Simple Exchange Method (Mitchell 1974a), is an algorithm that starts an $n$-point non-singular design constructed using the Dykstra Method by default. Each iteration consists of adding a point and dropping a point from the candidate points. The point that will be added is the point that will produce the maximum increase in $|X'X|$. But the result will be a design with $n + 1$ runs, not $n$. So, we need to drop a point from the $n + 1$ points. The point that will be dropped is the point which results in the least increase in $|X'X|$. The point exchange will continue until no further improvement is achieved or the improvement is sufficiently small.

The DETMAX algorithm starts as a Wynn-Mitchell algorithm in allowing the design to vary from $n$ to $n \pm 1$ (excursion size = 1) (Mitchell 1974b). This process continues until no further improvement in the optimality criterion can be achieved. Then, the DETMAX algorithm will allow a larger and larger excursion, meaning that the algorithm will add or subtract more than one point at a time ($n \pm k$) in order to affect improvement in the optimality criterion. The DETMAX algorithm continues to increase the excursion size until it reaches an upper bound for $k$ that is set
by the researcher. Mitchell recommended setting the upper bound of $k$ to 6 for discrete, finite
design spaces (Mitchell 1974b). However, it was noted by Cook and Nachtsheim (1980) that all
design spaces are convex and a maximum excursion size of 4 was effective for obtaining the
optimal design and a larger excursion did not yield an improved design.

The Modified Fedorov algorithm, or Simultaneous Switching algorithm, starts with an initial
design with the required size (Cook and Nachtsheim 1980). The initial design is constructed by
the Dykstra Method. To improve the design, each iteration and each point in the design will be
compared with each candidate point and the pair of points from the design (Fedorov 1972). The
candidate list that will produce the maximum increase in the optimality criterion will be switched.
The algorithm is expensive because it is necessary to make $nd \times nc$ comparisons in order to make
a single point exchange, where $nd$ is the number of design points and $nc$ is the number of the
candidate points. For that reason, modification to the algorithm was proposed. Cook and
Nachtsheim (1980) suggested to not carry the comparison until the end, but to stop as soon as the
improvement in the optimality criterion exceeds some given number, say $\delta$. However, the
problem with this approach is determining how to select an effective $\delta$. If $\delta$ is too large, then little
will be gained in the reduction of comparisons. But if $\delta$ is too low, then the comparison will be
terminated too early, leading to a very small improvement in the optimality criterion. These
challenges led to a new algorithm that is called the Modified Fedorov algorithm (Cook and
Nachtsheim 1980). The new algorithm will exchange each point in the design by comparing each
point with the candidate points and make a simultaneous exchange with the candidate point that
improves the optimality criterion the most. This means in one iteration, the Modified Fedorov
algorithm will make $nd$ exchanges versus the original Fedorov algorithm which makes one
exchange in each iteration. Since the development of the Fedorov and Modified Fedorov
algorithms in 1972 and 1980, respectively, computing technology has improved which made their
differences less important and also has made Fedorov algorithms more compatible with other algorithms.

One issue that arises with the use of optimality criteria and search algorithms is that the optimal design depends on the model. A change in the model requires a fundamental change in the design matrix to accommodate the new model. In more sophisticated cases, there are more treatments which are related in a more complicated way. For instance, treatments could have interaction among them, making it virtually impossible to guess the treatment combinations for the optimal design. In a causal structure model, there will also be multiple endogenous variables, which add more parameters to explain the relationship among the endogenous variables. The treatments may also have different effects on the endogenous variables. This demonstrates that one weakness of the univariate optimal design is that it considers one endogenous variable at a time. It does not take into account the endogenous parameters and the complicated intrarelationships both among and within the endogenous and exogenous variables.

Structural equations are based on models that are different from those used in the univariate optimal design. They take into account the complicated relationships among the endogenous and exogenous variables simultaneously. However, current literature about optimality criteria and search algorithms only discusses univariate optimal design. There has been no work done in the field of optimal design for causal structure models. This shows the need to develop the methodology and algorithms for structural equation models, which will be done in subsequent chapters of this dissertation.

2.2  Optimal Design for Univariate Mixed Models

2.2.1  Introduction

Historically, optimal design theory for regression models assumes uncorrelated error in most cases. However, this assumption may be invalid in many situations because the experimental runs
are unable to be carried out under homogenous conditions. Goos and Vandbroek (2001) provide several examples to demonstrate the problem of assuming uncorrelated errors. For example, the raw material used in a production process is obtained in batches in which the quality can vary considerably from one batch to another. To account for this variation among the batches, a random batch effect should be added to the regression model. As another example in the semiconductor industry, it is of interest to investigate the effect of several factors on the resistance in computer chips. Here measurements are taken using silicon wafers randomly drawn from a large lot. Therefore, the wafer effect should be considered as a random effect. Yet another example still is a plant science experiment where multiple fields are used. A final example is a chemistry experiment where runs are executed on different days or in different laboratories.

Chasalow (1992) also provides another effective example which demonstrates the problem of assuming uncorrelated error. An experiment in optometry examines the dependence of corneal hydration control on the CO₂ level in a gaseous environment which was applied through goggles covering the eyes of a human subject. Because a response is measured for each eye, one human subject provides a block of two possibly correlated observations by virtue of the subject having a pair of eyes.

2.2.2 Optimal Design for Mixed Models

2.2.2.1 Optimal Design for Random Mixed Models or Split-Plot Models

This section discusses blocked and split-plot experiments since split-plot experiments are special cases of blocked experiments. These experiments include at least two variance components. The objective would be to find treatment combinations in each block that will obtain the optimal parameters estimates with respect to one of the optimality criterion that were previously discussed in Section 2.1.2 using the algorithms from Section 2.1.3.
The model for the data that comes from random blocked or split-plot experiments is \( y = X\beta + Z\mu + \varepsilon \) where \( y \) is an \( n \times 1 \) vector of observations, \( X \) is an \( n \times q \) design matrix of rank \( q \), and \( \beta \) is a \( q \times 1 \) vector of unknown coefficients that are estimable. Additionally, \( q \) is the number of the fixed effects, \( b \) is the number of blocks, \( \mu = (u_1, \cdots, u_b)' \) is a \( b \times 1 \) vector of random effects, \( \varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)' \) is an \( n \times 1 \) vector of residual error, where \((\underline{\varepsilon}, \underline{\mu}) \sim N \left( [0], \begin{bmatrix} \sigma^2_\varepsilon I & 0 \\ 0 & \sigma^2_\mu I \end{bmatrix} \right)\). Assume that an experiment consists of \( n \) experimental runs arranged in \( b \) blocks of sizes \( k_1, \ldots, k_b \) with \( n = \sum_{i=1}^{b} k_i \). For a response surface where \( X \) contains the polynomial expansions of the levels of the \( m \) number of factors at the \( n \) experimental runs, \( Z \) is of the form \( Z = diag[1_{k_1}, \cdots, 1_{k_b}] \) where \( 1_{k_i} \) is a \( k_i \times 1 \) vector of ones. Then, \( V = Var \left( y \right) = \sigma^2_\varepsilon I + \sigma^2_\mu ZZ' \) and \( diag[V_1, \cdots, V_b] = Var \left( y \right) = \sigma^2_\varepsilon I + \sigma^2_\mu ZZ' \), where \( V_i = \sigma^2_\varepsilon \left( I_{k_i \times k_i} + \frac{\sigma^2_\mu}{\sigma^2_\varepsilon k_i} \right) \).

For a mixed model, the goal is to estimate the fixed parameter effects, or \( \beta \). However, in order to be able to test for significance and to construct a confidence interval, the random effect parameters \( \sigma^2_\varepsilon \) and \( \sigma^2_\mu \) need to be estimated. The best linear unbiased estimator for the fixed parameter effects is the generalized least squares (GLS) estimator \( \hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y \) where \( Var \left( \hat{\beta} \right) = (X'V^{-1}X)^{-1} \).

The lack of theoretical results on the subject of blocked and split-plot experiments in literature led to the popularity of search algorithms (Cheng 1995; Atkins and Cheng 1999). Following the footsteps of the process that was developed for optimal design in the case of completely randomized experiments, the objective becomes to choose \( X \) that will minimize the \( |Var \left( \hat{\beta} \right)| = |(X'V^{-1}X)^{-1}| \), which is equivalent to selecting the \( X \) that maximizes \( |(X'V^{-1}X)| \). It is not necessary to use \( D \)-optimality. Any of the other optimality criteria that were discussed in Section
2.1.2 could have been used to select the $X$ that will maximize the information matrix. But since $D$-optimality is the most popular criterion (Mitchell 1974b), it is a good place to start.

However, the common problem for all optimality criteria is that $V$ is unknown. That fact presents a new challenge and shows that the more complicated the model, the more challenges there are that can arise. The first solution for the new challenge was to assume that $V = Var(y) = \sigma^2 + \sigma^2 ZZ'$ is known by making assumptions about $\sigma^2$ and $\sigma^2$. The second solution was to rewrite

$$V = Var(y) = \sigma^2 + \sigma^2 ZZ' = \sigma^2 (I + \eta ZZ') = \sigma^2 \left(I + \frac{\rho}{1-\rho} ZZ'\right),$$

where $\eta = \frac{\sigma^2}{\sigma^2}$ and $\rho = \frac{\sigma^2}{\sigma^2 + \sigma^2}$. (Goos and Vandebroek 2001; Goos and Vandebroek 2003; Goos and Jones 2007). This allows us to rewrite $|(X'V^{-1}X)^{-1}| = \sigma^{-2q} |X'(I + \eta ZZ')^{-1}X|$ as in the completely randomized experiment. It can be assumed without loss of generality $\sigma^2 = 1$, thus $|(X'V^{-1}X)^{-1}|$ can be expressed in terms of variance ratio $\eta$ or in terms of the correlation coefficient $\rho$. Therefore, the $D$-optimal design for a blocked or a split-plot experiment depends on the variance ratio. However, Goos and Vandebroek (2001; 2003) and Goos and Jones (2011) argued that the dependence is minor and presented algorithms for constructing locally optimal designs for a given $\eta$ or $\rho$. In some special cases, the optimal design is globally optimal and neither depends on the variance ratio nor the degree of correlation of which the most popular is the orthogonal design (Goos and Vandebroek 2001).

The previous method which is based on optimization of $|(X'V^{-1}X)|$ is referred to in literature as the traditional method. This method has two weaknesses (Mylona, Goos, and Jones 2014). The first weakness is that the $D$-optimal design depends on $\eta$ or $\rho$ and these values are not known prior to the experiment. The second weakness is that the optimal design focuses only on the fixed effect parameters and ignores the estimate of the random effect parameters. These weaknesses could have severe consequences on the estimate of the variance components, which severely affect the inferential statistics on the fixed effect parameters. Mylona, Goos, and Jones (2014)
provided scenarios for optimal designs that were constructed using the traditional $D$-optimality criterion and in these designs, the random effects $\sigma^2_u$ and $\sigma^2_\varepsilon$ are not estimable.

Consequently, the two weaknesses led to a new composite $D$-optimality criterion. The reason for calling it composite is because this criterion takes into consideration both the fixed effect and the random effect parameters. It is important to not fall into the practice of focusing on some parameters and leaving others. Regarding causal models, the same argument can be made for describing the relationship among the endogenous parameters in the optimality criterion since they are also parameters of interest.

Mylona, Goos, and Jones (2014) proposed a new composite design criterion that includes both information matrices for the fixed and the random effects. For the covariance information matrix, they considered the residual maximum likelihood (REML) estimates and not the maximum likelihood (ML) estimates since the ML information matrix for the variance components does not depend on the design matrix and the REML information matrix does. The new composite criterion is given by $
abla = \frac{\alpha}{q} \log |X'V^{-1}X| + \frac{1-\alpha}{2} \log |N|$, where $\alpha$ is a real number between 0 and 1, and is a weight attached for the fixed effects estimation, $1-\alpha$ represents a weight for the variance component estimation, and $N = \begin{bmatrix} Var(\hat{\sigma}_u^2) & Cov(\hat{\sigma}_u^2, \hat{\sigma}_\varepsilon^2) \\ Cov(\hat{\sigma}_u^2, \hat{\sigma}_\varepsilon^2) & Var(\hat{\sigma}_\varepsilon^2) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} tr(N_{11}) & tr(N_{12}) \\ tr(N_{21}) & tr(N_{22}) \end{bmatrix}$.

Furthermore, where $N_{11} = \{(V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}X)^{-1}X'V^{-1}X\}^2$, $N_{12} = N_{21} = \{(V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1})^2\}$, $N_{22} = \{(V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1})^2\}$.

The new criterion still requires an input on the relative magnitude of the variance components, $\eta$ or $\rho$, which is unknown prior to the experiment. So, to account for the uncertainty, Mylona, Goos, and Jones (2014) used a Bayesian approach that was introduced by Chaloner and Larntz (1989). They referred to this criterion as Bayesian Composite $D$-optimality Criterion. In it, they used three approaches with the criteria. The first approach is a log-normal prior distribution for $\eta$
where \( D_B = \int_0^{\infty} \Phi(\eta) \ast \frac{1}{\eta \sigma \sqrt{2\pi}} e^{-(\log(\eta) - \mu)^2/(2\sigma^2)} \, d\eta \). The second approach uses a beta prior distribution for \( \rho \) where \( D_B = \int_0^1 \Phi(\rho) \ast \frac{\Gamma(k+\lambda)}{\Gamma(k)\Gamma(\lambda)} \rho^{k-1}(1-\rho)^{\lambda-1} \, d\rho \). The third approach uses two different prior distributions for \( \sigma_u^2 \) and \( \sigma_e^2 \) where \( \int_0^{\infty} \int_0^{\infty} \Phi(\sigma_u^2, \sigma_e^2) \ast p(\sigma_u^2) \ast q(\sigma_e^2) \, d\sigma_u^2 \, d\sigma_e^2 \).

The third approach is the least preferred option since it is necessary to have two prior distributions and to evaluate double integrals. For these reasons, Mylona, Goos, and Jones (2014) focused on the first and second approaches, and had three main conclusions. First, the Bayesian Composite \( D \)-optimality Criterion produced more efficient designs that resulted in fewer zero and large estimates for the variance components, especially in the case where the number of parameters is close to the number of runs. The second conclusion was that in terms of the effect of the prior distribution(s), new optimal designs are not highly sensitive to the prior distribution assumed. This means that the most important thing is to include the information matrix of the variance components and not to focus on the exact prior distribution. Finally, they found \( \alpha \) between 0.5 and 0.75 led to satisfactory results.

A special case of the mixed model is the split-plot model. Split-plot experiments are blocked experiments where the blocks themselves serve as experimental units for a subset of the factors (Fisher 1925; Yates 1936). Therefore, a split-plot design has two levels of experimental units where the blocks are called the whole plots while the experimental units within the blocks are labelled as split-plots.

In recent years, split-plot optimal designs have received special attention in literature and many advances have been made (Macharia and Goos 2010). The importance of split-plot optimal designs stems from their applications in industry (Box, Hunter, & Hunter 2005). Industrial designs are complex and usually include factors that are hard to change, are costly, and/or are
time-consuming to adjust or readjust (Jones and Goos 2007). Jones and Goos (2012b) detail many specific examples and applications in industrial experimentation with split-plot designs.

Goos (2006) and Jones and Nachtsheim (2009) point out that there are several approaches to obtain a split-plot response surface optimal design. The first approach, which was developed by Bingham and Sitter (1999, 2001, 2003), Loeppky and Sitter (2002), and Bingham, Schoen, and Sitter (2004) uses a minimum aberration for split-plot fractional factorial design. The second approach obtains equivalence of ordinary least squares (OLS) and generalized least squares (GLS) estimators (Kowalski and Montgomery 2005). The optimal design that was produced through this approach is called the equivalent-estimate optimal design. The third approach uses the variance of the GLS estimates (Goos 2002; Goos and Vandebroek 2001; Goos and Vandebroek 2003; Goos and Vandebroek 2004; Jones and Goos 2007; Jones and Goos 2009; Jones and Nachtsheim 2009).

Macharia and Goos (2010) discussed the second and third approaches and compared the $D$-efficiency of the design. They discussed the advantages and disadvantages of each approach and introduced an algorithm that produced an optimal design that is both $D$-optimal and equivalent-estimate optimal simultaneously. A second algorithm was introduced by Jones and Goos (2012b), which produced equivalent-estimation designs for scenarios where Macharia and Goos failed. They also claimed that their new algorithm produced equivalent-estimation designs that outperformed those of Macharia and Goos.

Jones and Goos (2012a) compared $I$-optimality and $D$-optimality for split-plot response surface designs and argued in favor of using $I$-optimality criteria for split-plot response surface designs since the goal is most likely prediction. $I$-optimality design minimizes the average prediction variance over the design region. They also concluded that the performance of the $D$-optimality
split-plot designs in terms of the $I$-optimality criterion perform substantially worse as compared to $I$-optimal design in terms of the $D$-optimality criterion.

### 2.2.2.2 Optimal Design for Mixed Models with Fixed Blocks

The first algorithm that allows for blocking was developed by Harville (1974). However, in that era the computing capacity was limited and was one of the challenges that motivated the development of other algorithms that were more efficient. Examples of alternative algorithms were proposed by Atkinson and Donev (1989) and Cook and Nachtsheim (1989) where both algorithms produced a $D$-optimal regression design in the presence of fixed blocks. Trinca and Gilmour (2000) also proposed an exchange algorithm for a prespecified treatment combination of candidate points to a number of fixed blocks. While they all have the same objective function, the difference between these algorithms is the exchange method which affects the rate of convergence.

For the methods described above, the objective function is derived by assuming the blocks are fixed instead of random. The model would be $y = X\beta + Zu + \varepsilon$ where $\varepsilon \sim N(0, \sigma^2 I)$. In this scenario, the number of observations $n$ must be greater than or equal to $q + b$ the number of fixed effects. Meanwhile, in the random mixed model, $n$ must be greater than or equal to $q$. There is no interest in the estimate of the block effects in the fixed model and these parameters are considered nuisance parameters. For that reason, the objective is to maximize the precision for $\hat{\beta}$. It is necessary to rewrite the model and partition the information matrix to focus on the precision of $\hat{\beta}$ alone. To achieve this objective, we start with the original model $y = X\beta + Zu + \varepsilon = \left[\begin{array}{c} X \\ Z \end{array}\right]\left[\begin{array}{c} \beta' \\ u' \end{array}\right]' + \varepsilon$. This can be simplified as $y = F\xi + \varepsilon$. The full information matrix would
be \( F'F = [X \ Z]'[X \ Z] = \begin{bmatrix} M_{11} & M_{12} \\ X'X & X'Z \\ Z'X & Z'Z \\ M_{21} & M_{22} \end{bmatrix} \). Cook and Nachtsheim (1989) argued that the \( D \)-optimal design for \( \beta \) would be the design \( X \) such that 
\[
\max_X |M_{11} - M_{12}M_{22}^{-1}M_{21}| = \max_X |X'X - X'Z(Z'Z)^{-1}Z'X|.
\]
This objective function was used to obtain the optimal designs and also to explain the relationship between the optimal designs for mixed models with fixed blocks and the optimal designs for random mixed models.

Through the discussion from the previous sections, it can be seen that changing the types of linear models led to the development of new algorithms and new objective functions. The search algorithms and objective function for a completely randomized optimal design was different than the search algorithms and objective function for a blocked optimal design. That is to say that the optimal design depends on the model.

In order to demonstrate the dependency of optimal design on the model, consider the following two simple examples. In the first example, we will construct a \( D \)-optimal design for 20 observations for the model \( y_i = \mu + \beta_1 x_i + \varepsilon_i \), where \( x_i \in \{-1, 0, 1\} \), \( i = 1, \ldots, 20 \) and \( \varepsilon_i \sim N(0, \sigma^2) \). The \( D \)-optimal design matrix is

\[
X = \begin{bmatrix} \mu & x_i \\ 1_{10 \times 1} & 1_{10 \times 1} \\ 1_{10 \times 1} & -1_{10 \times 1} \end{bmatrix}.
\]

In a completely randomized optimal design to estimate the model parameters \( \mu \) and \( \beta_1 \), the treatment will be applied evenly at the level \(-1\) and \(1\). That is to say that the treatment level \(-1\) is applied on 10 experimental units and the treatment level 1 is applied on 10 experimental units in order to obtain the most precise estimates for the treatment effect and the intercept.
In the second example, we will construct a $D$-optimal design for 20 observations for the model 
\[ y_i = \mu + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i, \]
where \( x_i \in \{-1, 0, 1\}, \ i = 1, \cdots, 20 \) and \( \varepsilon_i \sim N(0, \ \sigma^2) \). The $D$-
optimal design matrix is
\[
X = \begin{bmatrix}
\mu & x_i & x_i^2 \\
1_{7 \times 1} & 1_{7 \times 1} & 1_{7 \times 1} \\
1_{6 \times 1} & 0_{6 \times 1} & 0_{6 \times 1} \\
1_{7 \times 1} & -1_{7 \times 1} & 1_{7 \times 1}
\end{bmatrix}
\]

In a completely randomized experiment to estimate the model parameters $\mu$, $\beta_1$, and $\beta_2$, the
treatment must be applied to all three levels. The treatment level 1 is applied on 7 experimental
units, the treatment level $-1$ is applied on 7 experimental units, and the treatment level 0 is
applied on 6 experimental units in order to obtain the most precise estimates for the treatment
effect and the intercept.

The previous two examples demonstrate how much the optimal design depends on the model. In
the first example, the optimal design for the model was achieved by applying the treatment evenly
at the levels $-1$ and $1$. In contrast, the optimal design for the second model in which there is a
quadratic term, the optimal design was achieved by applying the treatment at all three levels.
Adding the quadratic term changed the optimal design and this shows that a change in the model
required a fundamental change in the design matrix to accommodate the new model. In the same
way that univariate optimal design for linear models necessitated the development of search
algorithms and objective functions, the same need exists for structural equations, as well.

2.3 Structural Equations

2.3.1 Introduction

A univariate linear model is defined by $y = X\beta + \varepsilon$. However, most often there is more than one
dependent variable involved in a natural system. The linear model can be extended to a
multivariate model where there are \( p \) dependent variables, with \( p > 1 \). Assume that there are \( q \) independent variables. A multivariate linear model is defined by \( Y = X\Gamma + E \), where \( Y = \begin{bmatrix} y_1' \\ \vdots \\ y_n' \end{bmatrix} \) is an \( n \times p \) matrix that includes responses of the \( p \) dependent variables and \( y_k \) is a \( p \times 1 \) vector of responses for the \( k^{th} \) observation. \( X = \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix} \) is an \( n \times q \) design matrix that includes \( q \) independent variables and \( x_k \) is a \( q \times 1 \) vector of predetermined values. \( \Gamma \) is a \( q \times p \) matrix of unknown coefficients that are estimable. \( E = \begin{bmatrix} \varepsilon_1' \\ \vdots \\ \varepsilon_n' \end{bmatrix} \) is an \( n \times p \) matrix where \( \varepsilon_k \) is a \( p \times 1 \) vector of random residuals such that \( \varepsilon_k \sim N(0, \Sigma) \) where \( \varepsilon_j \) and \( \varepsilon_j \) are independent, and \( \Sigma \) is a \( p \times p \) nonnegative definite matrix. The multivariate linear model has advantages over the univariate model by allowing more realistic modeling for the system because it takes into account more than one dependent variable at a time. It also gives additional insight over the univariate model since it allows one to test hypotheses across variables, which is not possible to be accomplished through the univariate approach.

But the multivariate linear model assumes that the relationship among the \( p \) dependent variables is too simplistic. For example, the \( k^{th} \) observation vector from a multivariate model can be expressed by \( y_k = \Gamma' x_k + \varepsilon_k \), where \( Var(\varepsilon_k) = \Sigma \) describes the association among the \( p \) dependent variables. However, in most natural systems, the dependent variables are related in a much more complicated way where one dependent variable could directly or indirectly influence another dependent variable. Multivariate modeling does not address the direct intrarelationships among dependent variables. Structural equation modeling is considered as a natural extension and generalization for multivariate linear modeling.
Structural equation modeling (SEM), also called causal structure modeling, allows qualitative cause-effect information to be combined with statistical data (Jöreskog 1970; Bollen 1989). This combination provides quantitative assessment of the cause-effect relationships among and within the endogenous and exogenous variables (Bollen 1989). SEMs are widely-used and are particularly helpful in social and behavioral science, as well as in economics, agriculture, epidemiology and many other areas (Bollen, 1989). SEMs were originally referred to as “path analysis” or “causal structures” until Jöreskog (1970), who generalized these models into one overall framework, termed them “structural equation models” (Jöreskog 1970; Haavelmo 1943; Koopmans 1953; Koopmans, Rubin, and Leipnik 1950; Wright 1921).

Similar to linear models in the univariate case, we will need to estimate the endogenous and exogenous parameters for a causal structure. However, in order to obtain the estimates, we must first discuss the estimation methods for a causal structure. The concept of identification for a causal structure must also be discussed, as well.

2.3.2 The Reduced Form of a Causal Structure

The variables in the SEMs are classified into two main types, endogenous variables and exogenous variables, which can be denoted by \( \{y_1, y_2\} \) and \( \{x_1, x_2\} \), respectively. The endogenous variables are those variables which are determined within the system and are the dependent variables in a multivariate linear model (Koopmans, Rubin, and Leipnik 1950). Examples of endogenous variables in an agricultural experiment involving maize would include yield, ears per meter squared, rows per ear, kernels per ear, and kernel weight (Milander et al. 2016). In a heart study from medicine, endogenous variables might include hypertension, coronary heart disease, and blood glucose (Mi et al. 2011).

The exogenous variables are those which represent forces outside the confines of the system and are the independent variables in a multivariate linear model (Koopmans, Rubin, and Leipnik
1950). Examples of exogenous variables in an agricultural experiment include fertilizers, hybrid types, planting depth, irrigation, and seeding rate (Milander et al. 2016). In a medical experiment, examples of exogenous variables would include age, body mass index, gender, and cigarettes per day (Mi et al. 2011). In the context of optimal design, the exogenous variables will be represented by the design matrix \( X \).

A causal structure can be written as a set of simultaneous equations, with each equation in the form of

\[
\sum_{j=1}^{p} y_{kj} b_{ji} + \sum_{j=1}^{q} x_{kj} y_{ji} + \epsilon_{ki} = 0 \quad \text{where } i = 1, 2, \ldots, p \text{ and } k = 1, 2, \ldots, n \text{ and where } p \text{ is the number of endogenous variables, } q \text{ is the number of exogenous variables, and } n \text{ is the number of observations (Schmidt 1976). Additionally, } y_{k1}, \ldots, y_{kp} \text{ represent the endogenous variables or the responses for the } k^{th} \text{ observation, } b_{1i}, \ldots, b_{pi} \text{ are the endogenous parameters of the } i^{th} \text{ equation, } x_{k1}, \ldots, x_{kq} \text{ represent the exogenous variables or the predetermined factors for the } k^{th} \text{ observation, } y_{1i}, \ldots, y_{qi} \text{ are the endogenous parameters of the } i^{th} \text{ equation, and } \epsilon_{ki} \text{ is the disturbance of the } i^{th} \text{ equation for the } k^{th} \text{ observation. Written in matrix form, the model is } YB + X\Gamma + E = 0 \text{ where } Y \text{ is an } n \times p \text{ matrix that includes responses of the } p \text{ endogenous variables, } B \text{ is a } p \times p \text{ matrix of unknown coefficients, } X \text{ is an } n \times q \text{ design matrix, } \Gamma \text{ is a } q \times p \text{ matrix of unknown coefficients, } E = \begin{bmatrix} \epsilon_1' \\ \vdots \\ \epsilon_n' \end{bmatrix} \text{ where } \epsilon_i \text{ is a } p \times 1 \text{ vector of random residuals such that } \epsilon_i \sim N(0, \Sigma) \text{ where } \epsilon_i \text{ and } \epsilon_j \text{ are independent, and } \Sigma \text{ is a } p \times p \text{ nonnegative definite matrix. The reduced form can be written by solving for } Y, \text{ assuming that } B^{-1} \text{ exists. So, the reduced form is a multivariate linear model } Y = X\Pi + H \text{ where } \Pi = -\Gamma B^{-1} \text{ and } H = -EB^{-1} \text{ (Schmidt 1976). We will demonstrate how to use the reduced form for a causal structure through three simple systems, which will also serve to introduce the concept of identification.}
System 1 in Figure 2.1 has two exogenous variables and two endogenous variables. Each exogenous variable affects an endogenous parameter independently and is denoted by arrows. The first endogenous variable affects the second endogenous variable as shown in Fig. 2.1.

**Figure 2.1:** Causal structure where each exogenous variable directly affects only one endogenous variable.

Throughout, assume that all of the variables have been centered and there is no interest in the intercepts. The $k^{th}$ observation can then be expressed mathematically as $y_{k1} = y_{11}x_{k1} + \varepsilon_{k1}$ and $y_{k2} = b_{12}y_{k1} + y_{22}x_{k2} + \varepsilon_{k2}$ where $[\varepsilon_{k1} \varepsilon_{k2}] \sim N(0, \Sigma)$ for $k = 1 \ldots n$. These can be rewritten in a more familiar form of

$$
\begin{bmatrix}
\gamma_{11}' \\
\gamma_{22}'
\end{bmatrix}
\begin{bmatrix}
x_{k1}' \\
x_{k2}'
\end{bmatrix}
+ 
\begin{bmatrix}
\varepsilon_{k1}' \\
\varepsilon_{k2}'
\end{bmatrix}
= 
\begin{bmatrix}
y_{k1}' \\
y_{k2}'
\end{bmatrix}

$$

By transposing both sides of the previous equation, then $y_{k}'(\Delta - I)' + x_{k}'\Gamma + \varepsilon_{k}' = 0$. Let $B = (\Delta - I)'$ and assume that $\varepsilon_{k}' \sim N(0, \Sigma)$, then $y_{k}'B + x_{k}'\Gamma + \varepsilon_{k}' = 0$. This equation is for one observation. Now stack all of the observations in a matrix

$$
\begin{bmatrix}
y_{1}' \\
\vdots \\
y_{n}'
\end{bmatrix}
B + 
\begin{bmatrix}
x_{1}' \\
\vdots \\
x_{n}'
\end{bmatrix}
\Gamma + 
\begin{bmatrix}
\varepsilon_{1}' \\
\vdots \\
\varepsilon_{n}'
\end{bmatrix}
= 0
$$

Then, $YB + X\Gamma + E = 0$ which is called the general structural model where all of the terms are defined previously in Section 2.3.2. In the example above for Figure 2.1, $B = \begin{bmatrix}
-1 & b_{12} \\
0 & -1
\end{bmatrix}$ and $\Gamma = \begin{bmatrix}
y_{11} & 0 \\
0 & y_{22}
\end{bmatrix}$. Another way to express the general structural model is by post-multiplying both sides by $B^{-1}$, making $Y = X(B^{-1}\Gamma + (-EB)^{-1})$. Therefore,
\[ Y = X\Pi + H \quad \text{where} \quad H = \begin{bmatrix} h_1' \\ \vdots \\ h_n' \end{bmatrix}, \quad \text{and} \quad h_k' \quad \text{are identically independently distributed as normal with mean zero and covariance matrix} \ B^{-1}\Sigma[B^{-1}]' \quad \text{which, again, is the reduced form (Schmidt 1976).} \]

System 2 in Figure 2.2 has two exogenous variables and two endogenous variables. The first exogenous variable affects only one endogenous variable. The second exogenous variable affects both endogenous variables. The first endogenous variable affects the second endogenous variable as shown in Fig. 2.2.

**Figure 2.2:** Causal structure where the second exogenous variable directly affects both endogenous variables.

The \( k \)th observation can then be expressed mathematically as \( y_{k1} = \gamma_{11}x_{k1} + \gamma_{21}x_{k2} + \varepsilon_{k1} \) and \( y_{k2} = b_{12}y_{k1} + \gamma_{22}x_{k2} + \varepsilon_{k2} \). Then, the structural form for System 2 is \( YB + X\Gamma + E = 0 \),

where \( B = \begin{bmatrix} -1 & b_{12} \\ 0 & -1 \end{bmatrix} \) and \( \Gamma = \begin{bmatrix} \gamma_{11} & 0 \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \).

System 3 in Figure 2.3 has two exogenous variables and two endogenous variables. The first exogenous variable affects two endogenous variables. The second exogenous variable affects only one endogenous variable. The first endogenous variable affects the second endogenous variable as shown in Fig. 2.3:
Figure 2.3: Causal structure where the first exogenous variable affects both endogenous variables.

The $k^{th}$ observation can then be expressed mathematically as $y_{k1} = y_{11}x_{k1} + \varepsilon_{k1}$ and $y_{k2} = b_{12}y_{k1} + y_{22}x_{k2} + y_{12}x_{k1} + \varepsilon_{k2}$. Then, the reduced form for System 3 is $YB + X\Gamma + E = 0$, where $B = \begin{bmatrix} -1 & b_{12} \\ 0 & -1 \end{bmatrix}$ and $\Gamma = \begin{bmatrix} y_{11} & y_{12} \\ 0 & y_{22} \end{bmatrix}$.

In all three systems, the reduced form $Y = X\Pi + H$ is useful because the ordinary least squares (OLS) estimator $\hat{\Pi}$ can be used to obtain estimates for the endogenous and exogenous parameters $B$ and $\Gamma$, respectively.

Our objective for a structural equation model is to estimate the endogenous parameters, the exogenous parameters, and the covariance components. For the model in the first system, the goal would be to estimate $B$, $\Gamma$, and $\Sigma$ while maintaining the constraints on $B$ and $\Gamma$. For example, $b_{ii} = -1$ in all three of the previous examples where $b_{ii}$ are the diagonal entries of $B$.

2.3.3 Identification of the Parameters of a Causal Structure

The first estimation technique uses the reduced form $Y = X\Pi + H$, which looks like a multivariate linear model. The OLS estimate $\hat{\Pi} = (X'X)^{-1}X'Y$, but the goal here is to estimate $B$ and $\Gamma$, and $\hat{\Pi}$ is only useful in that regard. Therefore, the next step would be to use $\hat{\Pi} = -\hat{\Gamma}B^{-1}$ to estimate $B$ and $\Gamma$, which is not always possible since often there are more parameters than equations. This leads to the concept of identification. A parameter of a model is identified if and only if it can be estimated from the reduced system or $\hat{\Pi}$ (Schmidt 1976). To demonstrate the
concept of identification, the three systems in Section 2.3.2 will be used to discuss the identification of the parameters of each system.

First, it will be shown that the three parameters $\gamma_{11}$, $\gamma_{22}$, and $b_{12}$ in System 1 are identifiable but that the system is overidentified. Since $\hat{\Pi} = -\hat{\Gamma}\hat{B}^{-1}$, then $\begin{bmatrix} \hat{n}_{11} & \hat{n}_{12} \\ \hat{n}_{21} & \hat{n}_{22} \end{bmatrix} = \left(\hat{X}'\hat{X}\right)^{-1}\hat{X}'\hat{Y}$

$$-\begin{bmatrix} \hat{\gamma}_{11} & 0 \\ 0 & \hat{\gamma}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \hat{\gamma}_{11} & 0 \\ 0 & \hat{\gamma}_{22} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1}.$$ Now multiply the right-hand side of the equation to obtain $\begin{bmatrix} \hat{n}_{11} & \hat{n}_{12} \\ \hat{n}_{21} & \hat{n}_{22} \end{bmatrix} = \begin{bmatrix} \hat{\gamma}_{11} & \hat{\gamma}_{11} \hat{b}_{12} \\ \hat{\gamma}_{21} & \hat{\gamma}_{21} \hat{b}_{12} + \hat{\gamma}_{22} \end{bmatrix}$. So, $\hat{\gamma}_{11} = \hat{n}_{11}$, $\hat{\gamma}_{22} = \hat{n}_{22}$, and $\hat{b}_{12} = \hat{n}_{12}$. The three parameters $\gamma_{11}$, $\gamma_{22}$, and $b_{12}$ of the reduced form of System 1 are estimable. However, the first element in the second row on the left-hand side is a zero and does not require an estimate. Such a parameter that has two different estimates is called overidentified (Schmidt 1976). This precipitated the need for another estimating method that will address the overidentification challenge.

Next, it will be shown that the four parameters $\gamma_{11}$, $\gamma_{21}$, $\gamma_{22}$, and $b_{12}$ in System 2 are identifiable but that the system is exactly identified. Since $\hat{\Pi} = -\hat{\Gamma}\hat{B}^{-1}$, then $\begin{bmatrix} \hat{n}_{11} & \hat{n}_{12} \\ \hat{n}_{21} & \hat{n}_{22} \end{bmatrix} = \left(\hat{X}'\hat{X}\right)^{-1}\hat{X}'\hat{Y}$

$$-\begin{bmatrix} \hat{\gamma}_{11} & 0 \\ \hat{\gamma}_{21} & \hat{\gamma}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \hat{\gamma}_{11} & \hat{\gamma}_{11} \hat{b}_{12} \\ \hat{\gamma}_{21} & \hat{\gamma}_{21} \hat{b}_{12} + \hat{\gamma}_{22} \end{bmatrix}$. So, $\hat{\gamma}_{11} = \hat{n}_{11}$, $\hat{\gamma}_{21} = \hat{n}_{21}$, $\hat{b}_{12} = \hat{n}_{12}$, and $\hat{\gamma}_{22} = \hat{n}_{22} - \frac{\hat{n}_{21} \hat{n}_{12}}{\hat{n}_{11}}$. The four parameters $\gamma_{11}$, $\gamma_{21}$, $\gamma_{22}$, and $b_{12}$ of the reduced form of System 2 are estimable and each parameter has a unique estimate. Therefore, this system is called exactly identified (Schmidt 1976).
Finally, for System 3, not all parameters are identified. Since \( \hat{\Pi} = -\hat{\Gamma} \hat{B}^{-1} \), then

\[
\begin{bmatrix}
\hat{\Pi}_{11} & \hat{\Pi}_{12} \\
\hat{\Pi}_{21} & \hat{\Pi}_{22}
\end{bmatrix}
= \left( X'X \right)^{-1} X'Y
\]

\[- \begin{bmatrix}
\hat{y}_{11} & \hat{y}_{12} \\
0 & \hat{y}_{22}
\end{bmatrix}
\begin{bmatrix}
-1 & \hat{b}_{12} \\
0 & -1
\end{bmatrix}
= \begin{bmatrix}
\hat{y}_{11} & (\hat{y}_{11} \hat{b}_{12} + \hat{y}_{12}) \\
0 & \hat{y}_{22}
\end{bmatrix} \]

So, \( \hat{\Pi}_{11} = \hat{\Pi}_{11}, \hat{y}_{22} = \hat{\Pi}_{22}, \) and \( \hat{\Pi}_{12} = \hat{y}_{11} \hat{b}_{12} + \hat{y}_{12} \). The last equation has two unknowns, which are \( b_{12} \) and \( y_{12} \). Thus, we cannot estimate uniquely \( b_{12} \) and \( y_{12} \). Thus, the system is called underidentified because not all of the parameters can be estimated from the reduced form (Schmidt 1976).

### 2.3.4 Causal Structure Estimating Methods

In the case of overidentification where a parameter has more than one estimate, there is a need for new estimators that would not have that problem. This new class of estimators includes Indirect Least-Squares (ILS), Two-Stage Least-Squares (2SLS) (Theil 1953; Basmann 1957) and Limited Information Maximum Likelihood (LIML) (Anderson and Rubin 1949). Both 2SLS and LIML address the overidentification challenge. However, these estimators would not be much help in terms of optimal design since they estimate the parameter of one equation at a time. The goal here is to obtain the optimal design for the entire system, not to obtain different optimal designs for different equations in the system.

The last class of estimators addresses the overidentification issue and estimates the entire system of parameters simultaneously, which increases the efficiency of the estimation since it takes into account the correlation among the equations of the system. There are two estimators that achieve those objectives and they are Three-Stage Least-Squares (3SLS) (Zellner and Theil 1962) and Full Information Maximum Likelihood (FIML) (Anderson and Rubin 1949).

Some definitions are necessary for the development of the Three-Stage Least Squares (3SLS) estimator. \( Y_{(j)} \) is an \( n \times (p_j - 1) \) matrix that consists of the endogenous variables that are in the \( j^{th} \) equation. \( y_{(j)} \) is an \( n \times 1 \) vector that consists of the responses for the \( j^{th} \) endogenous variables.
variable, which is to say that \( Y_{(j)} \) is the \( j \)th column of \( Y \) where \( Y \) was previously defined in Section 2.3.1. \( X_{(j)} \) is an \( n \times q_j \) matrix that consists of the exogenous variables that are in the \( j \)th equation, \( b_{(j)} \) are the endogenous parameters of the \( j \)th equation, and \( \gamma_{(j)} \) are the exogenous parameters of the \( j \)th equation. \( \varepsilon_{(j)} \) is an \( n \times 1 \) vector of random error for the \( j \)th endogenous variable which is the \( j \)th column of \( E \) where \( E \) was defined previously in Section 2.3.1. Each equation in the system can be rewritten using vector notations. So, \( Y_{(j)} = Y_{(j)}b_{(j)} + X_{(j)\gamma_{(j)}} + \varepsilon_{(j)} \) and \( \varepsilon_{(j)} \sim N \left( 0, \sigma^2_{(j)} I \right) \). Now rewrite the \( j \)th equation as \( Y_{(j)} = [Y_{(j)} \quad X_{(j)}] \left[ b_{(j)} \quad \gamma_{(j)} \right]' + \varepsilon_{(j)} \). Therefore, \( Y_{(j)} = W_{(j)}\delta_{(j)} + \varepsilon_{(j)} \). As in the univariate case, multiply both sides by \( X' \) to obtain the normal equations. So, \( X'Y_{(j)} = \)

\[
X'W_{(j)}\delta_{(j)} + X'\varepsilon_{(j)}.
\]

Now stack the previous equations in a matrix to obtain

\[
\begin{bmatrix}
X'Y_{(1)} \\
\vdots \\
X'Y_{(p)}
\end{bmatrix} = \begin{bmatrix}
X'W_{(1)} & 0 & 0 & \delta_{(1)} \\
0 & \ddots & 0 & \vdots \\
0 & 0 & X'W_{(p)} & \delta_{(p)}
\end{bmatrix} + \begin{bmatrix}
X'\varepsilon_{(1)} \\
\vdots \\
X'\varepsilon_{(p)}
\end{bmatrix}.
\]

Using the Kronecker product, \((I \otimes X)\gamma^* = (I \otimes X)W^* \delta^* + (I \otimes X)\varepsilon^* \) where \( \gamma^* = [Y_{(1)'}, \ldots, Y_{(p)'}]' \), \( W^* = \begin{bmatrix}
W_{(1)} & 0 & 0 \\
0 & \ddots & 0 \\
0 & 0 & W_{(p)}
\end{bmatrix} \), \( \delta^* = \begin{bmatrix}
\delta_{(1)}' \ldots, \delta_{(p)}'
\end{bmatrix}' \) and \( \varepsilon^* = [\varepsilon_{(1)'}, \ldots, \varepsilon_{(p)'}]' \). Then, to estimate \( \delta^* \), use the Generalized Least Square (GLS) to obtain \( \delta^*_{3SLS} = [W^* (\Sigma^{-1} \otimes X[X'X]^{-1} X') W^*]^{-1} W^* [\Sigma^{-1} \otimes X[X'X]^{-1} X']^{-1} \gamma^* \). Besides the estimates, the standard error is needed in order to compute confidence intervals and to test hypotheses. So, the covariance matrix of the 3SLS estimators is \( Var(\hat{\delta}^*_{3SLS}) = [W^* (\Sigma^{-1} \otimes X[X'X]^{-1} X') W^*]^{-1} + O \left( \frac{1}{n} \right) \) where \( \hat{W}^* = E(W) \) (Zellner and Theil 1962; Sargan 1964). The derivation of the covariance matrix estimate of the 3SLS estimators can be found in
Appendix 2A. Thus, the estimate of the covariance matrix would be $\hat{\text{Var}}(\hat{\delta}_{3SLS}^*) = \left[\hat{W}^* (\Sigma^{-1} \otimes X'X)^{-1} X' \hat{W}^*\right]^{-1}$. The reason that we are interested in $\hat{\text{Var}}(\hat{\delta}_{3SLS}^*)$ is because our objective is to obtain the design $X$ that minimizes $|\hat{\text{Var}}(\hat{\delta}_{3SLS}^*)|$. Since normality is assumed throughout the research, then $\left[\hat{\text{Var}}(\hat{\delta}_{3SLS}^*)\right]^{-1}$ is called the information matrix estimate and is denoted by $\hat{M}_{3SLS} = \left[\hat{\text{Var}}(\hat{\delta}_{3SLS}^*)\right]^{-1}$. Thus, our objective would be to obtain the optimal design $X$ that maximizes the determinant of the information matrix estimates.

The Full Information Maximum Likelihood (FIML) procedure is another procedure for estimating the parameters of a simultaneous linear equation system and was established at the infancy stage of Econometrics (Anderson and Rubin 1949). Despite the fact that FIML estimators have many desirable properties, their normal equations are non-linear, making them hard to solve. The nonlinearity led to the necessity for many procedures that were developed to solve FIML normal equations (Koopmans, Rubin, and Leipnik 1950; Chernoff and Divinsky 1953; Brown 1959; Eisenpress 1962). All of these procedures require a numerical computational iterative procedure, which carries a computational burden. Because of that computational burden, the 2SLS and the 3SLS estimators became more popular in the 1960’s and 1970’s, but which was overcome due to advanced computing technology.

It is necessary to introduce the necessary definitions, notations, and methodology in order to obtain the FIML information matrix. For the structural equation model in the form $YB + X\Gamma + E = 0$ where the rows of $E$ are independently and normally distributed with vector mean zero and a $p \times p$ variance matrix $\Sigma$, the log likelihood function can be given by $\log L = \text{Constant} + n \log |B| + \frac{n}{2} \log |\Sigma| - \frac{1}{2} tr[(YB + X\Gamma)'(YB + X\Gamma)\Sigma^{-1}]$. First, to obtain the FIML estimators, we find the partial derivatives and set them to zero as follows:

$\frac{\partial \log L}{\partial B} = n(B')^{-1} - Y'(YB + X\Gamma)\Sigma^{-1} \Rightarrow n(\hat{B}')^{-1} - Y'(Y\hat{B} + X\hat{\Gamma})\hat{\Sigma}^{-1} = 0,$

(1)
\[ \frac{\partial \log L}{\partial \Gamma} = -X'(YB + X\Gamma)\Sigma^{-1} \Rightarrow -X'(Y\hat{B} + X\hat{\Gamma})\Sigma^{-1} = 0, \quad (2) \]

Because of the invariance properties of the ML estimators (Casella and Berger 2001), it would simplify the derivative if we derive the likelihood with respect to \( \Sigma^{-1} \) instead of \( \Sigma \). So,

\[ \frac{\partial \log L}{\partial \Sigma^{-1}} = \frac{n}{2} \Sigma - \frac{1}{2} (YB + X\Gamma)'(YB + X\Gamma) \Rightarrow \frac{n}{2} \hat{\Sigma} - \frac{1}{2} (Y\hat{B} + X\hat{\Gamma})'(Y\hat{B} + X\hat{\Gamma}) = 0. \quad (3) \]

These equations are difficult to solve for two reasons. The first is because of the presence of \( \hat{B} \) and \( (\hat{B}')^{-1} \) in the first equation, which leads to nonlinearity. The second reason is because of the restrictions in \( B \) and \( \Gamma \). Some of the values are 0 or 1, which means that the left-hand side in the first and second equations are derived with respect to only the unknown quantities, or the derivatives with respect to the unknown parameters, are equated to zero. Durbin (1988) proposed a transformation to the maximum likelihood equations that overcame these two problems. It simplified the computations and also made it easier to study the properties of FIML estimators and their advantages over 3SLS estimators.

Start by expanding (Eq. 3) to obtain

\[ \frac{n}{2} \hat{\Sigma} - \frac{1}{2} \hat{B}'Y'(Y\hat{B} + X\hat{\Gamma}) - \frac{1}{2} \hat{\Gamma}'X'(Y\hat{B} + X\hat{\Gamma}) = 0. \]

To simplify the equation, multiply both sides by 2, which results in

\[ n\hat{\Sigma} - \hat{B}'Y'(Y\hat{B} + X\hat{\Gamma}) - \hat{\Gamma}'X'(Y\hat{B} + X\hat{\Gamma}) = 0. \]

Next, pre-multiply by \( (\hat{B}')^{-1} \) and post-multiply by \( \hat{\Sigma}^{-1} \) to obtain

\[ n (\hat{B}')^{-1}\hat{\Sigma}^{-1} - (\hat{B}')^{-1} \hat{B}'Y'(Y\hat{B} + X\hat{\Gamma})\Sigma^{-1} - (\hat{B}')^{-1}\hat{\Gamma}'X'(Y\hat{B} + X\hat{\Gamma})\Sigma^{-1} = 0. \]
Finally, re-arrange to obtain

\[
\begin{align*}
\frac{n (\hat{B}')^{-1} - Y'(Y\hat{B} + X\hat{\Gamma})\hat{\Sigma}^{-1}}{=0} &= (\hat{B}')^{-1} \hat{\Gamma}'X'(Y\hat{B} + X\hat{\Gamma})\hat{\Sigma}^{-1}. \\
\end{align*}
\]

The left-hand side of the equation is equal to zero from (Eq. 1).

Therefore,

\[
(\hat{B}')^{-1} \hat{\Gamma}'X'(Y\hat{B} + X\hat{\Gamma})\hat{\Sigma}^{-1} = 0.
\]

(4)

Let \( W = [Y \ X], \hat{\gamma} = -X\hat{\Gamma}\hat{B}^{-1}, \bar{W} = [-X\hat{\Gamma}\hat{B}^{-1} \ X], \) and let \( C = \begin{bmatrix} B \end{bmatrix} \) and \( \bar{C} = \begin{bmatrix} \hat{B} \end{bmatrix} \). These new notations will allow us to rewrite (Eq. 2) and (Eq. 4) in terms of the new notations. These equations can be written as

\[
\begin{bmatrix}
(\hat{B}')^{-1} \hat{\Gamma}'X'(Y\hat{B} + X\hat{\Gamma})\hat{\Sigma}^{-1} \\
-\hat{\Gamma}'X'(Y\hat{B} + X\hat{\Gamma})\hat{\Sigma}^{-1}
\end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.
\]

Then by rewriting the previous equation,

\[
\begin{bmatrix}
(\hat{B}')^{-1} \hat{\Gamma}'X' \\
-\hat{\Gamma}'X'
\end{bmatrix}(Y\hat{B} + X\hat{\Gamma})\hat{\Sigma}^{-1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.
\]

Next,

\[
\begin{bmatrix}
-\hat{\Gamma}'X' \\
\bar{W}'\bar{W}^{-1}
\end{bmatrix} [Y \ X] \begin{bmatrix} \hat{B} \\ \Gamma \end{bmatrix} \hat{\Sigma}^{-1} = 0.
\]

Finally,

\[
\bar{W}'\bar{W} \hat{C} \hat{\Sigma}^{-1} = 0.
\]

(5)

These are under the same restrictions as (Eq. 2) and (Eq. 4), which implies that the only unknown elements of \( \hat{C} \) in the left-hand side are equated to zero.
Let \( \hat{c}(j) \) be the \( j^{th} \) column of \( \hat{C} \) where \( j = 1, \ldots, P \). This leads us to conclude that the \( j^{th} \) element of \( \hat{c}(j) \) is one since the leading diagonal of \( B \) consists of 1's. Also, other elements of \( \hat{c}(j) \) are zero by the original restrictions on \( B \) and \( \Gamma \). Let \( m_j \) be the number of unknown elements in \( \hat{c}(j) \) and let

\(-\hat{\delta}(j)\) be the \( m_j \times 1 \) vectors of these unknowns. In other words, \(-\hat{\delta}(j)\) represents the unknown coefficients of the \( j^{th} \) equation in the original model. Let the columns of \( W = [Y \ X] \) corresponding to unknown elements of \( \hat{c}(j) \) be rearranged as \( n \times m_j \) matrix called \( Q(j) \). Also, let \( \gamma(j) \) represent the \( j^{th} \) column of \( Y_{n \times p} \). The new notations will allow us to rewrite (Eq. 5) without restrictions. Note that \( W\hat{c}(j) = \gamma(j) - Q(j)\hat{\delta}(j) \) since the coefficient of \( \gamma(j) \) is one, \( Q(j) \) has coefficients of \(-\hat{\delta}(j)\), and the rest of columns of \( W \) have coefficients of zero. Thus,

\[ W\hat{C} = [\gamma(1) - Q(1)\hat{\delta}(1), \gamma(2) - Q(2)\hat{\delta}(2), \ldots, \gamma(p) - Q(p)\hat{\delta}(p)]. \]

By substituting this quantity in to (Eq. 5), it becomes

\[ W'\Sigma_{i=k}^{p} \left( \gamma(k) - Q(k)\hat{\delta}(k) \right) [\sigma^{k1}, \sigma^{k2}, \ldots, \sigma^{kp}] = 0. \tag{6} \]

where \( \sigma^{kj} \) is the \( kj^{th} \) element of \( \Sigma^{-1} \). Consequently, the \( j^{th} \) column of the left-hand side of (Eq. 6) is equal to \( W'\Sigma_{k=1}^{p} \left( \gamma(k) - Q(k)\hat{\delta}(k) \right) \sigma^{kj} = 0 \). Use similar notations as before and let \( \hat{Q}(j) \) be the columns of \( \hat{W}' \) that correspond to unknown elements of the vector \( \hat{c}(j) \). For that reason,

\[ \hat{Q}(j) \Sigma_{k=1}^{p} \left( \gamma(k) - Q(k)\hat{\delta}(k) \right) \sigma^{kj} = 0 \] is equivalent to the previous equation. The full system is obtained by stacking all of the equations and using similar technique as the 3SLS. Let

\[ Q = \begin{bmatrix} Q(1) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & Q(p) \end{bmatrix}, \]

\[ \hat{Q} = \begin{bmatrix} \hat{Q}(1) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \hat{Q}(p) \end{bmatrix}, \]
\[
\mathbf{y}^* = \begin{bmatrix} y_1 \ \vdots \ y_p \end{bmatrix},
\]
\[
\mathbf{\delta}^* = \begin{bmatrix} \delta_1 \ \vdots \ \delta_p \end{bmatrix},
\]

and
\[
\mathbf{G} = \begin{bmatrix} \sigma_{11} \mathbf{I} & \sigma_{12} \mathbf{I} & \cdots & \sigma_{1p} \mathbf{I} \\
\sigma_{21} \mathbf{I} & \sigma_{22} \mathbf{I} & \cdots & \sigma_{2p} \mathbf{I} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{p1} \mathbf{I} & \sigma_{p2} \mathbf{I} & \cdots & \sigma_{pp} \mathbf{I} \end{bmatrix}.
\]

Then Eq. 6 simplifies to
\[
\mathbf{Q}' \mathbf{G} (\mathbf{y}^* - \mathbf{Q} \mathbf{\delta}^*) = 0,
\]
which implies that,
\[
\mathbf{\hat{Q}}' \mathbf{G} \mathbf{\delta}^* = \mathbf{\hat{Q}}' \mathbf{G} \mathbf{y}^*.
\]

The advantage of this notation is that the elements of \( \mathbf{\delta} \) are unknown and it is an unrestricted set of equations. The restrictions on \( \mathbf{B} \) and \( \Gamma \) have been resolved by the notations. However, there is a problem in solving for \( \mathbf{\delta} \) since \( \mathbf{G} \) is unknown. Additionally, \( \mathbf{\hat{Q}} \) consists of columns of \( \mathbf{\hat{W}} = [-X\hat{\Gamma}\hat{B}^{-1} \ X] \), which is unknown because of \( \hat{\Gamma} \) and \( \hat{B} \). Therefore replace these values by initial values and replace

\( \mathbf{G} \) by its ML estimate, \( \mathbf{\hat{G}} \mathbf{\otimes} = \frac{1}{n} (\mathbf{Y} \hat{\mathbf{B}} + X\hat{\Gamma})' (\mathbf{Y} \hat{\mathbf{B}} + X\hat{\Gamma}) \mathbf{\otimes} \mathbf{I} \). Thus, (Eq. 7) Becomes.

\[
\mathbf{\hat{Q}}' \mathbf{G} \mathbf{\hat{\delta}}^* = \mathbf{\hat{Q}}' \mathbf{G} \mathbf{y}^*.
\]
Let \( r \) be the number of iterations. Now, use an iterative “Newton-Raphson” algorithm to solve

\[
\hat{\delta}^*_{(r+1)} = \left( \hat{Q}'_r \hat{G}_r \hat{Q}_r \right)^{-1} \hat{Q}'_r \hat{G}_r \hat{y}^* \] until convergence is achieved. Since we are interested in minimizing the estimated variance of the estimates, Durbin (1988) proves that

\[
\widetilde{\text{Var}} \left( \hat{\delta}^*_{ML} \right) = \left( \hat{Q}'_r \hat{Q}_r \right)^{-1}. \] This result is mathematically justified in Appendix 2B and will be used later in Chapter 3 to obtain the optimal design \( X \) that maximizes the determinant of the information matrix estimates.

2.4 Summary

In univariate optimal design theory, a function of the variance of the estimates, or of the information matrix, and search algorithms were used to obtain an optimal design. However, there is no current literature to support a similar objective for optimal design for a causal structure. Therefore, the 3SLS information matrix, the FIML information matrix, and the search algorithms that were discussed in Section 2.1.3 will be used to obtain an optimal design for a given causal structure in the next chapters.

2.5 References


Appendix 2A: Derivation of the asymptotic 3SLS information matrix.

To justify the asymptotic covariance matrix of $\delta_{3SLS}^*$, we start with

$$
\begin{bmatrix}
X'y_{(1)} \\
\vdots \\
X'y_{(p)}
\end{bmatrix}
$$

as noted in Section 2.3.4. Then, using the Kronecker product,

$$(I \otimes X')Y^* = (I \otimes X')W^* \delta^* + (I \otimes X')\xi^*,$$

where $E(\xi^* \cdot \xi^{*\prime}) = \Sigma \otimes I$. The elements of $\Sigma$ are of $O(n)$ and the elements of $(I \otimes X')\xi^*$ and $(I \otimes X')W$ are of $O(n)$ in probability (Zellner and Theil 1962). Next:

$$
\text{Var}[(I \otimes X')\xi^*] = (I \otimes X')\text{Var}(\xi^*)(I \otimes X')
$$

$$
= (I \otimes X')[\Sigma \otimes I](I \otimes X)
$$

$$
= (\Sigma \otimes [X'X])
$$

$$
= \begin{bmatrix}
\sigma_{11}[X'X] & \sigma_{12}[X'X] & \cdots & \sigma_{1p}[X'X] \\
\sigma_{21}[X'X] & \sigma_{22}[X'X] & \cdots & \sigma_{2p}[X'X] \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{p1}[X'X] & \sigma_{p2}[X'X] & \cdots & \sigma_{pp}[X'X]
\end{bmatrix}.
$$

Assuming that $\Sigma = \begin{bmatrix}
\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\
\sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\
\sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp}
\end{bmatrix}$, let $\Sigma^{-1} = \begin{bmatrix}
\sigma_{11}^{-1} & \sigma_{12}^{-1} & \cdots & \sigma_{1p}^{-1} \\
\sigma_{21}^{-1} & \sigma_{22}^{-1} & \cdots & \sigma_{2p}^{-1} \\
\sigma_{p1}^{-1} & \sigma_{p2}^{-1} & \cdots & \sigma_{pp}^{-1}
\end{bmatrix}$. But $\sigma_{ij}$ are unknown, therefore let $\hat{\delta}_{2SLS}^{ij} = S^{ij}$ where $S^{ij}$ differs from $\sigma_{ij}$ by $O(n^{-1})$ (Zellner and Theil 1962). Using the previous notation:

$$
\hat{\Sigma} \otimes [X'X] = \Sigma \otimes [X'X] + \Delta_1.
$$

Then the elements of $\hat{\Sigma}$ are $O(n)$ and the elements of $\Delta_1$ are of $O(n^2)$ in probability. Now, to
estimate \( \hat{\delta} \), use the Generalized Least Square (GLS) \( \hat{\delta}_{SLS}^{*} \) as:

\[
\left( (I \otimes X')W^* \right) \left( \Sigma^{-1} \otimes [X'X]^{-1} \right) \left( (I \otimes X')W^* \right)^{-1} \left( (I \otimes X')W^* \right) \left[ \Sigma^{-1} \otimes [X'X]^{-1} \right] \left( (I \otimes X')W^* \right)^{-1} \left( (I \otimes X')W^* \right) \left( \Sigma^{-1} \otimes [X'X]^{-1} \right) \left( (I \otimes X')W^* \right)^{-1} \left( (I \otimes X')W^* \right) \Sigma^{-1} \otimes [X'X]^{-1}. \]

Since \( (I \otimes X')y^* = (I \otimes X')W^* \hat{\delta} + (I \otimes X')\varepsilon^* \), multiply both sides by

\[
\left( (I \otimes X')W^* \right) \left( \Sigma^{-1} \otimes [X'X]^{-1} \right) \left( (I \otimes X')W^* \right)^{-1} \left( (I \otimes X')W^* \right) \left( \Sigma^{-1} \otimes [X'X]^{-1} \right), \]

which results in

\[
\hat{\delta}_{SLS}^{*} - \delta \tag{10}
\]

From Eq. 1, \( \hat{\Sigma}^{-1} \otimes [X'X]^{-1} = (\Sigma \otimes [X'X] + \Delta_1)^{-1} \), which can be simplified to

\[
\Sigma^{-1} \otimes [X'X]^{-1} + \Delta_2 \tag{11}
\]

where \( \Delta_2 \) is of \( O(n^{-\frac{3}{2}}) \) since \( \Sigma \) is of \( O(n) \) and \( \Delta_1 \) is of \( O(n^{\frac{1}{2}}) \) in probability. \( W^* \) can be decomposed to endogenous and exogenous variables since \( W_{(i)} = (X_{(i)} Y_{(i)}) \). Let \( \tilde{W}^* = E(W^*|X) \). For that reason,

\[
(I \otimes X')W^* = (I \otimes X') \tilde{W}^* + \Delta_3 \tag{12}
\]

\( \Delta_3 \) is a matrix that consists of zeros except at the diagonal blocks, which consists of the reduced form residual multiplied by \( X' \) (Zellner and Theil 1962). For that reason, the elements of \( \Delta_3 \) are of \( O(n^{-\frac{1}{2}}) \) in probability. From Eq. 3 and Eq. 4 we can conclude that \( (I \otimes X') \tilde{W}^* \)' * \( \hat{\Sigma}^{-1} \otimes [X'X]^{-1} \) = \( \left[ (I \otimes X') \tilde{W}^* + \Delta_3 \right] \left[ \Sigma^{-1} \otimes [X'X]^{-1} + \Delta_2 \right] \), which results in

\[
\left( (I \otimes X') \tilde{W}^* \right)' \Sigma^{-1} \otimes [X'X]^{-1} + \Delta_4 \tag{13}
\]

Where \( \left( (I \otimes X') \tilde{W}^* \right)' \Sigma^{-1} \otimes [X'X]^{-1} \) is of \( O(1) \) and \( \Delta_4 \) is of \( O(n^{-\frac{3}{2}}) \) in probability. Then,

\[
\left( (I \otimes X')W^* \right)' \Sigma^{-1} \otimes [X'X]^{-1} (I \otimes X')W^*
\]

\[
= \left[ \left( (I \otimes X') \tilde{W}^* \right)' \Sigma^{-1} \otimes [X'X]^{-1} + \Delta_4 \right] \left( (I \otimes X') \tilde{W}^* + \Delta_3 \right)
\]
\[
\left( (I \otimes X') \hat{W}^* \right)^\prime \Sigma^{-1} \otimes [X'X]^{-1} (I \otimes X') \hat{W}^* + \Delta_5
\]

Where \( (I \otimes X') \hat{W}^* \left( I \otimes X' \right)^{1/2} \) is of \( O(n) \) and \( \Delta_5 \) is of \( O(n^{-1/2}) \) in probability.

Taking the inverse of both sides, then

\[
\left[ (I \otimes X') \hat{W}^* \right)^\prime \Sigma^{-1} \otimes [X'X]^{-1} (I \otimes X') \hat{W}^* + \Delta_6
\]

Where \( \left[ (I \otimes X') \hat{W}^* \right)^\prime \Sigma^{-1} \otimes [X'X]^{-1} (I \otimes X') \hat{W}^* \) is of \( O(n^{-1}) \) and \( \Delta_6 \) is of \( O(n^{-1/2}) \) in probability.

Now multiply Eq. 4 and Eq. 6. \( \left[ (I \otimes X') \hat{W}^* \right)^\prime \Sigma^{-1} \otimes [X'X]^{-1} (I \otimes X') \hat{W}^* \) is of \( O(n^{-1}) \) and \( \Delta_6 \) is of \( O(n^{-1/2}) \) in probability. Next, from Eq. 2:

\[
\hat{\delta}_S = (I \otimes X') \hat{W}^* \left( I \otimes X' \right)^{1/2} \left( I \otimes X' \right)^{1/2} \right] (I \otimes X') \xi
\]

Where \( \left[ (I \otimes X') \hat{W}^* \right)^\prime \Sigma^{-1} \otimes [X'X]^{-1} (I \otimes X') \hat{W}^* \) is of \( O(n^{-1/2}) \) and \( \Delta_7 \xi \) is of \( O(n^{-1}) \) in probability. Since \( \xi \) is a vector variable which are of \( O(n^{-1}) \) in probability, the leading term of \( (\hat{\delta}_S^{\prime} \hat{\delta}_S - \hat{\delta}^\prime \hat{\delta})' \) is
\[
\left( (I \otimes X') \hat{W}^* \right)' \Sigma^{-1} \otimes [X'X]^{-1} (I \otimes X') \hat{W}^* \right]^{-1} \left( (I \otimes X') \hat{W}^* \right)' \Sigma^{-1} \otimes [X'X]^{-1} \xi^* .
\]

The expected value of the leading term is \[
\left( (I \otimes X') \hat{W}^* \right)' \Sigma^{-1} \otimes [X'X]^{-1} (I \otimes X') \hat{W}^* \right]^{-1} \] which is of \( O(n^{-1}) \) from Eq. 6 and all other terms are of higher order of smallness. Therefore, the asymptotic covariance matrix of \( \delta_{3SLS}^* \) is \[
\left( (I \otimes X') \hat{W}^* \right)' \Sigma^{-1} \otimes [X'X]^{-1} (I \otimes X') \hat{W}^* \right]^{-1} \] (Zellner and Theil 1962).
Appendix 2B: Derivation of the FIML information matrix.

Starting with the notations as denoted by Durbin (1988), let the maximum likelihood estimate of $\hat{W} = [-XBA \ X]$ be denoted by $\hat{W}$ not by $\tilde{W}$. Similarly, as noted in Section 2.3.4, the estimate of $\hat{Q}$ which is the arrangement of $\hat{W}$ is denoted by $\tilde{Q}$. Finally, the maximum likelihood estimate of $\delta^*$ is $\hat{\delta}^*$.

Using the results from Section 2.3.4, $\tilde{Q}' \tilde{G} \delta^* = \tilde{Q}' \tilde{G} y^*$. However, $A, B$, and $\tilde{G}$ are unknown. Therefore, an iterative approach needs to be used. Start by the 2SLS estimates which are of $O(n^{-\frac{1}{2}})$ in probability. To obtain the second estimates, let $\hat{\delta}^* = \hat{\delta}_{(1)}^* + d\hat{\delta}^*$. Assuming the first Taylor expansions, let $\tilde{Q}_{(1)}$ and $\tilde{G}_{(1)}$ be the values of $\tilde{Q}$ and $\tilde{G}$ evaluated at $\hat{\delta}_{(1)}^*$. Then,

$$
\tilde{Q}_{(1)} \tilde{G}_{(1)} \left( \tilde{Q}_{(1)} \hat{G}_{(1)} - y^* \right) + Q_{(1)} \tilde{G}_{(1)} Q d\hat{\delta}^* + (d\tilde{Q}' \tilde{G}_{(1)} + \tilde{Q}_{(1)} d\tilde{G}) (\tilde{Q}_{(1)} \hat{G}_{(1)} - y^*) = 0.
$$

But since $d\tilde{Q}' \tilde{G}_{(1)}$ and $\tilde{Q}_{(1)} d\tilde{G}$ are small in comparison to $\tilde{Z}_{(1)} \tilde{G}_{(1)}$, we ignore the third term and conclude that $d\delta = (Q_{(1)} \tilde{G}_{(1)} Q)^{-1} Q_{(1)} \tilde{G}_{(1)} \left( y^* - Q \delta_{(1)}^* \right)$. Therefore, $\hat{\delta}_{(2)}^* = \hat{\delta}_{(1)}^* + d\hat{\delta}^* = \hat{\delta}_{(1)}^* + (Q_{(1)} \tilde{G}_{(1)} Q)^{-1} \tilde{Q}_{(1)} \tilde{G}_{(1)} y^*$.

Repeating the previous procedure, we can generalize the results for the $(r + 1)$ estimates where $\hat{\delta}_{(r+1)}^* = (\tilde{Q}_r \tilde{G}_r Q)^{-1} \tilde{Q}_r \tilde{G}_r y^*$.

The next task is to derive the asymptotic variance matrix of the estimates. The original model can be written as $y^* = Q \delta^* + \varepsilon^*$. Multiplying both sides by $(\tilde{Q}' \tilde{G} Q)^{-1} \tilde{Q}' \tilde{G}$, we obtain $\hat{\delta}^* = \hat{\delta}^* + (\tilde{Q}' \tilde{G} Q)^{-1} \tilde{Q}' \tilde{G} \varepsilon^*$. The right-hand side can be rewritten and therefore $\hat{\delta}^* - \hat{\delta}^* = (\tilde{Q}' \tilde{G} Q)^{-1} \tilde{Q}' \tilde{G} \varepsilon^* + K_1 \varepsilon^*$ where the elements of $K_1 = (\tilde{Q}' \tilde{G} Q)^{-1} \tilde{Q}' \tilde{G} - (\tilde{Q}' \tilde{G} Q)^{-1} \tilde{Q}' \tilde{G}$ are $O\left(n^{-\frac{1}{2}}\right)$ compared with the elements of $(\tilde{Q}' \tilde{G} Q)^{-1} \tilde{Q}' \tilde{G}$. However, $(\tilde{Q}' \tilde{G} Q)^{-1} \tilde{Q}' \tilde{G}$ still includes random variables. Hence
\[
\bar{Q}'GQ = \begin{bmatrix}
Q_{(1)} & 0 & \cdots & 0 \\
0 & Q_{(2)} & \ddots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & Q_{(p)}
\end{bmatrix}
\begin{bmatrix}
\sigma^{i1}I & \cdots & \sigma^{ip}I \\
\vdots & \ddots & \vdots \\
\sigma^{p1}I & \cdots & \sigma^{pp}I
\end{bmatrix}
\begin{bmatrix}
Q_{(1)} & 0 & \cdots & 0 \\
0 & Q_{(2)} & \ddots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & Q_{(p)}
\end{bmatrix}.
\]

The results of the product can be rewritten as \(\sigma^{ij} \bar{Q}'_{(i)} Q_{(j)} = \sigma^{ij} \bar{Q}'_{(i)} \bar{Q}_{(j)} + \sigma^{ij} \bar{Q}'_{(i)} \left( Q_{(j)} - \bar{Q}_{(j)} \right)\).

Each column of \(Q_{(j)} - \bar{Q}_{(j)}\) consists of either zeros or independent random variables with 0 mean and constant variance. Also, \(n^{-1}X'X\) and \(n^{-1}\bar{Q}'_{(i)} \bar{Q}_{(j)}\) both converge to a finite positive-definite matrix. For that reason, the elements of \(\bar{Q}'_{(i)} \left( Q_{(j)} - \bar{Q}_{(j)} \right)\) are \(O\left(n^{-\frac{1}{2}}\right)\) compared with the elements of \(\bar{Q}'_{(i)} \bar{Q}_{(j)}\). Thus, the elements of \(Q'GQ\) differ from the elements of \(\bar{Q}'G\bar{Q}\) by terms of relative order \(O\left(n^{-\frac{1}{2}}\right)\). As a result, \((\bar{Q}'G\bar{Q})^{-1} \bar{Q}'G = (\bar{Q}'G\bar{Q})^{-1} \bar{Q}'G + K_2\) where the elements of \(K_2\) are of \(O\left(n^{-\frac{1}{2}}\right)\) compared with the elements of \((\bar{Q}'G\bar{Q})^{-1} \bar{Q}'G\). Based on that, \(\hat{\delta} - \delta = (\bar{Q}'G\bar{Q})^{-1} \bar{Q}'G\hat{\varepsilon} + (K_1 + K_2)\hat{\varepsilon}^*\). Taking the leading term only, \(Var(\hat{\delta}) = E\left([\hat{\delta}^* - \delta^*][\hat{\delta}^* - \delta^*]'\right) + O(n^{-1}) = E\left((\bar{Q}'G\bar{Q})^{-1} \bar{Q}'G\hat{\varepsilon}^*\hat{\varepsilon}G(\bar{Q}'G\bar{Q})^{-1}\right) + O(n^{-1}) = (\bar{Q}'G\bar{Q})^{-1} + O(n^{-1}).\)

Therefore, the \(Var(\hat{\delta}_{FIML}) = (\bar{Q}'G\bar{Q})^{-1} + O(n^{-1}).\)
CHAPTER 3: OPTIMAL DESIGN FOR ENDOGENOUS AND EXOGENOUS PARAMETERS

3.1 Introduction

There are two types of optimal designs that will be discussed in this chapter, optimal designs for a causal structure with no restrictions and optimal designs for a causal structure with blocks. In Section 3.2, we will use the estimates of the covariance matrices to obtain the optimal design for both 3SLS and FIML estimators. In Section 3.3, we will derive the estimate of the covariance matrices for a causal structure with random blocks. We will demonstrate how to use those matrices to obtain an optimal design for 3SLS and FIML estimators. The estimate of the covariance matrices for a causal structure with fixed blocks will be derived in Section 3.4. Again, we demonstrate how to use those matrices to obtain an optimal design for 3SLS and FIML estimators.

3.2 Optimal Design for a Causal Structure Model

The completely randomized design is the simplest possible design where treatments are randomized to experimental units without any restrictions. Since it is the simplest design and has no restrictions, the completely randomized optimal design for a causal structure will be the default design and is referred to in this research as an optimal design for a causal structure.
3.2.1 3SLS Optimal Design for a Causal Structure

Consistent with the optimal design for the univariate models, our objective would be to obtain a design $X$ that minimizes the determinant of the estimate of the covariance matrix of $\hat{\delta}_{3SLS}^*$ or equivalently maximizes the inverse of the determinant of the estimate of the covariance matrix of $\hat{\delta}_{3SLS}^*$. Specifically,

$$\min_X |\text{Var}(\hat{\delta}_{3SLS}^*)| = \max_X |\tilde{W}^* (\Sigma \otimes X [X'X]^{-1}X') \tilde{W}^*|$$

(15)

where $\tilde{W}^* = E(W^*)$ is the expectation of $W^*$.

An alternative way to write the previous equation is $\min_X |\text{Var}(\hat{\delta}_{3SLS}^*)| = \max_X |\tilde{W}'' \tilde{V}^{-1} \tilde{W}^*|$

where $\tilde{V}^{-1} = \Sigma^{-1} \otimes X[X'X]^{-1}X'$ and $\tilde{V}^{-1} = \Sigma^{-1} \otimes X[X'X]^{-1}X$. The objective is to obtain the design matrix $X$ that will minimize the variance of the estimates or maximize the inverse of the variance of the estimates, which can also be referred to as the information matrix estimates since normality is assumed for all causal structure models in this research. For this reason, the terms information matrix and the inverse of the variance of the estimates can be used interchangeably.

However, the variance of the 3SLS estimates is an asymptotic result which assumes that the sample size is large. The optimal design that we will obtain in this research is for a small sample size. Therefore, the performance of the optimal design for a causal structure will be compared to the univariate optimal design by computing the variance of the estimates through simulation both for the optimal design for a causal structure and for the univariate optimal design. The design which has variances that are consistently the smallest would be the better design. In the following example, the objective will be to use $\text{Var}(\hat{\delta}_{3SLS}^*)$ to obtain the 3SLS optimal design.

Example #1: Using the 3SLS methodology that was previously discussed in Section 2.3.4, the optimal design for the causal structure below (Fig. 3.1) assumes that the treatments are qualitative
treatments for twenty observations and assumes that the true values for $\gamma_{11}$, $\gamma_{22}$, and $b_{12}$ are 8, 2, and 5, respectively:

**Figure 3.1:** The path model for Example #1.

The $k^{th}$ observation can then be expressed mathematically as $y_{k1} = y_{11} x_{k1} + \varepsilon_{k1}$ and $y_{k2} = b_{12} y_{k1} + y_{22} x_{k2} + \varepsilon_{k2}$ where $\begin{bmatrix} \varepsilon_{k1} \\ \varepsilon_{k2} \end{bmatrix} \sim N(0, \Sigma)$ for $k = 1 \ldots 20$. In vector notation, the model can be expressed as $\begin{bmatrix} y_{(1)} \\ y_{(2)} \end{bmatrix} = \begin{bmatrix} \gamma_{11} x_{(1)} + \varepsilon_{(1)} \\ b_{12} y_{(1)} + y_{22} x_{(2)} + \varepsilon_{(2)} \end{bmatrix}$ where $\begin{bmatrix} \varepsilon_{(1)} \\ \varepsilon_{(2)} \end{bmatrix} \sim N(0, \Sigma \otimes I)$. In this example, $y_{(1)}$ is a $20 \times 1$ vector that includes the responses of the first endogenous variable, $y_{(2)}$ is a $20 \times 1$ vector that includes the responses of the second endogenous variable, $x_{(1)}$ is a $20 \times 1$ vector that indicates whether or not the first treatment is applied on the $k^{th}$ experimental unit as denoted by $(x_{k1} = 1)$ or $(x_{k1} = 0)$, $x_{(2)}$ is a $20 \times 1$ vector that indicates whether or not the second treatment is applied on the $k^{th}$ experimental unit as denoted by $(x_{k2} = 1)$ or $(x_{k2} = 0)$, $\varepsilon_{(1)}$ is a $20 \times 1$ vector of random residuals for the first endogenous variable, $\varepsilon_{(2)}$ is a $20 \times 1$ vector of random residuals for the second endogenous variable, $0$ is a $40 \times 1$ vector of zeros, $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, and $I$ is a $20 \times 20$ identity matrix. To obtain an optimal design, the algorithm starts with multiple random designs in order to avoid local optimality. For this specific example, 50 initial random designs were used, each design with size 20. The initial designs were constructed randomly. To improve the design, the algorithm compares each point in the design with the candidate points and make a simultaneous exchange with the candidate point that improves the optimality criterion from Eq. 1 the most. The exchanges will continue until no further improvement is achieved or the improvement is
sufficiently small. This procedure will be repeated for each of the 50 initial designs. The final designs will be compared. The design \( X \) that has the smallest \( \text{Var}(\delta^*_{2SLS}) \) from among the 50 final designs will be selected as the optimal design.

To ensure that the optimal design is a global, or absolute, optimal design and not a local optimal design, the algorithm was re-run multiple times to confirm that determinant of the information matrix for the final optimal designs from each run were equivalent. For this example, 50 initial designs were enough to produce equivalent designs. The number of initial designs may be adjusted up or down depending on the complexity of the model, the number of endogenous and exogenous variables, and the sample size until multiple runs produce the same determinant of the information matrix, and thus achieving convergence.

Using the previous algorithm, a 3SLS optimal design for 20 observations for the model in Example #1 is given in Table 3.1.

**Table 3.1:** 3SLS and FIML optimal design for Example #1 with a determinant of 134,784.

<table>
<thead>
<tr>
<th>#</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 )</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( x_2 )</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Since the two treatments are qualitative, the four possible candidate points are \((0, 0), (0, 1), (1, 0)\) and \((1, 1)\). In the optimal design for the causal structure for the model in Fig. 3.1, the point \((1, 1)\) was replicated nine times, the point \((1, 0)\) was replicated nine times, and the point \((0, 1)\) was replicated twice. The point \((0, 0)\) was not replicated, which is expected because the model was assumed to have no intercept. We are interested in the magnitude of the coefficients and therefore there is no interest in the intercept.

Table 3.1 was based on the true value of \( \bar{W}^* \) where the true parameters are used in \( \bar{W}^* \). However, the true values are unknown. There are two approaches that address the issue of the true values being unknown. First, when prior data are available, estimates of the parameters can be used to
estimate $\hat{V}ar(\hat{\delta}^*_{3SLS})$ instead of the true parameters. One question that arises with the use of this approach is the robustness of the design. In order to address the sensitivity of the true values of the parameters on the design, simulation is used to obtain the optimal design based on the estimates of the parameters.

When there is no prior data available, a second approach is a Bayesian approach which assumes a prior distribution for the endogenous variables, which is similar to the approach that is used by Mylona, Goos, and Jones (2014) to address the variance components optimal design for blocked or split-plot experiments. This approach allows for the uncertainty of the endogenous parameters. Using the Bayesian approach to obtain a 3SLS optimal design for a causal structure will be the subject of future work.

In this dissertation research, the information matrix estimates, or the estimate of the asymptotic covariance matrix of the parameter estimates, will be used to obtain a 3SLS optimal design. Three sets of data (Appendix 3A) were simulated for Figure 3.1 and these data were then used to estimate the parameters where the estimates were used in $\hat{W}^*$ instead of the true values. Even though the estimates deviated from the true values, the optimal design did not, which is a good indicator for the robustness of the design.

Optimal design theory is currently limited to univariate or multivariate applications where the biggest weakness of these approaches is that they ignore the causal structure and the endogenous parameters. Because of the lack of theoretical results and the algorithms to produce optimal designs for a causal structure, the general practice is that univariate optimal designs are used for causal models, which leads to a loss of efficiency. To obtain a univariate optimal design for Example #1 above, it will be assumed that the treatment will affect both endogenous variables in order to avoid the situation where there is an optimal design for each equation. In this case, we obtain an optimal design for both treatment combinations. Using the modified Fedorov
algorithm, an optimal design in the univariate case for two qualitative treatments (i.e. \( y = y_{11}x_{1}(1) + y_{22}x_{2}(2) + \epsilon \)) can be found in Table 3.2 where the four possible candidate points are (0, 0), (0, 1), (1, 0) and (1, 1). The three candidate points (0, 1), (1, 0) and (1, 1) are replicated seven times, seven times, and six times, respectively. For the same reasons as in the optimal design for a causal structure, the point (0, 0) was not replicated since the intercept was not included in the model.

A classical univariate optimal design for 20 observations for the model \( y = y_{11}x_{1}(1) + y_{22}x_{2}(2) + \epsilon \) is given below in Table 3.2.

**Table 3.2**: Classical optimal design for univariate model \( y = y_{11}x_{1}(1) + y_{22}x_{2}(2) + \epsilon \).

| #  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|----|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|
| \( x_1 \) | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 |
| \( x_2 \) | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 |

The efficiency, which is the ratio of the determinants of the information matrices, is a measure which will show the percentage of increase or decrease in the determinant of the information matrix. \( D \)-efficiency will then be used to compare the overall improvement between both designs (Atkinson and Donev 1989; Atkins and Cheng 1999). The results from Table 3.3 compare the efficiency of the univariate optimal design from Table 3.2 to the efficiency of the new optimal design from Table 3.1.

**Table 3.3**: 3SLS results for three simulations where Column #2 represents the parameters and Columns #3 – 5 represent the estimates of those parameters based on the simulation.

<table>
<thead>
<tr>
<th>Parameters and Information Matrices</th>
<th>Parameter Values</th>
<th>Estimates Based on Simulation 1</th>
<th>Estimates Based on Simulation 2</th>
<th>Estimates Based on Simulation 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y_{11} )</td>
<td>8</td>
<td>7.91</td>
<td>8.25</td>
<td>8.34</td>
</tr>
<tr>
<td>( y_{22} )</td>
<td>2</td>
<td>2.31</td>
<td>0.19</td>
<td>1.30</td>
</tr>
<tr>
<td>( b_{12} )</td>
<td>5</td>
<td>5.01</td>
<td>5.10</td>
<td>5.03</td>
</tr>
<tr>
<td>(</td>
<td>\text{M}</td>
<td>=</td>
<td>\text{Var}(\delta_{\text{univariate}})</td>
<td>^{-1} )</td>
</tr>
<tr>
<td>(</td>
<td>\text{M}_{3SLS}</td>
<td>=</td>
<td>\text{Var}(\delta_{3SLS})</td>
<td>^{-1} )</td>
</tr>
<tr>
<td>( \text{eff} =</td>
<td>\text{M}</td>
<td>/</td>
<td>\text{M}_{3SLS}</td>
<td>)</td>
</tr>
<tr>
<td>( D_{\text{eff}} = (</td>
<td>\text{M}</td>
<td>/</td>
<td>\text{M}_{3SLS}</td>
<td>)^{1/3} )</td>
</tr>
</tbody>
</table>
When comparing the univariate optimal design to the 3SLS optimal design for a causal structure, there was an approximately 18% increase in the determinant of the asymptotic information matrix estimates. The univariate optimal design is about 94% as \( D \)-efficient as the 3SLS optimal design for a causal structure in all four cases.

One criticism over the comparison of the optimal designs obtained above is that they are based on the asymptotic information matrices. However, the optimal designs were obtained for small samples sizes (20 observations), so it is important to ensure that the results are consistent for small samples sizes, as well. To verify the comparison, we simulated data based on the causal structure optimal design and data based on the univariate optimal design 100,000 times each. We then estimated the parameters for both designs using the GLS for \( \hat{\delta}^*_{3SLS} \) and calculated the covariance matrix of the parameter estimates for both designs, whose elements are

\[
\text{cov}(\hat{\delta}_i, \hat{\delta}_j) = \sum_{i=1}^{10^5} \sum_{j=1}^{10^5} (\delta_i - \hat{\delta}_i)(\delta_j - \hat{\delta}_j)/(10^5 - 1)
\]

where \( \hat{\delta}_i \) is the \( i \)th element of \( \hat{\delta}^*_{3SLS} \). The determinants of the covariance matrices of the parameters estimates were computed. This process was repeated three times. The determinant of the covariance matrix of the parameters estimates for the 3SLS causal structure optimal design was consistently smaller than the determinant of the covariance matrix of the parameters estimates for the univariate optimal design as shown in Table 3.4.

**Table 3.4**: Determinants of the covariance matrices for the 3SLS causal structure optimal design and the 3SLS univariate optimal design based on three simulations each with 100,000 data sets.

<table>
<thead>
<tr>
<th></th>
<th>Results based on 1st 100,000 data sets</th>
<th>Results based on 2nd 100,000 data sets</th>
<th>Results based on 3rd 100,000 data sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Determinant of the covariance matrix for the causal structure optimal design</td>
<td>7,613,179.5</td>
<td>7,718,001.3</td>
<td>7,290,743.0</td>
</tr>
<tr>
<td>Determinant of the covariance matrix for the univariate optimal design</td>
<td>9,008,371.8</td>
<td>8,893,254.3</td>
<td>9,035,191.3</td>
</tr>
<tr>
<td>Efficiency</td>
<td>85%</td>
<td>87%</td>
<td>81%</td>
</tr>
<tr>
<td>( D )-efficiency</td>
<td>95%</td>
<td>95%</td>
<td>93%</td>
</tr>
</tbody>
</table>
The results from Table 3.4 were consistent with the results for the comparison of the asymptotic information matrices. The 3SLS causal structure optimal design consistently produced a smaller determinant for the covariance matrix of the parameters estimates than the 3SLS univariate optimal design. Specifically, the determinant of the covariance matrix of the parameters estimates for the 3SLS causal structure optimal design was about 15% - 19% smaller than the determinant of the covariance matrix of the parameters estimates for the 3SLS univariate optimal design. Based on those results, the 3SLS causal structure optimal design was 5% - 7% more $D$-efficient than the univariate optimal design.

It is important to note that whether the true parameters or their estimates are used, the 3SLS optimal design for a causal structure did not change. A natural extension of this work is to consider the FIML optimal design since for the FIML estimators, so this will be discussed in the next subsection.

### 3.2.2 FIML Optimal Design for a Causal Structure

Durbin (1988) proposed a transformation for the maximum likelihood equations that simplified the computations and also made it easier to study the properties of the FIML estimators and their advantages over the 3SLS estimators. Based on his results, our objective is to obtain a design matrix $X$ such that

$$\max_X |M_{FIML}| = \max_X |\bar{Q}' \bar{G} \bar{Q}|$$

where $\bar{Q}$ is defined in Appendix 2B.

As with the 3SLS optimal design, the objective would be to obtain the design matrix $X$ that will minimize the estimate of the determinant of the covariance of the FIML estimates or maximize the determinant of the FIML information matrix estimate. Again, the performance of the optimal design for a causal structure will be compared to the univariate optimal design by computing $\text{Var}(\hat{\delta}_{FIML})$ through simulation. The design which has variances that are consistently the smallest would be the better design. In the following example, the objective will be to use the estimate of the information matrix to obtain the FIML optimal design.
For the causal structure in Example #1, the 3SLS optimal design is also the optimal design for the FIML estimates. Similar to 3SLS, the true parameters are unknown, so the FIML estimates are used to replace the true parameters for the three simulations. In practice, the information matrix is estimated based on the parameters estimates. This is similar to the approach of Goos and Vandebroek (2001; 2003) and Goos and Jones (2011) in the case of optimal design for random blocks or split-plots where they suggest using the estimates of the variance parameters or a reasonable guess, arguing that the design minimally depends on those values. The results obtained for an optimal design for a causal structure in this dissertation research are consistent with those of Goos and Vandebroek and Goos and Jones where the optimal design did not change when replacing the true parameters for their estimates.

The results shown in Table 3.5 show the efficiency of the optimal design from Table 3.1 as compared to the efficiency of the univariate optimal design from Table 3.2.

### Table 3.5: FIML results for three simulations where Column #2 represents the parameters and Columns #3 – 5 represent the estimates of those parameters based on the simulation.

<table>
<thead>
<tr>
<th>Parameters and Information Matrices</th>
<th>Parameter Values</th>
<th>Estimates Based on Simulation 1</th>
<th>Estimates Based on Simulation 2</th>
<th>Estimates Based on Simulation 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{11}$</td>
<td>8</td>
<td>7.98</td>
<td>8.20</td>
<td>8.36</td>
</tr>
<tr>
<td>$\gamma_{22}$</td>
<td>2</td>
<td>2.05</td>
<td>1.17</td>
<td>2.04</td>
</tr>
<tr>
<td>$b_{12}$</td>
<td>5</td>
<td>5.03</td>
<td>4.98</td>
<td>4.98</td>
</tr>
<tr>
<td>$</td>
<td>M</td>
<td>=</td>
<td>Var(\delta_{\text{Univariate}})</td>
<td>^{-1}$</td>
</tr>
<tr>
<td>$</td>
<td>M_{\text{FIML}}</td>
<td>=</td>
<td>Var(\delta_{\text{FIML}})</td>
<td>^{-1}$</td>
</tr>
<tr>
<td>$\text{eff} =</td>
<td>M</td>
<td>/</td>
<td>M_{\text{FIML}}</td>
<td>$</td>
</tr>
<tr>
<td>$D_{\text{eff}} = (</td>
<td>M</td>
<td>/</td>
<td>M_{\text{FIML}}</td>
<td>)^{1/2}$</td>
</tr>
</tbody>
</table>

The results for the FIML causal structure optimal design are similar to the results for the 3SLS causal structure optimal design. However, there is a drastic difference between the optimal designs for the univariate optimal design from Table 3.2 and the optimal design for the causal structure. Additionally, the optimal design for the univariate case was about 6.6% less efficient than the design for the causal structure. The loss of efficiency demonstrates the importance of
taking the endogenous parameters into account to obtain an optimal design in order to study the causal structure.

For both the 3SLS and FIML optimal design, the goal of the three simulations is to show that the optimal design did not change even though the estimate of the parameters deviated from the true value. However, it may be problematic to base a conclusion based on only three simulations, therefore 300 data sets were simulated and then the parameters were estimated. These parameters were used to obtain an optimal design based on the estimate of the parameters. In each of the 300 simulations, the optimal design did not change and was equivalent to the optimal design in Table 3.1.

Another way to demonstrate the robustness of the design is to incrementally change the value of the parameter(s) from the true value, then obtain an optimal design based on the value of the parameters, and finally compare the optimal design to the optimal design that was obtained based on the true value. For the model in Fig. 3.1, the optimal design depends only on $\gamma_{11} = 8$ since $\hat{W}^*$ includes only $y_{1(1)} = y_{11}x_{1(1)} + \varepsilon_{(1)}$. Table 3.1 is the optimal design for $\gamma_{11} = 8$. Now, we will change the value of $\gamma_{11}$ incrementally based on the interval $8 \pm 6 = [14, 2]$. We first start with 8 and then decrease $\gamma_{11}$ incrementally by 0.1 (i.e. 8, 7.9, 7.8, ..., 2). Then, we obtain the optimal design based on these values and check whether the optimal design is equivalent to the optimal design in Table 3.1. Similarly, we start with 8 and then increase $\gamma_{11}$ incrementally by 0.1 (i.e. 8, 8.1, 8.2, ..., 14). Then, again, we obtain the optimal design based on these values and check whether the optimal design is equivalent to the optimal design in Table 3.1. The optimal designs did not change even though the value of the parameters changed up to 75% from the true value. These results are consistent with Goos and Vandebroek (2001; 2003) and Goos and Jones (2011) where they argued that the optimal design depended minimally on the specific values.
3.3 Optimal Design for a Causal Structure Model with Blocks

In applied research, blocked designs are likely the most commonly used experimental designs since they can be used to account for variation attributable to sources other than treatments and can considerably improve the precision of the experiment. In many industrial experiments, there are more factors and logistics to consider, so split-plot designs and incomplete block designs are commonly used to account for those logistical issues. The block size in these kinds of designs are dictated by those logistical restrictions. For this reason, an incomplete block design is used in this dissertation research. However, if there are no restrictions on block size, then a complete factorial block design could be considered and the same algorithms can be used to obtain the optimal combination of the treatments.

In this section, a causal structure in a blocked design is considered. The first objective is to establish the blocked causal structure and then to obtain the 3SLS and FIML estimators for the endogenous and exogenous parameters, which will be used to estimate their information matrices. The second objective of this section is to use the algorithm described in Section 3.2.1 in Example #1 to obtain an optimal design for both the endogenous and the exogenous parameters. The last objective is to compare the efficiency of the optimal design to the classical univariate mixed model optimal design.
3.3.1 3SLS Optimal Design for a Causal Structure with Random Blocks

The model with random blocks is similar to the model for a completely randomized design except that the blocks effect needs to be added to every endogenous variable. Therefore, for the \(i\)th endogenous variable as the dependent variable, the model can be written as

\[
y(i) = Y(i) + X(i)Y(i) + Zu(i) + \epsilon(i)
\]

where all of the terms have been previously defined in Section 2.3.4 except for \(Z\) which is an \(n \times b\) matrix of the form 

\[
Z = \text{diag} \left[ \begin{array}{c}
1_{k_1}, 1_{k_2}, ..., 1_{k_b}
\end{array} \right]
\]

where \(k_i\) is the size of the \(i\)th block and \(u(i)\) are the blocks effect on the \(i\)th endogenous variable where 

\[
\begin{bmatrix}
u(i) \\
u(j)
\end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ \sigma_{u(i)}^2 \\ \sigma_{u(j)}^2 \\ \sigma_{u(i)}^2 \end{bmatrix} \otimes I_b \right).
\]

If the blocks have the same size, then 

\[
Z = \text{diag} \left[ \begin{array}{c}
1_{k}, 1_{k}, ..., 1_{k}
\end{array} \right] = I_{b} \otimes 1_k\] and the elements of \(u(i)\) and \(\epsilon(i)\) are assumed to be mutually independent and normally distributed with zero mean and variances \(\sigma_{u(i)}^2\) and \(\sigma_{\epsilon(i)}^2\), consecutively.

Let \(W(i) = [Y(i) \quad X(i)]\) and \(\delta(i) = [b(i) \quad y(i)]'\). Then, the model with random blocks can be rewritten as 

\[
y(i) = W(i)\delta(i) + Zu(i) + \epsilon(i).\]

Using the alternative method to obtain the 3SLS estimates (Schmidt 1976), multiply both sides by \(X'\) to obtain the normal equations 

\[
X'y(i) = X'W(i)\delta(i) + X'Zu(i) + X'\epsilon(i).
\]

Then stack all of the endogenous equations and use matrix notation to obtain

\[
\begin{bmatrix}
X'y_{(1)} \\
X'y_{(2)} \\
\vdots \\
X'y_{(p)}
\end{bmatrix} = \begin{bmatrix}
X'W_{(1)} & 0 & \cdots & 0 \\
0 & X'W_{(2)} & 0 & \vdots \\
\vdots & \vdots & \ddots & 0 \\
0 & \cdots & 0 & X'W_{(p)}
\end{bmatrix} \begin{bmatrix}
\delta_{(1)} \\
\delta_{(2)} \\
\vdots \\
\delta_{(p)}
\end{bmatrix} + \begin{bmatrix}
X'Z & 0 & \cdots & 0 \\
0 & X'Z & 0 & \vdots \\
\vdots & \vdots & \ddots & 0 \\
0 & \cdots & 0 & X'Z
\end{bmatrix} \begin{bmatrix}
u_{(1)} \\
u_{(2)} \\
\vdots \\
u_{(p)}
\end{bmatrix} + \begin{bmatrix}
X'\epsilon_{(1)} \\
X'\epsilon_{(2)} \\
\vdots \\
X'\epsilon_{(p)}
\end{bmatrix}.
\]
Let \( \mathbf{y}^* = \begin{bmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_p^* \end{bmatrix} \).

\[
\mathbf{w}^* = \begin{bmatrix} W_1 & 0 & \cdots & 0 \\ 0 & W_2 & \cdots & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & W_p \end{bmatrix},
\]

\[
\mathbf{\delta}^* = \begin{bmatrix} \delta_1^* \\ \delta_2^* \\ \vdots \\ \delta_p^* \end{bmatrix},
\]

\[
\mathbf{z}^* = \begin{bmatrix} Z & 0 & \cdots & 0 \\ 0 & Z & \cdots & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & Z \end{bmatrix},
\]

\[
\mathbf{u}^* = \begin{bmatrix} u_1^* \\ u_2^* \\ \vdots \\ u_p^* \end{bmatrix},
\]

and

\[
\mathbf{\epsilon}^* = \begin{bmatrix} \epsilon_1^* \\ \epsilon_2^* \\ \vdots \\ \epsilon_p^* \end{bmatrix}.
\]

Now use the Kronecker product to rewrite the previous model as

\[
(I \otimes X)^{\mathbf{y}^*} = (I \otimes X)W^* \delta^* + (I \otimes X'Z)u^* + (I \otimes X')\mathbf{\epsilon}^*.
\]

Then to estimate \( \delta^* \), use the GLS estimator to obtain \( \delta_{3SLS}^* = \)

\[
[[ (I \otimes X')W^* ] V^{-1} (I \otimes X')W^* ]^{-1} [ (I \otimes X')W^* ] V^{-1} (I \otimes X')y^* ] \text{ where } V = Var[(I \otimes X'Z)u^* + (I \otimes X')\mathbf{\epsilon}^*].
\]

Let \( Var(u^*) = \Sigma_u \otimes I_b \) and \( Var(\mathbf{\epsilon}^*) = \Sigma \otimes I_n \) where \( \Sigma_u \) is a \( p \times p \) nonnegative
definite matrix. Then, 
\[ \text{Var}[(I \otimes X'Z)\tilde{u}^* + (I \otimes X')\tilde{v}^*] = (I \otimes X'Z)\text{Var}[\tilde{u}^*](I \otimes X'Z)' + (I \otimes X')\text{Var}[\tilde{v}^*](I \otimes X')' \]

\[ = (I \otimes X')\text{Var}[\tilde{v}^*](I \otimes X')' \]

Using the same approach as described in Section 3.1.1, the asymptotic information matrix would be 
\[ M_{3SLS}^{-1} = \text{Var}(\hat{\delta}_{3SLS}) = [(I \otimes X')\hat{W}^*]'V^{-1}[(I \otimes X')\hat{W}^*] \]

where \( V = Z^*G^*Z'' + R^* \).

Based on the information matrix for the 3SLS estimates for the causal structure optimal design with random blocks, our objective would be to use the algorithm described in Section 3.2.1 in Example #1 to obtain the design \( X \) such that 
\[ \max_X |M_{3SLS}| = \max_X [(I \otimes X')\hat{W}^*]'V^{-1}[(I \otimes X')\hat{W}^*] \].

Since the information matrix depends on the endogenous variables through \( \hat{\Sigma} \) and \( \hat{W}^* \), we face the same predicament as in the case of the completely randomized optimal design for a causal structure because the endogenous variables are unknown.

The solutions that were proposed for the completely randomized optimal design for a causal structure in Section 3.2.1 are appropriate for the causal structure optimal design with random blocks. In addition, for an experiment with random blocks that is conducted in multiple stages, the data of one block (year, lab, day, etc.) will be utilized to obtain initial estimates. These initial estimates will be then used to obtain the optimal design for the rest of the blocks (other years, other labs, other days, etc.). Requiring preliminary information is a standard part of univariate design. If there is no prior data, then a reasonable guess can be used similar to the approach of Goos and Vandebroek (2001; 2003) and Goos and Jones (2011). In the next example, the objective is to use the asymptotic information matrix estimates to obtain an optimal design for a causal structure with four random blocks where the size of each block is four.

**Example #2**: Assume that there are three quantitative exogenous variables with two endogenous variables for the causal structure given in Fig. 3.2. Also assume that the true values for \( \gamma_{11}, \gamma_{22}, \gamma_{31}, \text{and } b_{12} \) are 8, 2, 3, and 5, respectively.
Figure 3.2: The path model for Example #2.

An optimal design for a causal structure with four random blocks with block size equal to four under the given assumptions can be found in Table 3.6.

Table 3.6: A 3SLS causal structure optimal design with random blocks where each block size is equal to four.

<table>
<thead>
<tr>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_3$</td>
<td>$x_1$</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
</tbody>
</table>

The expected determinant of the information matrix is 5,222,400 and the previous design is not orthogonal. The expected determinant of the information matrix of the orthogonal design in Table 3.7 is 4,784,128. The orthogonal design is the optimal design for the univariate mixed model with three treatments and four blocks each block with size four.

Table 3.7: An orthogonal design for four blocks where each block size is equal to four.

<table>
<thead>
<tr>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_3$</td>
<td>$x_1$</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

The new determinant of the design is about 9.2% more efficient than the determinant of the orthogonal design, which makes the new optimal design approximately 2.4% more $D$-efficient than the orthogonal design. The significance in the increase in efficiency of the determinant may not be easily identified through our simple example. However, it is important to keep in mind
that in Example #2, there is only one endogenous parameter. In a more complex design, there will be more endogenous variables and if those variables are interconnected in more sophisticated way, then the increase of the determinant of the information matrix would be amplified. This will be demonstrated in Chapter 5.

It is worth noting the differences between the 3SLS causal structure optimal design and the orthogonal design. First, the orthogonal design completely confounds the three-way interaction \((x_{i1}x_{j2}x_{k3})\) with the blocks where \(x_{i1}\) is the \(i^{th}\) element of \(x_{(1)}\), \(x_{j2}\) is the \(j^{th}\) element of \(x_{(2)}\), and \(x_{k3}\) is the \(k^{th}\) element of \(x_{(3)}\). The number of replicates for each candidate point is 2. However, this is not the case in the 3SLS causal structure optimal design where the two-way interaction of \((x_{j2}x_{k3})\) is confounded with Blocks 1 and 3, the three-way interaction \((x_{i1}x_{j2}x_{k3})\) is confounded with Block 4, and the two-way interaction of \((x_{i1}x_{k3})\) is confounded with Block 2. In the 3SLS causal structure optimal design, the number of candidate points is not replicated equally. For example, the candidate point \((-1, -1, -1)\) was replicated three times, but the candidate point \((1, -1, -1)\) was replicated only once.

To understand the differences between the designs, it is important to look more critically at the asymptotic information matrix of the 3SLS optimal design for a causal structure and where information is gained versus the asymptotic information matrix for the orthogonal design and where information is sacrificed. Tables 3.8 and 3.9 show the asymptotic information matrices for the model in Fig 3.2 for the orthogonal design and the 3SLS optimal design, respectively.

**Table 3.8:** Asymptotic information matrix for Fig 3.2 based on the orthogonal design.

<table>
<thead>
<tr>
<th></th>
<th>(Y_{11})</th>
<th>(Y_{22})</th>
<th>(Y_{31})</th>
<th>(b_{12})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Y_{11})</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(Y_{22})</td>
<td>0</td>
<td>16</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(Y_{31})</td>
<td>0</td>
<td>0</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>(b_{12})</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1,168</td>
</tr>
</tbody>
</table>
Table 3.9: Asymptotic information matrix for Fig. 3.2 based on the 3SLS optimal design.

<table>
<thead>
<tr>
<th></th>
<th>( \gamma_{11} )</th>
<th>( \gamma_{22} )</th>
<th>( \gamma_{31} )</th>
<th>( b_{12} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_{11} )</td>
<td>16</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>( \gamma_{22} )</td>
<td>0</td>
<td>16</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( \gamma_{31} )</td>
<td>4</td>
<td>0</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>( b_{12} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1,360</td>
</tr>
</tbody>
</table>

When the information of the estimates is compared, we can see that the exogenous parameters in both models have the same information with the exception of \( \text{cov}(\hat{\gamma}_{11}, \hat{\gamma}_{31}) \). However, the endogenous parameters in the 3SLS optimal design gained more information with an increase of about 16% over the orthogonal design. These results are expected because univariate optimal designs do not take the precision of the endogenous parameters into account whereas optimal designs for causal structures do. Therefore, our new approach has the advantage of giving the optimal combination of treatments for the entire model and taking into account the precision of both the exogenous and endogenous parameters.

Another important estimation method to be considered is the maximum likelihood. Therefore, the FIML estimates for the endogenous and exogenous parameters and their information matrix will be derived.

3.3.2 FIML Optimal Design for a Causal Structure with Random Blocks

As in completely randomized optimal design for a causal structure, another important class of estimators are the maximum likelihood estimators. Because we are considering the entire system and not one equation at a time, we will obtain the FIML estimators for a causal structure with random blocks and then obtain the covariance matrix estimates for the estimators, or the information matrix estimates. The information matrix estimates will then be used to obtain an optimal design for the FIML estimators.
Starting with the notation from Section 3.3.1, $y^* - W^* \tilde{\delta}^* = Z^* u^* + \varepsilon^*$ where

$$[u^* \ \varepsilon^*]^T \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma_u \otimes I_b \otimes 1 \right).$$

Then, $\mathcal{L}(V, \tilde{\delta}^*, \gamma^*, W) = \frac{-np}{2} \log(2\pi) - \frac{1}{2} \log |V| - \frac{1}{2} \left( y^* - W^* \tilde{\delta}^* \right)' V^{-1} \left( y^* - W^* \tilde{\delta}^* \right)$ where $I_b$ is the $b \times b$ identity matrix and $I$ is an $n \times n$ identity matrix.

Derive the likelihood, $\frac{\partial \mathcal{L}}{\partial \tilde{\delta}^*} = W^{**} V^{-1} y^* - W^{**} V^{-1} W^* \delta^*$. Set the partial derivative equal to 0, giving $W^{**} V^{-1} y^* - W^{**} V^{-1} W^* \delta^* = 0$. Solve for $\delta^*$, making $\hat{\delta}^* = (W^{**} V^{-1} W^*)^{-1} W^{**} V^{-1} y^*$. Then, find the second partial derivative and take its expected value to obtain the information matrix, $\frac{\partial^2 \mathcal{L}}{\partial (\tilde{\delta}^*)' \partial \tilde{\delta}^*} = -W^{**} V^{-1} W^*$. Next, $Var^{-1}(\hat{\delta}^*) = E(W^{**} V^{-1} W^*)$.

However, in general, we cannot find the expected value since $W^*$ includes fixed and random variables. But, the following is a demonstration of how to find the information matrix of $\hat{\delta}^*$, which will be denoted hereon as $M(\hat{\delta}^*)$ for Example #2 above. This approach can be generalized to any other structural model.

By partitioning $V^{-1} = \begin{bmatrix} V_{11}^{-1} & V_{12}^{-1} \\ V_{21}^{-1} & V_{22}^{-1} \end{bmatrix}$, we obtain $W^{**} V^{-1} W^* = \begin{bmatrix} W'_1 V_{11}^{-1} W_1 & W'_1 V_{12}^{-1} W_2 \\ W'_2 V_{21}^{-1} W_1 & W'_2 V_{22}^{-1} W_2 \end{bmatrix}$.

Considering the values of $W_1 = [x'_1(1) \ x'_1(3)]'$ and $W_2 = [x'_2(2) \ x'_2(1)]'$, then

$$W^{**} V^{-1} W^* = \begin{bmatrix} \begin{bmatrix} W'_1 V_{11}^{-1} W_1 \\ x'_1(1) V_{11}^{-1} x_1(1) & x'_1(1) V_{12}^{-1} x_1(3) \\ \end{bmatrix} & \begin{bmatrix} W'_1 V_{12}^{-1} W_2 \\ x'_1(1) V_{12}^{-1} x_1(1) & x'_1(3) V_{12}^{-1} x_1(3) \end{bmatrix} \\ \begin{bmatrix} x'_2(1) V_{21}^{-1} x_2(1) & x'_2(1) V_{22}^{-1} x_2(3) \\ x'_2(2) V_{21}^{-1} x_2(1) & x'_2(2) V_{22}^{-1} x_2(3) \end{bmatrix} & \begin{bmatrix} W'_2 V_{22}^{-1} W_2 \\ x'_2(2) V_{22}^{-1} x_2(1) & x'_2(2) V_{22}^{-1} x_2(3) \end{bmatrix} \end{bmatrix}.$$
Let $\hat{y}(1) = E\left(y(1)\right)$. Then

$$
M_{\text{FIML}} = E(W^{*}V^{-1}W^*) = 
\begin{bmatrix}
  \sum_{1}V_{11}^{-1}x_{1} & \sum_{1}V_{12}^{-1}x_{2} & \sum_{1}V_{12}^{-1}\hat{y}_{1} \\
  \sum_{3}V_{11}^{-1}x_{3} & \sum_{3}V_{12}^{-1}x_{2} & \sum_{3}V_{12}^{-1}\hat{y}_{1} \\
  \sum_{2}V_{21}^{-1}x_{1} & \sum_{2}V_{22}^{-1}x_{2} & \sum_{2}V_{22}^{-1}\hat{y}_{1} \\
  \sum_{1}V_{21}^{-1}x_{1} & \sum_{1}V_{22}^{-1}x_{2} & \sum_{1}V_{22}^{-1}\hat{y}_{1} \\
  \sum_{2}V_{21}^{-1}x_{3} & \sum_{2}V_{22}^{-1}x_{2} & \sum_{2}V_{22}^{-1}\hat{y}_{1} \\
  \sum_{3}V_{21}^{-1}x_{3} & \sum_{3}V_{22}^{-1}x_{2} & \sum_{3}V_{22}^{-1}\hat{y}_{1} \\
  \sum_{1}V_{21}^{-1}x_{3} & \sum_{1}V_{22}^{-1}x_{2} & \sum_{1}V_{22}^{-1}\hat{y}_{1} \\
  \sum_{2}V_{21}^{-1}x_{3} & \sum_{2}V_{22}^{-1}x_{2} & \sum_{2}V_{22}^{-1}\hat{y}_{1} \\
  \sum_{3}V_{21}^{-1}x_{3} & \sum_{3}V_{22}^{-1}x_{2} & \sum_{3}V_{22}^{-1}\hat{y}_{1} \\
\end{bmatrix}
$$

Based on the information matrix above, a FIML optimal design for Example #2 is equivalent to the design in Table 3.6. The determinant of the information matrix is 5,280,295.4. The orthogonal design was not as efficient as the new optimal design since the determinant of the information matrix for the orthogonal design was 4,845,883.1. The increase in determinant was about 9%, which makes the new optimal design approximately 2.3% more $D$-efficient than the orthogonal design. These results are similar to the optimal design for Section 3.2.1 where both the 3SLS and the FIML estimates had the same optimal design. The new design was more efficient than the orthogonal design which is universally optimal in the univariate case.

### 3.4 Optimal Design for a Causal Structure with Fixed Blocks

The previous section assumes that the blocks are random. However, the blocks may be fixed. The optimality criteria for a fixed causal structure will be developed in the next section. Their properties will be studied and their efficiency will be compared with other designs. Our objective will be to obtain a design that focuses on the precision of these parameters while at the same time producing estimable block effects where the precision of the block parameters are not taken into account. This objective can be achieved by decomposing the information matrix and portioning out precision of the nuisance parameters. If there are $s$ parameters of interest and the rest are $g$ nuisance parameters, then the information matrix for the model would be

$$
M = 
\begin{bmatrix}
  \sum_{s}x_{s} \times s & \sum_{s}x_{s} \times g \\
  \sum_{s}x_{s} \times s & \sum_{s}x_{s} \times g \\
  \sum_{s}x_{s} \times s & \sum_{s}x_{s} \times g \\
\end{bmatrix}
$$
The optimal design for only the subset of parameters is called the \( D_s \)-optimum. It maximizes

\[
M_{s|g} = \left| M_{11} - M_{12} M_{22}^{-1} M_{21} \right|
\]

(Atkinson and Donev 1989; Goos and Vandebroek 2001).

The specific application in this dissertation research is the case of fixed blocks and the parameters of interest are the endogenous and exogenous parameters. However, the same approach can be generalized to any of the parameters. For instance, in many of the causal models, researchers may be interested only in the endogenous parameters, and the other exogenous parameters and blocks would be the nuisance parameters. For example, the interest of Fig. 1.1 was only in the endogenous parameters and that specific objective could be achieved by decomposing the information matrix and portioning out the exogenous parameters of precision.

The causal structure model for fixed blocks can be rewritten as

\[
y_\ast = W^* \delta^* + Z^* u^* + \varepsilon^*
\]

where all of the terms were defined previously in Section 3.3.2. Let \( W^* \delta^* + Z^* u^* = \)

\[
\begin{bmatrix} W^* \\ Z^* \end{bmatrix} \begin{bmatrix} \delta^* \\ u^* \end{bmatrix}
\]

Then, \( y_\ast = F^* \xi^* + \varepsilon^* \) assuming that \( \varepsilon^* \sim N(0, \Sigma \otimes I_n) \). Next, \( y_\ast - F^* \xi^* = \varepsilon^* \) and

\[
E \left( y_\ast - F^* \xi^* \right) = E(\varepsilon^*) = 0.
\]

Also, let \( V = Var \left( y_\ast - F^* \xi^* \right) = Var(\varepsilon^*) = \Sigma \otimes I \). The likelihood of this model is

\[
L(V, \xi^* | y^*, F^*) = \frac{\exp \left( -\frac{1}{2} V^{-1} \left( y^* - F^* \xi^* \right) \right)}{\sqrt{2\pi} |V|}.
\]

To obtain the information matrix, we find the second partial derivative with respect to \( \xi^* \), which is

\[
\frac{\partial^2 L}{\partial (\xi^*)^2} = -F^* V^{-1} F^*.
\]
3.4.1 3SLS Optimal Design for a Causal Structure with Fixed Blocks

\[ \mathbf{M}^{-1}_{3SLS} = \text{Var}(\hat{\delta}^{3SLS}) = [(I \otimes \mathbf{X}')\bar{\mathbf{W}}^*]'\bar{\mathbf{V}}^{-1}[(I \otimes \mathbf{X}')\bar{\mathbf{W}}^*] \] where \( \bar{\mathbf{V}} = \mathbf{Z}^*\mathbf{G}^*\mathbf{Z}^{**} + \mathbf{R}^* \). The covariance matrix estimates of \( \hat{\delta}^{3SLS} \) is \( \text{Var}(\hat{\delta}^{3SLS}) = \left( [(I \otimes \mathbf{X}')\bar{\mathbf{W}}^*]'\Sigma^{-1} \otimes [\mathbf{X}'\mathbf{X}]^{-1}(I \otimes \mathbf{X}')\bar{\mathbf{W}}^* \right)^{-1} \). Starting with the model in 3.2.1

\[
\begin{align*}
\begin{bmatrix}
X'y(1) \\
X'y(2) \\
\vdots \\
X'y(p)
\end{bmatrix} &= \begin{bmatrix}
X'W(1) & 0 & \cdots & 0 \\
0 & X'W(2) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & X'W(p)
\end{bmatrix} \begin{bmatrix}
\delta(1) \\
\delta(2) \\
\vdots \\
\delta(p)
\end{bmatrix} + \begin{bmatrix}
X'Z & 0 & \cdots & 0 \\
0 & X'Z & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & X'Z
\end{bmatrix} \begin{bmatrix}
\mathbf{u}(1) \\
\mathbf{u}(2) \\
\vdots \\
\mathbf{u}(p)
\end{bmatrix} + \begin{bmatrix}
X'\epsilon(1) \\
X'\epsilon(2) \\
\vdots \\
X'\epsilon(p)
\end{bmatrix}
\end{align*}
\]

use the Kronecker product to rewrite the model as \((I \otimes \mathbf{X})y^* = (I \otimes \mathbf{X})W^*\delta^* + (I \otimes \mathbf{X})Z^*\mathbf{u}^* + (I \otimes \mathbf{X})\epsilon^* \). Rewriting the model will give \((I \otimes \mathbf{X})y^* = (I \otimes \mathbf{X})[W^*\delta^* + Z^*\mathbf{u}^*] + (I \otimes \mathbf{X})\epsilon^* \). Now combine the blocks effect with the endogenous and exogenous parameters to obtain \(W^*\delta^* + Z^*\mathbf{u}^*\)

\[ \epsilon^* \sim N(0, \Sigma \otimes I) \]. Comparing the previous model by the model in Section 2.3.4, \( \text{Var}(\hat{\epsilon}^{3SLS}) = [\hat{F}^{**}(\Sigma^{-1} \otimes \mathbf{X}[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}')\hat{F}^{**}]^{-1} \). Let us denote \( \Sigma^{-1} \otimes \mathbf{X}[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}' \) by \( \bar{\mathbf{V}}^{-1} \). If \( \bar{\mathbf{V}}^{-1} \) is unknown, then it will be replaced by its estimate \( \bar{\mathbf{V}}^{-1} \) as denoted by \( \bar{\mathbf{V}}^{-1} \) as denoted by \( \bar{\mathbf{V}}^{-1} \). As in Section 3.2.2, the asymptotic information matrix would be \( \mathbf{M}(\hat{\epsilon}^{3SLS}) = \hat{F}^{**}\bar{\mathbf{V}}^{-1}\hat{F}^{**} = [\hat{W}^* \quad \hat{Z}^*]^{-1} \).

Multiply the matrices to obtain

\[
\mathbf{M}(\hat{\epsilon}^{3SLS}) = \hat{F}^{**}\bar{\mathbf{V}}^{-1}\hat{F}^{**} = \begin{bmatrix}
\mathbf{M}_{11} \\
\mathbf{M}_{12} \\
\mathbf{M}_{21} \\
\mathbf{M}_{22}
\end{bmatrix}
\]

Therefore, the information matrix for the endogenous and exogenous parameters is \( \mathbf{M}(\hat{\delta}^{3SLS}) = \mathbf{M}_{11} - \mathbf{M}_{12}\mathbf{M}_{22}^{-1}\mathbf{M}_{21} \). Next substitute the values of \( \mathbf{M}_{11}, \mathbf{M}_{12}, \mathbf{M}_{22}, \) and \( \mathbf{M}_{21} \) to obtain \( \mathbf{M}(\hat{\delta}^{3SLS}) = \).
\[ \hat{W}^*\hat{V}^{-1}\hat{W}^* - \hat{W}^*\hat{V}^{-1}Z^*(Z^*\hat{V}^{-1}Z^*)^{-1}Z^*\hat{V}^{-1}\hat{W}^*. \]

Our objective would be to obtain a design \( X \) such that
\[
\max_X |M(\hat{\delta}^*)| = \max_X |\hat{W}^*\hat{V}^{-1}\hat{W}^* - \hat{W}^*\hat{V}^{-1}(Z^*\hat{V}^{-1}Z^*)^{-1}Z^*\hat{V}^{-1}|.
\]

Assume that in Example #2 that the blocks are fixed. An optimal design for the causal structure based on the criteria above is equivalent to the design in Table 3.6. The determinant of the information matrix of the new optimal design is 5,222,400, whereas the orthogonal design information matrix determinant is 4,784,128. The 3SLS optimal causal structure design with fixed blocks increased the determinant by 9.2%, which means that the new design is approximately 2.4% more \( D \)-efficient than the orthogonal design. The 3SLS optimal causal structure design with fixed blocks was the same as the optimal causal structure design with random blocks for both 3SLS and FIML. This result is consistent due to the variance ratio, or degree of correlation, between the residual and the blocks, which is discussed further in Section 3.4.2.

### 3.4.2 FIML Optimal Design for a Causal Structure with Fixed Blocks

Similar to the discussion in Section 3.3.2, the 3SLS optimal design for a causal structure with fixed blocks is based on the asymptotic information matrix, which may not be the best method for an optimal design with small sample sizes. Therefore, it is necessary to consider the FIML optimal design for a causal structure with fixed blocks.

Starting with the notation from the previous section,
\[
-\frac{\partial^2 \mathcal{L}}{\partial (\hat{\xi}^*)^2} = F^*V^{-1}F^* = \begin{bmatrix} W^*V^{-1}W^* & W^*V^{-1}Z^* \\ Z^*V^{-1}W^* & Z^*V^{-1}Z^* \end{bmatrix}.
\]

Then,
\[
E\left(-\frac{\partial^2 \mathcal{L}}{\partial (\hat{\xi}^*)^2}\right) = E(F^*V^{-1}F^*) = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix},
\]

where
\[
M_{11} = E(W^*V^{-1}W^*), \quad M_{12} = E(Z^*V^{-1}W^*), \quad M_{21} = E(Z^*V^{-1}Z^*), \quad M_{22} = E(Z^*V^{-1}Z^*).
\]
where \( E(W^*|X) = \hat{W}^* \). But \( M_{11} = E(W^*V^{-1}W^*) \) depends on the causal structure. For our example, it can be found as discussed in Section 3.3.2. Therefore, the information matrix for the endogenous and exogenous parameters is \( M(\hat{\delta}^*) = M_{11} - M_{12}M_{22}^{-1}M_{21} = M(\hat{\delta}^*) = E(W^*V^{-1}W^*) - \hat{W}^*V^{-1}Z^*(Z^*V^{-1}Z^*)^{-1}Z^*V^{-1}\hat{W}^* \). Our objective would be to obtain a design \( X \) such that \( \max_X |M(\hat{\delta}^*)| = \max_X |E(W^*V^{-1}W^*) - \hat{W}^*V^{-1}Z^*(Z^*V^{-1}Z^*)^{-1}Z^*V^{-1}\hat{W}^*| \).

Assume that in Example #2 the blocks are fixed. An optimal design for the causal structure based on the criteria above is equivalent to the design in Table 3.6. The determinant of the information matrix for the new optimal design is 5,283,840. This is about a 9% increase in efficiency over the orthogonal design information matrix determinant of 4,849,664. This makes the new optimal design approximately 2.3% more \( D \)-efficient than the orthogonal design.

For our example, the optimal design did not change whether the blocks were considered random or fixed even though the optimality criteria are different. This motivated the need to understand the relationship between the two criteria. The following two corollaries can be proven and explain the relationship.

**Corollary #1:** If the blocks are random and \( \Sigma_u = \begin{bmatrix} \sigma_{u1}^2 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \sigma_{up}^2 \end{bmatrix} \), and \( \text{Var}(\varepsilon^*) = \begin{bmatrix} \sigma_1^2 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \sigma_p^2 \end{bmatrix} \otimes I \), then \( |W^*V^{-1}W^*| = \prod_{i=1}^p \frac{1}{\sigma_i^2} \left[ W'_iW(i) - \frac{\eta_i}{1 + k\eta_i} \sum_{j=1}^b \left( W'_{ij}1_k \right) \left( W'_{ij}1_k \right)' \right] \)

where \( W_{(ij)} \) is the part of \( W(i) \) that corresponds to the \( j^{th} \) block and \( \eta_i = \frac{\sigma_{ai}^2}{\sigma_i^2} \). Next,

\[
V = \begin{bmatrix} Z & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ Z & \ldots & 0 & 0 & \ldots & 0 \\ 0 & \ldots & Z & 0 & \ldots & 0 \\ 0 & \ldots & \sigma_{u1}^2I_b & 0 & \ldots & \sigma_{up}^2I_b \end{bmatrix} \otimes I
\]
\[
Z' \sigma^2_{u_i} Z + \sigma^2_i I \quad \ldots \quad 0 \\
\vdots \quad \ddots \quad \vdots \\
0 \quad \ldots \quad Z' \sigma^2_{u_p} Z + \sigma^2_p I
\]

where \( V = \text{diag}\{V_{11}, V_{12}, \ldots, V_{1b}, \ldots, V_{p1}, V_{p2}, \ldots, V_{pb}\} \). Then, \( V_{ij} = \sigma^2_i (l_k + \eta_i 1_k 1_k') \) where \( \eta_i = \frac{\sigma^2_{u_i}}{\sigma^2_i} \) and \( k \) is the block size. Next, \( V_{ij}^{-1} = \frac{1}{\sigma^2_i} \left( l_k - \frac{\eta_i}{1 + k\eta_i} 1_k 1_k' \right) \). Therefore,

\[
V^{-1} = \begin{bmatrix}
diag_{i=1}^b \left\{ \frac{1}{\sigma^2_i} \left( l_k - \frac{\eta_i}{1 + k\eta_i} 1_k 1_k' \right) \right\} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \text{diag}_{i=1}^b \left\{ \frac{1}{\sigma^2_p} \left( l_k - \frac{\eta_p}{1 + k\eta_p} 1_k 1_k' \right) \right\}
\end{bmatrix}
\]

As a consequence, \( W^{**} V^{-1} W^* = \begin{bmatrix} W_{(1)} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & W_{(p)} \end{bmatrix} \begin{bmatrix} V_{11}^{-1} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & V_p^{-1} \end{bmatrix} \begin{bmatrix} W_{(1)} & \cdots & 0 \end{bmatrix} \]

where

\[
V_{ij}^{-1} = \text{diag}_{i=1}^b \left\{ \frac{1}{\sigma^2_i} \left( l_k - \frac{\eta_i}{1 + k\eta_i} 1_k 1_k' \right) \right\} \].

Therefore, the previous equation can be simplified as

\[
|W^{**} V^{-1} W^*| = \prod_{i=1}^p |W_{(i)} V_{i1}^{-1} W_{(i)}| = \prod_{i=1}^p \left| \frac{1}{\sigma^2_i} \left( W_{(i)} W_{(i)} - \frac{\eta_i}{1 + k\eta_i} \Sigma_{j=1}^b (W'_{(ij)} 1_k) (W'_{(ij)} 1_k) \right) \right|.
\]

From the previous result, it can be seen why orthogonal designs may not be optimal. If the design is optimal, then one of the conditions is \( X'_i 1_k = 0 \). Therefore, in the univariate case, \(|X' V^{-1} X| = \frac{1}{\sigma^2_i} \left\{ X' X - \Sigma_{i=1}^b \frac{\eta_i}{1 + k\eta} (X'_i 1_k) (X'_i 1_k)' \right\} \) and if the design is orthogonal, then \( X'_i 1_k = 0 \). Thus, the information matrix of an orthogonal design is \( M_{orth} = \frac{1}{\sigma^2_i} X' X \) (Goos and Vandebroek 2001). If the observations are arranged such that for at least one \( X'_i 1_k \neq 0 \), then the difference between the two information matrices is \( M_{orth} - M_{not\ orth} = \Sigma_{i=1}^b \frac{\eta_i}{1 + k\eta} (X'_i 1_k) (X'_i 1_k)' \) since

\[
\Sigma_{i=1}^b \frac{\eta_i}{1 + k\eta} (X'_i 1_k) (X'_i 1_k)' \text{ is a nonnegative-definite matrix. Therefore, orthogonal blocking is optimal in the univariate case (Goos and Vandebroek 2001). However, for a causal structure, if } X'_i 1_k = 0 \text{ (an orthogonal design), it does not imply that } W'_{(ij)} 1_k = 0 \text{. As a consequence, blocking might not be optimal for a causal structure.}
\]
Corollary #2: If the blocks are fixed and $\text{Var}(\xi') = \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_p^2 \end{bmatrix} \otimes I$, then $|M(\hat{\xi}^*)| = \Pi_{i=1}^P \left| \frac{1}{\sigma_i^2} \{ W'_i W_i - \frac{1}{k} \Sigma_{j=1}^b (W'_{ij} 1_k) (W'_{ij} 1_k)' \} \right|$ where $W_{(ij)}$ is the part of $W_{(i)}$ that corresponds to the $j^{th}$ block. Then, the information matrix of the endogenous and exogenous parameters is $M(\hat{\xi}^*) = M_{11} - M_{12} M_{22}^{-1} M_{21}$ where

$$M_{22} = \begin{bmatrix} Z' & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Z' \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1^2} I & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sigma_p^2} I \end{bmatrix} \begin{bmatrix} Z \\ \vdots \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sigma_1^2} Z' Z & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sigma_p^2} Z' Z \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma_1^2} \text{diag}_i^b(k) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sigma_p^2} \text{diag}_i^b(k) \end{bmatrix}.$$

Then, $M(\hat{\xi}^*) = M_{11} - M_{12} M_{22}^{-1} M_{21}$

$$= \begin{bmatrix} W'_i \frac{1}{\sigma_i^2} W_i & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & W'_p \frac{1}{\sigma_p^2} W_p \end{bmatrix} \begin{bmatrix} W'_i \frac{1}{\sigma_i^2} Z & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sigma_p^2} Z' Z \end{bmatrix} = \begin{bmatrix} W'_i \frac{1}{\sigma_i^2} W_i - \frac{1}{\sigma_i^2} W'_i Z Z' Z^{-1} Z' W_i & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & W'_p \frac{1}{\sigma_p^2} W_p - \frac{1}{\sigma_p^2} W'_p Z Z' Z^{-1} Z' W_p \end{bmatrix}.$$

Therefore, $|M(\hat{\xi}^*)| = \Pi_{i=1}^P \left| W'_i \frac{1}{\sigma_i^2} W_i - \frac{1}{\sigma_i^2} W'_i Z(Z' Z)^{-1} Z' W_i \right|$. Thus, $|M(\hat{\xi}^*)| = \Pi_{i=1}^P \left| \frac{1}{\sigma_i^2} \{ W'_i W_i - \frac{1}{k} \Sigma_{j=1}^b (W'_{ij} 1_k) (W'_{ij} 1_k)' \} \right|$.

We can derive the following two conclusions from the previous two corollaries. First,

$$\lim_{{(\eta_1,\ldots,\eta_p) \rightarrow (\infty,\ldots,\infty)}} \frac{W'' V^{-1} W'}{\text{Random blk str}} = |M(\hat{\xi}^*)|,$$

Second,

$$\lim_{{(\eta_1,\ldots,\eta_p) \rightarrow (0,\ldots,0)}} \frac{W'' V^{-1} W'}{\text{Fixed blk str}} = \frac{|W'' W'|}{\text{No blk str}}.$$
From these results it can be seen that if $\eta_i$ is large, then a blocked causal structure design is optimal. It does not make a difference whether the blocks are fixed or random. However, if $\eta_i$ is small, then the completely randomized optimal causal structure design would be optimal for a causal structure with blocks. If the values of $\eta_i$ are neither too small nor too large, then the causal structure optimal design for random blocks may not be optimal for fixed blocks because the information matrix may be different depending on $\eta_i$ and the design. These results are also consistent with the univariate case (Goos and Vandebroek 2001). However, if the variance components are more complicated and there is a serial correlation among the blocks or among the observations, like a non-diagonal $\Sigma$ matrix or a non-diagonal $R$ matrix, respectively, then the design may change depending on whether the blocks are fixed or random.

Another important aspect to address is the comparison of the multivariate optimal design and the optimal design for a causal structure model. There is little available literature which addresses multivariate optimal design. More specifically, there is no literature for multivariate optimal design with blocks, split-plot multivariate optimal design, or multivariate optimal design for variance components. Despite this fact, multivariate optimal design is distinctly different from optimal design for a causal structure. The difference comes from the fact that causal structure optimal design considers the entire model while multivariate optimal design ignores the endogenous parameters. However, this dissertation research becomes more important and has greater impact because the methodology that was developed here can be used to obtain a multivariate optimal design since multivariate optimal design is a special case of optimal design for a causal structure.

The algorithms that were developed for optimal design for a causal structure in this dissertation research perform better than multivariate optimal design. What follows is the mathematical justification for this claim. The simple assumptions for Corollary #1 for the variance components will be used. Assume the model in Example #1 from Section 3.2.1 with two exogenous variables
and two endogenous variables. However, this case can be generalized by induction for more than two variables and more complex models.

Start with $W^*V^{-1}W^*$ where $W^* = \begin{bmatrix} W_{(1)} & 0 \\ 0 & W_{(2)} \end{bmatrix}$, $W_{(1)} = \mathbf{x}_{(1)}$, $W_{(2)} = \begin{bmatrix} \mathbf{x}_{(2)} \end{bmatrix}$, and $V = \begin{bmatrix} V_{11} & 0 \\ 0 & V_{22} \end{bmatrix}$. Then by substituting these values and multiplying we get,

$$W^*V^{-1}W^* = \begin{bmatrix} W_{(1)} & 0 \\ 0 & W_{(2)} \end{bmatrix} \begin{bmatrix} V_{11}^{-1} & 0 \\ 0 & V_{22}^{-1} \end{bmatrix} \begin{bmatrix} W_{(1)} & 0 \\ 0 & W_{(2)} \end{bmatrix} = \begin{bmatrix} W_{(1)}V_{11}^{-1}W_{(1)} & 0 \\ 0 & W_{(2)}V_{22}^{-1}W_{(2)} \end{bmatrix}.$$

Then, take the expected value in order to obtain the information matrix

$$E(W^*V^{-1}W^*) = \begin{bmatrix} E(W_{(1)}V_{11}^{-1}W_{(1)}) & 0 \\ 0 & E(W_{(2)}V_{22}^{-1}W_{(2)}) \end{bmatrix}.$$

So,

$$|E(W^*V^{-1}W^*)| = |E(W_{(1)}V_{11}^{-1}W_{(1)})||E(W_{(2)}V_{22}^{-1}W_{(2)})|$$

where, $E(W_{(1)}V_{11}^{-1}W_{(1)}) = \mathbf{x}_{(1)}\mathbf{x}_{(1)}^{-1}\mathbf{x}_{(1)}$ since $W_{(1)}$ has no endogenous variables. However, $W_{(2)}$ includes both endogenous and exogenous variables in that specific example. By replacing $W_{(2)}$ by its expected value and multiplying the quantity $W_{(2)}V_{22}^{-1}W_{(2)}$ and then taking the expected value for the product we obtain

$$|E(W_{(2)}V_{22}^{-1}W_{(2)})| = |E \begin{bmatrix} \mathbf{x}_{(2)} \mathbf{y}_{(2)} \end{bmatrix} V_{22}^{-1} \begin{bmatrix} \mathbf{x}_{(2)} \mathbf{y}_{(2)} \end{bmatrix}| = \begin{bmatrix} x_{(2)}'y_{(2)}V_{22}^{-1}x_{(2)} & x_{(2)}'y_{(2)}V_{22}^{-1}\hat{y}_{(1)} \\ \hat{y}_{(2)}'V_{22}^{-1}x_{(2)} & tr(V_{11}V_{22}^{-1}) + \hat{y}_{(1)}'V_{22}^{-1}\hat{y}_{(1)} \end{bmatrix}.$$

Using the fact that $|M| = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} = |M_{11}||M_{22} - M_{21}M_{11}^{-1}M_{12}|$ (Anderson 2003),

$$\begin{bmatrix} x_{(2)}'y_{(2)}V_{22}^{-1}x_{(2)} & x_{(2)}'y_{(2)}V_{22}^{-1}\hat{y}_{(1)} \\ \hat{y}_{(2)}'V_{22}^{-1}x_{(2)} & tr(V_{11}V_{22}^{-1}) + \hat{y}_{(1)}'V_{22}^{-1}\hat{y}_{(1)} \end{bmatrix} = \begin{bmatrix} x_{(2)}'y_{(2)}V_{22}^{-1}x_{(2)} & x_{(2)}'y_{(2)}V_{22}^{-1}\hat{y}_{(1)} \\ \hat{y}_{(2)}'V_{22}^{-1}x_{(2)} & tr(V_{11}V_{22}^{-1}) + \hat{y}_{(1)}'V_{22}^{-1}\hat{y}_{(1)} \end{bmatrix} - \begin{bmatrix} x_{(2)}'y_{(2)}V_{22}^{-1}x_{(2)} & x_{(2)}'y_{(2)}V_{22}^{-1}\hat{y}_{(1)} \\ \hat{y}_{(2)}'V_{22}^{-1}x_{(2)} & tr(V_{11}V_{22}^{-1}) + \hat{y}_{(1)}'V_{22}^{-1}\hat{y}_{(1)} \end{bmatrix}^{}^{-1} x_{(2)}'y_{(2)}V_{22}^{-1}\hat{y}_{(1)}.$$
Thus,

\[ |E(W'^*V^{-1}W^*)| = |x_1' V_{11}^{-1} x_1(1) || x_2' V_{22}^{-1} x_2(2) || (tr(V_{11} V_{22}^{-1}) + \hat{\gamma}_1' V_{22}^{-1} \hat{\gamma}(1)) - \\
\hat{\gamma}_1' V_{22}^{-1} x_2(2) (x_2' V_{22}^{-1} x_2(2))^{-1} x_2' V_{22}^{-1} \hat{\gamma}(1)|. \]

For the example \( \hat{\theta}^* = [\hat{\gamma}_{11}, \hat{\gamma}_{22}, \hat{b}_{12}]' \), the previous result can be rewritten in terms of the information matrices. Thus, \( |M(\hat{\theta}^*)| = |M(\hat{\gamma}_{11})||M(\hat{\gamma}_{22})||M(\hat{b}_{12})||\hat{\gamma}_{22}|. \)

Under the given assumptions, the determinant of the information matrix is the product of the information matrices for the parameters. Comparing this with the multivariate information matrix in addition to the assumptions above, it is assumed that there are no interrelationships among the endogenous variables. Thus, \( W^* = \begin{bmatrix} W_{1(1)} & 0 \\ 0 & W_{2(2)} \end{bmatrix} \), \( W_{1(1)} = x_1(1) \), \( W_{2(2)} = x_2(2) \) and

\[ E(W'^*V^{-1}W^*) = \begin{bmatrix} E(x_1' V_{11}^{-1} x_1(1)) & 0 \\ 0 & E(x_2' V_{22}^{-1} x_2(2)) \end{bmatrix}. \]

Finally, \( |E(W'^*V^{-1}W^*)| = |x_1' V_{11}^{-1} x_1(1) || x_2' V_{22}^{-1} x_2(2)|. \)

It can be seen that multivariate optimal design is a special case of causal structure optimal design since optimal design for a causal structure considers the information matrix for all multivariate parameters. However, the multivariate information matrix does not include any of the causal structure parameter. The multivariate optimal design may be less efficient than the optimal design for a causal structure since the loss of information on the exogenous parameters will be compensated for by the gain on the endogenous parameters.

### 3.5 Summary

Optimal design theory and applications largely focus on univariate designs with a smaller body of work done on multivariate designs. There is no current literature including work done for causal structure modeling and most natural phenomena include multiple endogenous variables which are
related in complicated ways. There are two weaknesses concerning univariate designs. First, univariate designs are incapable of addressing more than one endogenous variable that could be affected by different exogenous variables. The second weakness is that univariate designs ignore the endogenous parameters. Our new approach overcomes both of these weaknesses.

Optimality criteria were developed for both the 3SLS and the FIML optimal designs for a causal structure for both completely randomized and blocked designs. The 3SLS causal structure optimal design is based on the estimate of the asymptotic information matrix, but the FIML causal structure optimal design is not. Therefore, the FIML causal structure optimal design may be more appropriate for designs with small sample sizes.

Our results show that with a simple two-equation causal model, the optimal designs were the same for both FIML and 3SLS estimates. In the example of the optimal causal structure in a completely randomized experiment, the new optimal design increased the information matrix determinant by at least 20% over the univariate optimal design. Because of this increase, the univariate optimal design was only 93.4% as $D$-efficient as the new optimal design.

Similarly, for an optimal design for a causal model with random blocks, the preliminary endogenous parameters had a significant impact on the optimal design. The univariate mixed model is an orthogonal design which is universally optimal. However, it is not optimal for a blocked causal structure. As shown in the example of the blocked causal structure, the new optimal design in Table 3.6 increased the determinant of the information matrix by at least 9%, meaning that the orthogonal design was less efficient than the new optimal design. The orthogonal design in both cases for 3SLS and FIML estimates was consistently about 97.8% as $D$-efficient as the new optimal design in Table 3.6. These results were consistent whether the blocks were fixed or random.
The optimal designs that were produced using the new criteria show that there is a large
difference in efficiency and $D$-efficiency among the endogenous parameters estimates in
comparison to the endogenous parameters estimates using the univariate optimal design. The
results produced here show the importance of taking advantage of prior knowledge of the
intrarelationships of endogenous and exogenous parameters to develop an optimal design.
Moreover, in a multi-stage experiment that takes place over days, or years, or in different
laboratories, the results of the first stage can be used to determine the model of interest in order to
determine an optimal design for the rest of the experiment. For example, if blocking was done by
days, then the results of the first block can be used to determine an optimal design for the rest of
the experiment. These criteria will help to obtain optimal designs that will produce the most
precise estimates for the parameters of interest.

3.6 References

   Wiley & Sons, Inc.


5. Durbin, J. Maximum Likelihood Estimation of the Parameters of a System of Simultaneous

   John Wiley and Sons, Inc.


Appendix 3A: Simulation Data for Example #1.

\[ y_1 = \gamma_{11} x_1 + \epsilon_1 = 8 \times x_1 + RANNOR(J(20, \ 1, \ -1)) \]

\[ y_2 = b_{12} y_1 + \gamma_{22} x_2 + \epsilon_2 = 5 \times y_1 + 2 \times x_2 + RANNOR(J(20, \ 1, \ -1)) \]

<table>
<thead>
<tr>
<th></th>
<th>x1</th>
<th>x2</th>
<th>y1</th>
<th>y2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>8.1888015</td>
<td>39.941753</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>-0.797031</td>
<td>-3.407073</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>7.7272725</td>
<td>38.871138</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0.2282006</td>
<td>2.9872978</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>6.5661959</td>
<td>35.875938</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>8.3634771</td>
<td>43.359895</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
<td>-0.4531314</td>
<td>-0.5238347</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0</td>
<td>7.447288</td>
<td>37.495002</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>7.7676745</td>
<td>40.673612</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>1</td>
<td>1.6270485</td>
<td>10.395996</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>0</td>
<td>8.186618</td>
<td>40.884228</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1</td>
<td>9.0815066</td>
<td>46.25106</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>1</td>
<td>8.7976471</td>
<td>44.783606</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>0</td>
<td>0.3364256</td>
<td>4.9876448</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>1</td>
<td>7.4723439</td>
<td>40.446071</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>1</td>
<td>-1.666419</td>
<td>-6.960678</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>0</td>
<td>7.9286436</td>
<td>39.012416</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>1</td>
<td>7.4837984</td>
<td>40.239636</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>1</td>
<td>0.1233543</td>
<td>3.0981016</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>0</td>
<td>8.3666194</td>
<td>42.484332</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>x1</th>
<th>x2</th>
<th>y1</th>
<th>y2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>7.4917</td>
<td>36.2698</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>-1.5942</td>
<td>-6.5816</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>8.8291</td>
<td>45.1596</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0.1347</td>
<td>1.2905</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>6.6427</td>
<td>34.8429</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>10.1131</td>
<td>32.4445</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
<td>0.1283</td>
<td>3.8132</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0</td>
<td>6.9464</td>
<td>35.4441</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>7.3985</td>
<td>37.5921</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>1</td>
<td>-0.0349</td>
<td>0.4825</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>0</td>
<td>6.9548</td>
<td>37.5587</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1</td>
<td>7.8721</td>
<td>41.7751</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>1</td>
<td>9.6083</td>
<td>47.4664</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>0</td>
<td>-1.0976</td>
<td>-3.9758</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>1</td>
<td>8.2836</td>
<td>44.3541</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>1</td>
<td>1.4928</td>
<td>8.4968</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>0</td>
<td>8.1758</td>
<td>40.9745</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>1</td>
<td>9.2933</td>
<td>47.5727</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>1</td>
<td>1.3524</td>
<td>9.879</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>0</td>
<td>9.6038</td>
<td>47.1368</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>x1</th>
<th>x2</th>
<th>y1</th>
<th>y2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.65494</td>
<td>5.4918</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>-0.85199</td>
<td>-3.7691</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>9.31644</td>
<td>48.4541</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>8.28673</td>
<td>45.2127</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>7.47119</td>
<td>36.6587</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>7.57047</td>
<td>39.9873</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
<td>-0.64061</td>
<td>-1.8465</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0</td>
<td>9.37555</td>
<td>46.3297</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>9.15694</td>
<td>48.4716</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>1</td>
<td>0.13937</td>
<td>3.3175</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>0</td>
<td>7.8939</td>
<td>39.8941</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>0</td>
<td>8.49782</td>
<td>44.8599</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>1</td>
<td>7.48505</td>
<td>37.125</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>1</td>
<td>0.01591</td>
<td>3.6337</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>0</td>
<td>8.8905</td>
<td>45.4475</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>1</td>
<td>0.83026</td>
<td>6.8346</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>0</td>
<td>8.01341</td>
<td>39.7109</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>1</td>
<td>7.78786</td>
<td>40.9468</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>1</td>
<td>-0.26435</td>
<td>-9.3785</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>0</td>
<td>8.11854</td>
<td>41.3342</td>
</tr>
</tbody>
</table>
CHAPTER 4: OPTIMAL DESIGN FOR A CAUSAL STRUCTURE
FOR ENDOGENOUS PARAMETERS, EXOGENOUS
PARAMETERS, AND VARIANCE COMPONENTS

4.1 Introduction to Variance Components in Optimal Design for a Causal Structure with Mixed Effects

A weakness of the exchange algorithms for $D$-optimal designs in the univariate case is that they focus entirely on the precision of the fixed parameter estimates and ignore the estimation of variance components, which are required as part of the GLS estimates of the fixed effects. This approach leads to optimal estimates for the fixed effects but may produce poor estimates or, in the worst case, variance components that are not estimable (Mylona, Goos, and Jones 2014). Without the variance components estimates, inference is infeasible. However, in most cases, the variance components are estimable but their estimates lack precision which may negatively affect the quality of GLS estimates.

Because of the consequences of having non-estimable or poor estimates for the variance components, Khuri (1992) introduced a new design approach where the precision of the variance components estimates are taken into account. In general, however, there is only a very small portion of the design literature that focuses on the precision of variance components. The primary emphasis is put on the need to select designs for the simultaneous estimation of fixed effects and variance components for mixed models. In 2014, an algorithm was proposed for an optimal design for the variance components with a suggested extension in optimal designs for both fixed parameters and covariance parameters simultaneously (Loeza-Serrano and Donev 2014). That extension was presented by Mylona, Goos, and Jones (2014) where they proposed composite criteria which obtain a simultaneous optimal design for both the fixed parameters and covariance components, or random parameters. Their composite criterion was the weighted
average of the information matrices scaled by the number of the parameters. It used prior
distributions for the covariance parameters and a Gaussian quadrature technique to approximate
the Bayesian composite criteria.

A similar approach to the univariate case as described above can be used for causal structure with
a mixed model to account for endogenous and exogenous parameters and their variance
components. Our criteria would be defined as

\[
\Phi = \frac{(1 - \alpha)}{p + q} \log |M(\hat{\delta}^*)| + \frac{\alpha}{h} \log |M(\hat{\sigma})|
\]

where \(M(\hat{\delta}^*)\) is the information matrix for the endogenous and the exogenous parameters, \(M(\hat{\sigma})\)
is the information matrix for the covariance parameters, \((p + q)\) is the number of endogenous and
exogenous parameters, and \(h\) is the number of variance components with \(\alpha\) as a weight constant
with \(0 \leq \alpha \leq 1\).

In this dissertation research, our objective is to obtain an optimal design for a causal structure that
simultaneously accounts for the endogenous and exogenous parameters, and the variance
components. There are two proposed approaches to obtain the information matrix for the
variance components of a causal structure with mixed effects. As in Chapter 3, the first approach
uses 3SLS methodology and the second approach uses FIML methodology. Our objective is to
obtain an optimal design using both methodologies. The efficiency of these designs will be
compared to the efficiency of some previously published optimal designs for the composite
criteria in the univariate case. Lastly, since we are interested in the variance components,
restricted maximum likelihood (REML) estimators will be used to obtain an optimal design for a
causal structure.
4.2 3SLS Simultaneous Optimal Design for a Causal Structure for the Endogenous Parameters, Exogenous Parameters, and Variance Components Using a Composite Criterion

Writing the log likelihood of the model from Section 3.3.2, \( L(V, \delta^*|y^*, W^*) \) = 

\[
(2\pi)^{-\frac{np}{2}} |V|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} \left( y^* - W^* \delta^* \right)' V^{-1} \left( y^* - W^* \delta^* \right) \right). \]

Let \( \mathcal{L} = \log(L) \), then

\[
\mathcal{L}(V, \delta^*|y^*, W^*) = -\frac{np}{2} \log(2\pi) - \frac{1}{2} \log|V| - \frac{1}{2} \left( y^* - W^* \delta^* \right)' V^{-1} \left( y^* - W^* \delta^* \right). \]

Taking the partial derivatives, then \( \frac{\partial \mathcal{L}}{\partial \delta^*_i} = W^* V^{-1} y^* - W^* V^{-1} W^* \delta^* \) and \( \frac{\partial^2 \mathcal{L}}{\partial \delta^*_i \partial \delta^*_{i'}} = -W^* V^{-1} W^* \). Now replace \( W^* \) by \( \hat{W}^* = E(W^*) \). The estimate of the asymptotic information matrix for the endogenous and exogenous parameters is

\[
\hat{M}(\hat{\delta}^*) = \hat{W}^* V^{-1} \hat{W}^*. \tag{17}
\]

Assuming that \( \theta \) is one of the elements of \( V \), then \( \theta \) is either one of the parameters of \( R^* \) or one of the parameters of \( G^* \). Using the previous notations, if \( \theta \) is one of the parameters of \( R^* \), then \( \theta = \sigma_i^2 \) or \( \sigma_{ij} \). If \( \theta \) is one of the parameters of \( G^* \), then using the previous notations, \( \theta = \sigma_{vt}^2 \) or \( \sigma_{vij} \).

Thus,

\[
M(\hat{\delta}) = \frac{1}{2} \begin{bmatrix}
\left\{ \text{tr}\left( P^* Z^* \frac{\partial G^*}{\partial \theta_{d_1}} Z' \right) \right\}_{i,l'=1}^{v_d} & \left\{ \text{tr}\left( P^* Z^* \frac{\partial G^*}{\partial \theta_{d_1}} Z_1 \right) \right\}_{i=1, j'=1}^{v_d, v_r} \\
\left\{ \text{tr}\left( P^* Z^* \frac{\partial R^*}{\partial \theta_{r_j}} Z_1 \right) \right\}_{r=1, d=1}^{v_r, v_d} & \left\{ \text{tr}\left( P^* Z^* \frac{\partial R^*}{\partial \theta_{r_j}} Z' \right) \right\}_{i=1, j=1}^{v_r, v_r}
\end{bmatrix}
\]

\[
\text{where } P^* = V^{-1} - V^{-1} \hat{W}^* (\hat{W}^* V^{-1} \hat{W}^*)^{-1} \hat{W}^* V^{-1}, Z^* = I \otimes Z, G^* = \Sigma_u \otimes I_b, \text{ and } R^* = \Sigma \otimes I.
\]

Mylona, Goos, and Jones (2014) provided a univariate optimal design for the fixed and random (the covariance) parameters using composite criteria for two treatments with four blocks and each
block size equal to two. Since the covariance parameters were involved, the candidate points that they considered for each treatment were $-1$, $0$, or $1$, which amounts to nine possible candidate points for both treatments. This design will be used as a comparison for the optimal design for a causal structure with two treatments and two endogenous parameters.

Example #3: Use composite criterion $\Phi$ and the information matrix in Eq. 16 and the values of $M(\tilde{\delta})$ and $M(\tilde{\sigma})$ in Eq. 17 and Eq. 18, respectively, for the causal structure shown in Fig. 4.1 to obtain a 3SLS optimal design with random blocks for four blocks with block size equal to two.

Assume that $\sigma_{u_1}^2 = 0.25$, $\sigma_{u_2}^2 = 0.40$, $\sigma_{u_{12}} = 0.10$, $\sigma_1^2 = 1$, $\sigma_2^2 = 1$, and $\sigma_{12} = 0.50$.

**Figure 4.1:** The path model for Example #3.

Let $u^* = \begin{bmatrix} u_{1(1)} \\ u_{2(2)} \end{bmatrix}$ and $\epsilon^* = \begin{bmatrix} \epsilon_{1(1)} \\ \epsilon_{2(2)} \end{bmatrix}$ where $(u^*, \epsilon^*) \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} G^* & 0 \\ 0 & R^* \end{bmatrix} \right)$. A 3SLS optimal design with random blocks can be obtained using the information matrix that was developed above. Since the variance structures among the endogenous variables can differ, the following two cases will be considered based on the simplest and most complex variance structures. However, the method can be generalized to obtain an optimal design for other variance structures.

Case #1: Assume that the endogenous variables are not contemporaneously correlated ($\sigma_{12} = 0$ and $\sigma_{u_{12}} = 0$) where

$$G^* = \begin{bmatrix} \sigma_{u_1}^2 I_b & 0 \\ 0 & \sigma_{u_2}^2 I_b \end{bmatrix} = \begin{bmatrix} \sigma_{u_1}^2 & 0 \\ 0 & \sigma_{u_2}^2 \end{bmatrix} \otimes I_b.$$
\[ R^* = \begin{bmatrix} \sigma_1^2 I & 0 \\ 0 & \sigma_2^2 I \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \otimes I, \]

and

\[ \Phi = \frac{\alpha}{3} \log |M(\delta^*)| + \frac{\alpha}{4} \log |M(\delta)|. \]

Assume that there are four blocks, each of size two. Two different alphas will be chosen based on the recommendation of Mylona, Goos, and Jones (2014). Table 4.1 is the optimal design when \( \alpha = 0.5 \) and Table 4.2 is the optimal design when \( \alpha = 0.75 \).

Table 4.1: A 3SLS optimal design for four blocks with block size equal to two with \( \alpha = 0.5 \) with no contemporaneous correlation among the endogenous parameters or the observation within block.

<table>
<thead>
<tr>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 )</td>
<td>( x_2 )</td>
<td>( x_1 )</td>
<td>( x_2 )</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
</tr>
</tbody>
</table>

The value of the composite criteria for the endogenous parameters, exogenous parameters, and the variance components is equal to 2.29.

Table 4.2: A 3SLS optimal design for four blocks with block size equal to two with \( \alpha = 0.75 \) with no contemporaneous correlation among the endogenous parameters or the observation within block.

<table>
<thead>
<tr>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 )</td>
<td>( x_2 )</td>
<td>( x_1 )</td>
<td>( x_2 )</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

The value of the composite criteria for the endogenous parameters, exogenous parameters, and the variance components is equal to 1.62.
Case #2: Assume that the endogenous variables are contemporaneously correlated ($\sigma_{12} \neq 0$ and $\sigma_{u_{12}} \neq 0$) where

$$G^* = \begin{bmatrix} \sigma_{u_1}^2 l_b & \sigma_{u_{12}} l_b \\ \sigma_{u_{12}}^2 l_b & \sigma_{u_2}^2 l_b \end{bmatrix} = \begin{bmatrix} \sigma_{u_1}^2 & \sigma_{u_{12}} \\ \sigma_{u_{12}} & \sigma_{u_2}^2 \end{bmatrix} \otimes l_b,$$

$$R^* = \begin{bmatrix} \sigma_1^2 l & \sigma_{12} l \\ \sigma_{12} l & \sigma_2^2 l \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \otimes l,$$

and the composite criterion is

$$\Phi = \frac{(1-\alpha)}{3} \log |M(\hat{\delta}^*)| + \frac{\alpha}{6} \log |M(\hat{\sigma})|.$$

Table 4.3 shows the optimal design assuming that there are four blocks each of which are size two when $\alpha = 0.5$.

**Table 4.3:** A 3SLS optimal design for four blocks with block size equal to two with $\alpha = 0.5$ with contemporaneous correlation among the endogenous parameters and the observation within block.

<table>
<thead>
<tr>
<th></th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_1$</td>
</tr>
<tr>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_1$</td>
</tr>
<tr>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_1$</td>
</tr>
<tr>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_1$</td>
</tr>
</tbody>
</table>

The value of the composite criteria for the endogenous parameters, exogenous parameters, and the variance components is equal to 2.49.

The covariance structure did not have any impact on the optimal design in this example. In both cases for the same $\alpha$, the optimal design did not change. However, the choice of $\alpha$ had a large impact on the optimal design. Changing $\alpha$ from 0.5 to 0.75 changed the optimal design dramatically. The previous approach may work for the univariate case, but there are two disadvantages for a causal structure optimal design. First, the optimal design depends on $\alpha$.

Second, and more seriously, the endogenous parameters estimates are correlated to the variance.
components estimates in the causal structure model. Using composite criteria would ignore the correlation among these estimates and, therefore, the information matrices should not be separated.

For these reasons, composite criteria would not be appropriate in the case of causal structure models given the fact that $W^*$ is not fixed. The entire information matrix will be included in a causal structure model for the endogenous parameters, exogenous parameters, and the variance estimates with the correlation of the estimates. In this case, the objective would be to obtain a design that would maximize the determinant of that information matrix.

### 4.3 FIML Simultaneous Optimal Design for a Causal Structure for the Endogenous Parameters, Exogenous Parameters, and Variance Components

One of the main differences between the univariate case and the causal structure model is that the endogenous and exogenous parameters estimates are correlated. Also, the endogenous and variance covariance parameters estimates are correlated whereas in the univariate case, the fixed and random estimates are independent. The composite criterion does not take into account the correlation between the endogenous parameters estimates and the variance covariance parameters estimates.

The next approach would consider the information matrix for all of the parameters in order to take into account the correlation among them. The model for the $i^{th}$ endogenous variable is

$$y_{(i)} = Y_{(i)}h_{(i)} + X_{(i)}Y_{(i)} + Z_{(i)}u_{(i)} + \varepsilon_{(i)}.$$  

(19)
The elements of $u_i$ and $\varepsilon_i$ are assumed to be mutually independent and normally distributed with zero mean and variances $\sigma_u^2$ and $\sigma^2_i$, consecutively. Let $W_{(i)} = [Y_{(i)}^T \ X_{(i)}^T]$ and $\hat{\eta} = [\hat{b}^T \ \hat{y}^T]^T$.

Rewriting Eq. 19 will give us $Y_{(i)} = W_{(i)} \hat{\eta}_{(i)} + Z_{u_{(i)}} + \varepsilon_{(i)}$. Writing the previous equation in full system form gives:

$$
\begin{bmatrix}
Y_{(1)} \\
Y_{(2)} \\
\vdots \\
Y_{(p)} \\
\end{bmatrix} 
= 
\begin{bmatrix}
W_{(1)} & 0 & \cdots & 0 \\
0 & W_{(2)} & 0 & \vdots \\
\vdots & \vdots & \ddots & 0 \\
0 & \cdots & 0 & W_{(p)} \\
\end{bmatrix}
\begin{bmatrix}
\hat{\eta}_{(1)} \\
\hat{\eta}_{(2)} \\
\vdots \\
\hat{\eta}_{(p)} \\
\end{bmatrix} 
+ 
\begin{bmatrix}
Z & 0 & \cdots & 0 \\
0 & Z & 0 & \vdots \\
\vdots & \vdots & \ddots & 0 \\
0 & \cdots & 0 & Z \\
\end{bmatrix}
\begin{bmatrix}
\hat{u}_{(1)} \\
\hat{u}_{(2)} \\
\vdots \\
\hat{u}_{(p)} \\
\end{bmatrix} 
+ 
\begin{bmatrix}
\hat{\varepsilon}_{(1)} \\
\hat{\varepsilon}_{(2)} \\
\vdots \\
\hat{\varepsilon}_{(p)} \\
\end{bmatrix}.
$$

Now rewrite the previous model to get $y^* - W^* \hat{\delta}^* = Z^* u^* + \varepsilon^*$ where $u^* \sim \mathcal{N}(0, \Sigma_u \otimes I_b)$ where $\Sigma_u$ is a $p \times p$ nonnegative definite matrix, and $I_b$ is a $b \times b$ identity matrix. Also, $\varepsilon^* \sim \mathcal{N}(0, \Sigma \otimes I)$. where $\Sigma$ is a $p \times p$ nonnegative definite matrix, and $I$ is an $n \times n$ identity matrix. Thus, $E \left( y^* - W^* \hat{\delta}^* \right) = E \left( Z^* u^* + \varepsilon^* \right) = 0$ and

$$V = \text{Var}(y^* - W^* \hat{\delta}^*) = \text{Var}(Z^* u^* + \varepsilon^*) = Z^* \text{Var}(u^*) Z^{**} + \text{Var}(\varepsilon^*) = Z^* (\Sigma_u \otimes I_b) Z^{**} + \Sigma \otimes I.$$

Now re-expressing the likelihood and log likelihood gives $L(V, \delta^* | y^*, W^*) = (2\pi)^{-np/2} |V|^{-1/2} \exp \left[ -\frac{1}{2} \left( y^* - W^* \delta^* \right)' V^{-1} (y^* - W^* \delta^*) \right]$ and $L(V, \delta^* | y^*, W^*) = -\frac{np}{2} \log(2\pi) - \frac{1}{2} \log |V| - \frac{1}{2} \left( y^* - W^* \delta^* \right)' V^{-1} \left( y^* - W^* \delta^* \right)$, respectively.
The information matrix for all parameters is of the form
\[ M(\hat{\delta}, \psi) = \begin{bmatrix} \text{Var}(\hat{\delta}) & \text{Cov}(\hat{\delta}, \psi) \\ \text{Cov}(\hat{\delta}, \psi)' & \text{Var}(\psi) \end{bmatrix}^{-1}. \]

Obtaining the first partial derivative gives us
\[ \frac{\partial L}{\partial \hat{\delta}} = W''V^{-1}y - W''V^{-1}W^*\hat{\delta}, \]
thus the second partial derivative is
\[ \frac{\partial^2 L}{\partial \hat{\delta} \partial \delta'} = -W''V^{-1}W^*. \]
Taking the expectation for the previous results gives
\[ \text{Var}(\hat{\delta}) = E(W''V^{-1}W^*). \]
To obtain the information matrix for \( \hat{\delta} \), we must know the causal structure model. The causal structure model shown in Figure 4.1 from Section 4.2 will be used as a demonstration. However, the same method can be generalized to any causal structure.

Example #4: Redo Example #3 using ML estimates for the variance components. Assume there are four blocks with size two each.

The system can be written in matrix form as
\[ \begin{bmatrix} Y(1) \\ Y(2) \end{bmatrix} = \begin{bmatrix} X(1) & 0 & 0 & Y(1) \\ 0 & X(2) & 0 & Y(2) \end{bmatrix} + \begin{bmatrix} Z1 \\ 0 \\ 0 \\ Z1 \end{bmatrix} \begin{bmatrix} \hat{u}(1) \\ \hat{u}(2) \end{bmatrix} + \begin{bmatrix} \varepsilon(1) \\ \varepsilon(2) \end{bmatrix}, \]
thus
\[ y^* = W^*\delta^* + Z^*\hat{u} + \varepsilon^* \]
where \( V = \text{Var}(y^* - W^*\delta^*) = Z^*\text{Var}(\hat{u})Z^* + \text{Var}(\varepsilon^*) \)
and \( V = Z^* \begin{bmatrix} \sigma_{u1}^2 & \sigma_{u12} \\ \sigma_{u12} & \sigma_{u2}^2 \end{bmatrix} Z \) \( Z^* + \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_{22}^2 \end{bmatrix} \otimes I = Z^*GZ^* + R^* \). The variance of the estimates is \( \text{Var}(\hat{\delta}) = E(W''V^{-1}W^*) \)
where
\[ W^* = \begin{bmatrix} W(1) \\ 0 \\ W(2) \end{bmatrix} = \begin{bmatrix} W1 \\ \hat{x}(1) \\ 0 \\ \hat{x}(2) \\ Y(1) \end{bmatrix}. \]

Let \( V^{-1} = \begin{bmatrix} V_{11}^{-1} & V_{12}^{-1} \\ V_{21}^{-1} & V_{22}^{-1} \end{bmatrix} \). Therefore, \( W''V^{-1}W^* = \begin{bmatrix} \hat{x}(1) \\ \hat{x}(2) \end{bmatrix} \begin{bmatrix} i' & \hat{X}(1) \\ \hat{X}(2) \end{bmatrix} \begin{bmatrix} \hat{X}(1) \\ \hat{X}(2) \end{bmatrix} \begin{bmatrix} \hat{X}(1) \\ \hat{X}(2) \end{bmatrix} \).

Multiplying the right-hand side can be expressed as
\[
W''V^{-1}W^* = \begin{bmatrix} \hat{X}_1'V_{11}^{-1}\hat{X}_1 & \hat{X}_1'V_{12}^{-1}\hat{X}_2 & \hat{X}_1'V_{12}^{-1}Y(1) \\ \hat{X}_2'V_{21}^{-1}\hat{X}_1 & \hat{X}_2'V_{22}^{-1}\hat{X}_2 & \hat{X}_2'V_{22}^{-1}Y(1) \\ \hat{Y}_1'V_{21}^{-1}\hat{X}_1 & \hat{Y}_1'V_{22}^{-1}\hat{X}_2 & \hat{Y}_1'V_{22}^{-1}Y(1) \end{bmatrix}.
\]
Let $\hat{y}_{(1)} = E\left(y_{(1)}\right)$. Then,

$$E(W^*V^{-1}W^*) = \begin{bmatrix}
\hat{x}'(1)V_{11}^{-1}\hat{x}(1) & \hat{x}'(1)V_{12}^{-1}\hat{x}(2) & \hat{x}'(1)V_{12}^{-1}\hat{y}(1) \\
\hat{x}'(2)V_{21}^{-1}\hat{x}(1) & \hat{x}'(2)V_{22}^{-1}\hat{x}(2) & \hat{x}'(2)V_{22}^{-1}\hat{y}(1) \\
\hat{x}'(1)V_{21}^{-1}\hat{x}(1) & \hat{x}'(1)V_{22}^{-1}\hat{x}(2) & tr(V_{11}V_{22}^{-1}) + \hat{y}'(1)V_{22}^{-1}\hat{y}(1)
\end{bmatrix}.$$ 

Therefore, the information matrix $M(\hat{\delta}_{FIML})$ is

$$M(\hat{\delta}_{FIML}) = \begin{bmatrix}
\hat{x}'(1)V_{11}^{-1}\hat{x}(1) & \hat{x}'(1)V_{12}^{-1}\hat{x}(2) & \hat{x}'(1)V_{12}^{-1}\hat{y}(1) \\
\hat{x}'(2)V_{21}^{-1}\hat{x}(1) & \hat{x}'(2)V_{22}^{-1}\hat{x}(2) & \hat{x}'(2)V_{22}^{-1}\hat{y}(1) \\
\hat{x}'(1)V_{21}^{-1}\hat{x}(1) & \hat{x}'(1)V_{22}^{-1}\hat{x}(2) & tr(V_{11}V_{22}^{-1}) + \hat{y}'(1)V_{22}^{-1}\hat{y}(1)
\end{bmatrix}.$$ 

Using the results from Searle (1971), assume that $\theta$ is one of the elements of $V$, then $\theta$ is one of the parameters of $R^*$ using the previous notations, either $\theta = \sigma_i^2$ or $\sigma_{ij}$, or $\theta$ is one of the parameters of $G^*$ using the previous notations $\theta = \sigma_{ui}$ or $\sigma_{uij}$. The derivative of the variance matrix $V$ with respect to $\theta$ is $\frac{\partial V^{-1}}{\partial \theta} = -V^{-1}\frac{\partial V}{\partial \theta}V^{-1}$ and $\frac{\partial \ln|V|}{\partial \theta} = tr\left(V^{-1}\frac{\partial V}{\partial \theta}\right)$. Then, the derivative of the likelihood is

$$\frac{\partial L}{\partial \theta_i} = -\frac{1}{2} tr\left(V^{-1}\frac{\partial V}{\partial \theta} + \frac{1}{2} \left(y^* - W^*\hat{\delta}^*\right)'V^{-1}\frac{\partial V}{\partial \theta}V^{-1}\left(y^* - W^*\hat{\delta}^*\right)\right).$$

Since the second term is a scalar, then

$$\left(y^* - W^*\hat{\delta}^*\right)'V^{-1}\frac{\partial V}{\partial \theta}V^{-1}\left(y^* - W^*\hat{\delta}^*\right) = tr\left(V^{-1}\frac{\partial V}{\partial \theta}V^{-1}\left(y^* - W^*\hat{\delta}^*\right)\left(y^* - W^*\hat{\delta}^*\right)\right).$$

Using the properties of the trace operator, then

$$\frac{\partial L}{\partial \theta_i} = \frac{1}{2} tr\left(-V^{-1}\frac{\partial V}{\partial \theta_i} + V^{-1}\frac{\partial V}{\partial \theta_i}V^{-1}\left(y^* - W^*\hat{\delta}^*\right)\left(y^* - W^*\hat{\delta}^*\right)\right).$$
The second partial derivative is

\[
\frac{\partial^2 L}{\partial \theta_k \partial \theta_l} = \frac{1}{2} \text{tr} \left( V^{-1} \frac{\partial V}{\partial \theta_l} V^{-1} \frac{\partial V}{\partial \theta_k} V^{-1} \frac{\partial^2 V}{\partial \theta_k \partial \theta_l} \right) + \left\{ -V^{-1} \frac{\partial V}{\partial \theta_l} V^{-1} \frac{\partial V}{\partial \theta_k} V^{-1} + V^{-1} \frac{\partial^2 V}{\partial \theta_k \partial \theta_l} V^{-1} - V^{-1} \frac{\partial V}{\partial \theta_k} V^{-1} \frac{\partial V}{\partial \theta_l} \right\} \left( y^* - W^* \delta^* \right) \right). 
\]

Taking the expectation of the second derivative,

\[
-E \left( \frac{\partial^2 L}{\partial \theta_k \partial \theta_l} \right) = \frac{1}{2} \text{tr} \left( V^{-1} \frac{\partial V}{\partial \theta_l} V^{-1} \frac{\partial V}{\partial \theta_k} V^{-1} - V^{-1} \frac{\partial^2 V}{\partial \theta_k \partial \theta_l} \right) + \left\{ -V^{-1} \frac{\partial V}{\partial \theta_l} V^{-1} \frac{\partial V}{\partial \theta_k} V^{-1} + V^{-1} \frac{\partial^2 V}{\partial \theta_k \partial \theta_l} V^{-1} - V^{-1} \frac{\partial V}{\partial \theta_k} V^{-1} \frac{\partial V}{\partial \theta_l} \right\} V. 
\]

By simplifying the second derivative, \( E \left( \frac{\partial^2 L}{\partial \theta_k \partial \theta_l} \right) = \frac{1}{2} \text{tr} \left( V^{-1} \frac{\partial^2 V}{\partial \theta_k \partial \theta_l} \right) \). Thus,

\[
\text{Var}(\delta) = 2 \left[ \left\{ \text{tr} \left( V^{-1} Z^* \frac{\partial G^*}{\partial \theta_l} Z^* V^{-1} \frac{\partial G^*}{\partial \theta_j} Z^* \right) \right\}_{i,j=1}^m \left\{ \text{tr} \left( V^{-1} Z^* \frac{\partial G^*}{\partial \theta_l} Z^* V^{-1} \frac{\partial R^*}{\partial \theta_j} \right) \right\}_{i=1,j=1}^n \right]^{-1}.
\]

Therefore, the information matrix for the covariance estimate is

\[
M(\delta) = \frac{1}{2} \left[ \left\{ \text{tr} \left( V^{-1} Z^* \frac{\partial G^*}{\partial \theta_l} Z^* V^{-1} \frac{\partial G^*}{\partial \theta_j} Z^* \right) \right\}_{i,j=1}^m \left\{ \text{tr} \left( V^{-1} Z^* \frac{\partial G^*}{\partial \theta_l} Z^* V^{-1} \frac{\partial R^*}{\partial \theta_j} \right) \right\}_{i=1,j=1}^n \right].
\]

To obtain the covariance of the endogenous and exogenous parameter estimates and the covariance estimates, \( \frac{\partial \mathcal{L}}{\partial \delta} = -\frac{1}{2} + 2 \left[ \frac{\partial}{\partial \delta} (y^* - W^* \delta^*) \right] V^{-1} \left( y^* - W^* \delta^* \right) \). Then, \( \frac{\partial \mathcal{L}}{\partial \delta^*} = -W^* V^{-1} \left( y^* - W^* \delta^* \right) \). Next, obtain the second partial derivative, which gives us \( \frac{\partial^2 \mathcal{L}}{\partial \delta^* \partial \theta} = \).
\[ W^* V^{-1} \frac{\partial^2 \nu}{\partial \theta^2} V^{-1} (y^* - W^* \delta^*) \]

Since \( W^* \) is not fixed, then \(-E \left( \frac{\partial^2 L}{\partial \delta^* \partial \theta} \right) \neq 0 \) as in the univariate case. The next objective is to obtain the correlation among these estimates for Example #4 above.

Obtaining the second partial derivative, \(- \frac{\partial^2 L}{\partial \delta^* \partial \theta} = W^{**} V^{-1} \frac{\partial \nu}{\partial \theta} V^{-1} (y^* - W^* \delta^*) \). Substituting for the value \( W^* \), then

\[
- \frac{\partial^2 L}{\partial \delta^* \partial \theta} = \begin{bmatrix} x_1' & 0 & 0 \\ 0 & x_2' & y_2' \\ G_{11} & G_{12} & \begin{bmatrix} y_1' \\ y_2' \\ y_2' \end{bmatrix} - \begin{bmatrix} x_1' \\ x_2' \\ y_1' \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \end{bmatrix}. 
\]

Simplify the right-hand side of the previous equation to obtain

\[
- \frac{\partial^2 L}{\partial \delta^* \partial \theta} = \begin{bmatrix} x_1' G_{11} & x_1' G_{12} \\ x_2' G_{21} & x_2' G_{22} \\ y_1' G_{12} & y_1' G_{22} \end{bmatrix} \begin{bmatrix} y_1' - y_{11} x_1' \\ y_2' - y_{11} x_1' \end{bmatrix} - \begin{bmatrix} x_1' G_{11} & x_1' G_{12} \\ x_2' G_{21} & x_2' G_{22} \\ y_1' G_{12} & y_1' G_{22} \end{bmatrix} \begin{bmatrix} y_1' - y_{11} x_1' \\ y_2' - y_{11} x_1' \end{bmatrix}.
\]

Now multiply the right-hand side of the equation to obtain

\[
- \frac{\partial^2 L}{\partial \delta^* \partial \theta} = \begin{bmatrix} x_1' G_{11} & x_1' G_{12} \\ x_2' G_{21} & x_2' G_{22} \\ y_1' G_{12} & y_1' G_{22} \end{bmatrix} \begin{bmatrix} y_1' - y_{11} x_1' \\ y_2' - y_{11} x_1' \end{bmatrix} - \begin{bmatrix} x_1' G_{11} & x_1' G_{12} \\ x_2' G_{21} & x_2' G_{22} \\ y_1' G_{12} & y_1' G_{22} \end{bmatrix} \begin{bmatrix} y_1' - y_{11} x_1' \\ y_2' - y_{11} x_1' \end{bmatrix}.
\]

Then,

\[
- \frac{\partial^2 L}{\partial \delta^* \partial \theta} = \begin{bmatrix} x_1' G_{11} (y_1' - y_{11} x_1') + x_1' G_{12} (y_2' - y_{11} x_1') - b_{12} y_1' \\ x_2' G_{21} (y_1' - y_{11} x_1') + x_2' G_{22} (y_2' - y_{11} x_1') - b_{12} y_1' \end{bmatrix}.
\]

Taking the expectations gives

\[
-E \left( \frac{\partial^2 L}{\partial \delta^* \partial \theta} \right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} E \left( y_1' G_{12} y_1 - y_1' G_{12} y_{11} x_1 + y_1' G_{22} y_2 - y_1' G_{22} y_{11} x_1 + y_1' G_{22} y_{22} x_2 - y_1' G_{22} y_{22} x_2 - y_1' G_{22} b_{12} y_1 \right) \end{bmatrix}. 
\]
Let $V_{ij} = \text{cov}(Z_u(i) + \varepsilon_{(i)}, Z_u(j) + \varepsilon_{(j)})$ and $\hat{V}_{ij} = \text{cov}(y_{(i)}, y_{(j)})$. Simplify the last element of the previous matrix to obtain

$$E \left( y'_{(1)} G_{12} y_{(1)} - y'_{(1)} G_{12} y_{(1)} x_{(1)} + y'_{(1)} G_{22} y_{(2)} - y'_{(1)} G_{22} y_{(2)} x_{(2)} - y'_{(1)} G_{22} b_{12} y_{(1)} \right)$$

$$= \text{tr}(V_{11} G_{12}) + E \left( y'_{(1)} G_{12} E (y_{(3)}) - E \left( y'_{(1)} G_{12} y_{(3)} x_{(1)} + \text{tr}(V_{12} G_{22}) + E \left( y'_{(1)} G_{22} y_{(2)} - E \left( y'_{(1)} G_{22} y_{(2)} x_{(2)} - \text{tr}(b_{12} \hat{V}_{11} G_{22}) - E \left( y'_{(1)} G_{22} b_{12} E (y_{(3)}) \right) \right) \right) \right)$$

$$= \hat{V}_{11} G_{12} + \hat{V}_{12} G_{22} - b_{12} \hat{V}_{11} G_{22}.$$ 

Therefore,

$$-E \left( \frac{\partial^2 L}{\partial \hat{\alpha}^2 \hat{\theta}} \right) = \begin{bmatrix} 0 & 0 \\ \text{tr}(V_{11} G_{12} + \hat{V}_{12} G_{22} - b_{12} V_{11} G_{22}) & \end{bmatrix}$$

where $\hat{V}_{11} = V_{11}$ since $y_{(1)}$ has only exogenous variables and

$$\hat{V}_{12} = \text{cov}(y_{(1)}, y_{(2)}) = \text{cov}(y_{(1)}, b_{12} y_{(1)} + y_{22} x_{(2)} + Z u_{(1)} + \varepsilon_{(1)})$$

$$= \text{cov}(y_{(1)}, b_{12} y_{(1)}) + \text{cov}(y_{(1)}, y_{22} x_{(2)}) + \text{cov}(y_{(1)}, Z u_{(1)}) + \text{cov}(y_{(1)}, \varepsilon_{(1)})$$

$$= b_{12} V_{11} + 0 + \text{cov}(y_{11} x_{(1)} + u_{(1)} + \varepsilon_{(1), u_{(1)}}) + \text{cov}(y_{11} x_{(1)} + u_{(1)} + \varepsilon_{(1), x_{(1)}})$$

$$= b_{12} V_{11} + \text{cov}(Z u_{(1)}, Z u_{(2)}) + \text{cov}(\varepsilon_{(1), \varepsilon_{(2))})$$

$$= b_{12} V_{11} + Z \sigma_{u_{12}}^2 Z' + \sigma_{\varepsilon_{12}}^2 I_n$$

$$= b_{12} V_{11} + V_{12}.$$
Thus,

\[ tr(V_{11}G_{12} + V_{12}G_{22} - b_{12}V_{11}G_{22}) = tr(V_{11}G_{12} + [b_{12}V_{11} + V_{12}]G_{22} - b_{12}V_{11}G_{22}) \]

\[ = tr(V_{11}G_{12} + V_{12}G_{22}). \]

Next, the expected value is simplified to

\[ -E \left( \frac{\partial^2 L}{\partial \delta^2} \right) = \begin{bmatrix} 0 \\ 0 \\ tr(V_{11}G_{12} + V_{12}G_{22}) \end{bmatrix} \]

where \( G_1 = V^{-1} \frac{\partial V}{\partial \sigma^2} V^{-1} = V^{-1} \left( \begin{bmatrix} \sigma_1^2 \\ 0 \\ 0 \end{bmatrix} \otimes I \right) V^{-1} = \begin{bmatrix} G_{111} \\ G_{121} \\ G_{122} \end{bmatrix} \). Therefore,

\[ -E \left( \frac{\partial^2 L}{\partial \delta^2} \right) = \begin{bmatrix} 0 \\ 0 \\ tr(V_{11}G_{111} + V_{12}G_{122}) \end{bmatrix}, \]

assuming that

\[
G^* = \begin{bmatrix}
\sigma_{u_1}^2 l_b & \sigma_{u_{12}} l_b \\
\sigma_{u_{12}} l_b & \sigma_{u_2}^2 l_b
\end{bmatrix} = \begin{bmatrix}
\sigma_{u_1}^2 & \sigma_{u_{12}} \\
\sigma_{u_{12}} & \sigma_{u_2}^2
\end{bmatrix} \otimes l_b \quad \text{and} \quad R^* = \begin{bmatrix}
\sigma_1^2 l & \sigma_{12} l \\
\sigma_{12} l & \sigma_2^2 l
\end{bmatrix} = \begin{bmatrix}
\sigma_1^2 & \sigma_{12} \\
\sigma_{12} & \sigma_2^2
\end{bmatrix} \otimes l.
\]

The full information matrix for Example #4 can be found in Appendix 4A. The optimal design for the previous example based on the new criteria is shown in Table 4.4.

**Table 4.4:** An ML optimal design for four blocks with block size equal to two.

<table>
<thead>
<tr>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 )</td>
<td>( x_2 )</td>
<td>( x_1 )</td>
<td>( x_2 )</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

The determinant of the information matrix is equal to 898,636,208.

The previous optimal design is not dependent on \( \alpha \), however, the ML estimate for the variance components and the correlations of the estimates of the endogenous parameters and the
covariance parameters does not depend on the design $X$. For that reason, REML estimators will be considered.

4.4 Residual Maximum Likelihood (REML) Simultaneous Optimal Design for a Causal Structure for the Endogenous Parameters, Exogenous Parameters, and Variance Components

Starting with notations similar to those of Durbin (1988) and as was discussed in Chapter 2, rewrite the model in order to use REML estimators, giving

$$YB + X\Gamma = ZU + E.$$ 

By transposing both sides, $B'Y' + \Gamma'X' = U'Z' + E'$, which is equivalent to $B'Y'I + IT'X' = IU'Z' + E'$. Using the $\text{Vec}$ operator,

$$\text{Vec}(B'Y'I) + \text{Vec}(IT'X') = \text{Vec}(IU'Z') + \text{Vec}(E').$$

But according to the $\text{Vec}$ operator properties

$$\text{Vec}(B'Y'I) = (I \otimes B')\text{Vec}(Y'), \text{Vec}(IT'X') = (X \otimes I)\text{Vec}(\Gamma'),$$

and

$$\text{Vec}(IU'Z') = (Z \otimes I)\text{Vec}(U').$$

Therefore,

$$(I \otimes B')\text{Vec}(Y') + (X \otimes I)\text{Vec}(\Gamma') = (Z \otimes I)\text{Vec}(U') + \text{Vec}(E').$$

Next,

$$\text{Vec}(Y') + (I \otimes B')^{-1}(X \otimes I)\text{Vec}(\Gamma') = (I \otimes B')^{-1}[(Z \otimes I)\text{Vec}(U') + \text{Vec}(E')]$$

and

$$\text{Vec}(Y') + (I \otimes [B']^{-1})(X \otimes I)\text{Vec}(\Gamma') = (I \otimes B')^{-1}[(Z \otimes I)\text{Vec}(U') + \text{Vec}(E')]$$

But using the Kronecker product properties

$$(I \otimes [B']^{-1})(X \otimes I) = (X \otimes [B']^{-1}).$$
Therefore,

$$\text{Vec}(Y') + (X \otimes [B']^{-1})\text{Vec}(\Gamma') = (I \otimes B')^{-1}[(Z \otimes I)\text{Vec}(U') + \text{Vec}(E')]$$.

Next,

$$(X \otimes [B']^{-1}) = (XI \otimes [B']^{-1})$$

$$= (X \otimes I)(I \otimes [B']^{-1})$$.

From the previous results, we conclude that

$$\text{Vec}(Y') \underbrace{=}_{y^*} - (X \otimes I)(I \otimes B')^{-1} \text{Vec}(\Gamma') + (I \otimes B')^{-1} \left[ (Z \otimes I) \text{Vec}(U') + \text{Vec}(E') \right]$$,

which can be rewritten as

$$y^* = -\tilde{X}^* \tilde{B}^* \tilde{y}^* + \tilde{B}^* \tilde{Z}^* \tilde{u}^* + \tilde{\varepsilon}^*$$.

Let $\text{Var}[\text{Vec}(U')] = I_b \otimes \Sigma_u$ and $\text{Var}[\text{Vec}(E')] = I \otimes \Sigma$ (See the justification in Appendix 4B). Therefore,

$$\begin{pmatrix} u^* \\ \varepsilon^* \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} I_b \otimes \Sigma_u & 0 \\ 0 & I \otimes \Sigma \end{pmatrix} \right) = N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tilde{G}^* & 0 \\ 0 & \tilde{R}^* \end{pmatrix} \right)$$,

where $\tilde{G}^* = I_b \otimes \Sigma_u$, $\tilde{R}^* = I \otimes \Sigma$, $\Sigma_u = \begin{pmatrix} \sigma_{u_{11}} & \sigma_{u_{12}} & \cdots & \sigma_{u_{1p}} \\ \sigma_{u_{12}} & \sigma_{u_{22}} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{u_{1p}} & \vdots & \cdots & \sigma_{u_{pp}} \end{pmatrix}$ and $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \vdots & \cdots & \sigma_{pp} \end{pmatrix}$.

Since $E \left( \tilde{y}^* \right) = -\tilde{X}^* \tilde{B}^* \tilde{y}^*$ and $V = \text{Var}(\tilde{y}^*) = \text{Var}(\tilde{B}^* \tilde{Z}^* \tilde{u}^* + \tilde{\varepsilon}^*)$, then $V = \tilde{B}^* \tilde{Z}^* \tilde{G}^* \tilde{Z}^* \tilde{u}^* + \tilde{\varepsilon}^*$.

Therefore, the likelihood function is
\[
L(\mathcal{V}, \kappa^*|\tilde{\mathcal{V}}^*, X^*) = (2\pi)^{-\frac{np}{2}} |\mathcal{V}|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \left( \tilde{\mathcal{V}}^* + X^* B^* \mathcal{Y}^* \right)' V^{-1} \left( \tilde{\mathcal{V}}^* + X^* B^* \mathcal{Y}^* \right) \right].
\]

Pick \( \tilde{\kappa} \) such that \( \tilde{\kappa}' X^* = 0 \). Then, \( \tilde{\kappa}' = \mathcal{E}' \left[ I - X^* (R^* X^*)^{-1} X^* \right] \). Next, \( \tilde{\kappa}' \tilde{\mathcal{Y}}^* = -\mathcal{E}' X^* B^* \mathcal{Y}^* + \mathcal{E}' [Z^* \mathcal{Y}^* + \mathcal{E}^*] \). Therefore, the expected value and variance are \( E \left( \tilde{\kappa}' \tilde{\mathcal{Y}}^* \right) = 0 \) and \( Var \left( \tilde{\kappa}' \tilde{\mathcal{Y}}^* \right) = \mathcal{E}' \mathcal{Y}^* \mathcal{E}^* \sim N(0, \tilde{\kappa}' \tilde{\mathcal{Y}}^* \mathcal{E}^*) \) and the log likelihood is \( L(\mathcal{V} | \tilde{\mathcal{Y}}^*, \tilde{\kappa}) = \frac{-np}{2} \log(2\pi) - \frac{1}{2} \log \left| \mathcal{V} \right| + \frac{1}{2} \left( \mathcal{E}' \mathcal{Y}^* \mathcal{E}^* \right)' \left( \mathcal{E}' \mathcal{Y}^* \mathcal{E}^* \right)^{-1} \left( \mathcal{E}' \mathcal{Y}^* \mathcal{E}^* \right) \). Assume that \( \theta \) is one of the elements of \( \mathcal{V} \), then \( \theta \) is one of the parameters of \( R^* \) using the previous notations, either \( \theta = \sigma^2 \) or \( \sigma_{ij} \), or \( \theta \) is one of the parameters of \( G^* \) using the previous notations \( \theta = \sigma^2_{ui} \) or \( \sigma_{uij} \). Then, the partial derivative of \( \log |\mathcal{V}| \) is \( \frac{\partial \log |\mathcal{V}|}{\partial \theta} = \text{tr}(\mathcal{V}^{-1} \frac{\partial \mathcal{V}}{\partial \theta}) \) and \( \frac{\partial \mathcal{V}}{\partial \theta} = -\mathcal{V}^{-1} \frac{\partial \mathcal{V}}{\partial \theta} \mathcal{V}^{-1} \), thus \( \frac{\partial L}{\partial \theta} = -\frac{1}{2} \frac{\partial \log |\mathcal{V}|}{\partial \theta} - \left( \mathcal{E}' \mathcal{Y}^* \mathcal{E}^* \right)' \left( \mathcal{E}' \mathcal{Y}^* \mathcal{E}^* \right)^{-1} \left( \mathcal{E}' \mathcal{Y}^* \mathcal{E}^* \right) \). Next, \( \frac{\partial L}{\partial \theta_k} = \frac{1}{2} \text{tr} \left( \mathcal{E}' \mathcal{Y}^* \mathcal{E}^* \right)^{-1} \mathcal{E}' \frac{\partial \mathcal{V}}{\partial \theta_k} \mathcal{E}^*. \) Setting the previous equation equal to zero, this results in
\[
\text{tr} \left( \mathcal{E}' \mathcal{Y}^* \mathcal{E}^* \right)^{-1} \mathcal{E}' \frac{\partial \mathcal{V}}{\partial \theta_k} \mathcal{E}^* = \tilde{\mathcal{Y}}^* \mathcal{E}' \left( \mathcal{E}' \mathcal{Y}^* \mathcal{E}^* \right)^{-1} \left( \mathcal{E}' \mathcal{Y}^* \mathcal{E}^* \right)^{-1} \mathcal{E}' \tilde{Y}^*. \tag{20}
\]

Let \( \hat{\theta} \) be the optimal solution for the likelihood function. Also, let \( \hat{\mathcal{V}} = \mathcal{V} |_{\theta = \hat{\theta}} \) let \( \hat{P} = \mathcal{E}' \mathcal{Y}^* \mathcal{E}^* \mathcal{E}' \mathcal{Y}^* \mathcal{E}^* \mathcal{E}' \mathcal{Y}^* \mathcal{E}^* \) and \( \hat{\kappa} = \mathcal{E}' \mathcal{Y}^* \mathcal{E}^* \mathcal{E}' \mathcal{Y}^* \mathcal{E}^* \mathcal{E}' \mathcal{Y}^* \mathcal{E}^* \). Then
\[
\text{tr} \left( \hat{\mathcal{V}} |_{\theta = \hat{\theta}} \frac{\partial \mathcal{V}}{\partial \theta_k} \right) = \tilde{\mathcal{Y}}^* \mathcal{E}' \left( \mathcal{E}' \mathcal{Y}^* \mathcal{E}^* \right)^{-1} \left( \mathcal{E}' \mathcal{Y}^* \mathcal{E}^* \right)^{-1} \mathcal{E}' \tilde{\mathcal{V}}^*. \tag{21}
\]

If \( \theta_k \) belongs to \( \tilde{R}^* \), then
\[
\frac{\partial \mathcal{V}}{\partial \theta_k} = \mathcal{E}' \frac{\partial \mathcal{V}}{\partial \theta_k} \mathcal{E}^* \tag{22}
\]
Under these conditions, Eq. 20 above would become

\[
tr\left(\hat{Z}'\hat{B}'\hat{\mathbf{P}}\hat{B}'\hat{Z}'\frac{\partial \hat{G}^*}{\partial \theta_k}\bigg| \theta = \hat{\theta}\right) = \frac{\bar{Z}'\hat{B}'\hat{Z}'\frac{\partial \hat{G}^*}{\partial \theta_k}}{\theta = \hat{\theta}} \hat{Z}'\hat{B}'\hat{\mathbf{P}}\hat{Z}'
\]

and if \( \theta_k \) belongs to \( \hat{R}^* \) or \( \theta = \sigma_i^2 \) or \( \sigma_\ell \), then

\[
tr\left(\hat{B}'\frac{\partial \hat{R}}{\partial \theta_k}\bigg| \theta = \hat{\theta}\right) = \frac{\bar{Z}'\hat{B}'\hat{\mathbf{P}}\hat{Z}'}{\theta = \hat{\theta}}. \quad \text{Since \( b_i \) is one of the parameters of \( \hat{B} \) where \( \hat{B} = (I \otimes \hat{B}')^{-1} \), then}
\]

\[
\bar{V} = \hat{B} [\hat{Z}' \hat{G}' \hat{Z}'' + \hat{R}'] (I \otimes \hat{B}^{-1})^{-1} \bar{V} = (I \otimes \hat{B}')^{-1} [\hat{Z}' \hat{G}' \hat{Z}'' + \hat{R}'] (I \otimes \hat{B})^{-1} \]

Therefore,

\[
\frac{\partial \bar{V}}{\partial b_i} = - (I \otimes \hat{B}')^{-1} \left( I \otimes \frac{\partial \hat{B}'}{\partial b_i} \right) \bar{V} = (I \otimes \hat{B}')^{-1} [\hat{Z}' \hat{G}' \hat{Z}'' + \hat{R}'] (I \otimes \hat{B}^{-1})^{-1}
\]

\[
= (I \otimes \hat{B}')^{-1} [\hat{Z}' \hat{G}' \hat{Z}'' + \hat{R}'] (I \otimes \hat{B})^{-1} \left( I \otimes \frac{\partial \hat{B}'}{\partial b_i} \right) (I \otimes \hat{B})^{-1}.
\]

The second partial derivative is

\[
\frac{\partial^2 \mathcal{L}}{\partial b_i \partial \theta_k} = \frac{1}{2} tr \left\{ \frac{\partial}{\partial b_i} \left( \frac{\partial \mathcal{L}}{\partial \theta_k} \right) - \frac{\partial^2 \bar{V}}{\partial b_i \partial \theta_k} \right\}.
\]

Next,

\[
\frac{\partial^2 \mathcal{L}}{\partial b_i \partial \theta_k} = \frac{1}{2} tr \left\{ \frac{\partial}{\partial b_i} \left( \frac{\partial \mathcal{L}}{\partial \theta_k} \right) - \frac{\partial^2 \bar{V}}{\partial b_i \partial \theta_k} \right\} + \left( \frac{\partial \bar{V}}{\partial b_i} \right) \frac{\partial \bar{V}}{\partial \theta_k} - \left( \frac{\partial \bar{V}}{\partial b_i} \right) \frac{\partial \bar{V}}{\partial \theta_k} - \left( \frac{\partial \bar{V}}{\partial b_i} \right) \frac{\partial \bar{V}}{\partial \theta_k}.
\]
Simplifying the partial derivative further,

\[
\frac{\partial^2 L}{\partial b_i \partial \theta_k} = \frac{1}{2} \text{tr} \left \{ \left[ \kappa' \tilde{V} \tilde{K} \right]^{-1} \frac{\partial \tilde{V}}{\partial b_i} \left[ \kappa' \tilde{V} \tilde{K} \right]^{-1} \kappa' \frac{\partial \tilde{V}}{\partial \theta_k} \tilde{K} - \left[ \kappa' \tilde{V} \tilde{K} \right]^{-1} \kappa' \frac{\partial^2 \tilde{V}}{\partial b_i \partial \theta_k} \tilde{K} \right \}
\]

\[
+ \left \{ -\left[ \kappa' \tilde{V} \tilde{K} \right]^{-1} \frac{\partial \tilde{V}}{\partial b_i} \left( \kappa' \tilde{V} \tilde{K} \right)^{-1} \kappa' \frac{\partial \tilde{V}}{\partial \theta_k} \left( \kappa' \tilde{V} \tilde{K} \right)^{-1} \right \} \left( \kappa' \tilde{V}' \right) \left( \kappa' \tilde{V}' \right)^t.
\]

Since \( E \left[ \left( \kappa' \tilde{V}' \right) \left( \kappa' \tilde{V}' \right)^t \right] = \kappa' \tilde{V} \tilde{K} \), then

\[
E \left( -\frac{\partial^2 L}{\partial b_i \partial \theta_k} \right) = \frac{1}{2} \text{tr} \left \{ -\left[ \kappa' \tilde{V} \tilde{K} \right]^{-1} \frac{\partial \tilde{V}}{\partial b_i} \left[ \kappa' \tilde{V} \tilde{K} \right]^{-1} \kappa' \frac{\partial \tilde{V}}{\partial \theta_k} \tilde{K} + \left[ \kappa' \tilde{V} \tilde{K} \right]^{-1} \kappa' \frac{\partial^2 \tilde{V}}{\partial b_i \partial \theta_k} \tilde{K} \right \}
\]

\[
+ \left \{ -\left[ \kappa' \tilde{V} \tilde{K} \right]^{-1} \frac{\partial \tilde{V}}{\partial b_i} \left( \kappa' \tilde{V} \tilde{K} \right)^{-1} \kappa' \frac{\partial \tilde{V}}{\partial \theta_k} \left( \kappa' \tilde{V} \tilde{K} \right)^{-1} \right \} \left( \kappa' \tilde{V}' \right) \left( \kappa' \tilde{V}' \right)^t.
\]

Next, \( E \left( -\frac{\partial^2 L}{\partial b_i \partial \theta_k} \right) = \frac{1}{2} \text{tr} \left \{ \bar{p} \frac{\partial \tilde{V}}{\partial \theta_k} \bar{p} \frac{\partial \tilde{V}}{\partial b_i} \right \} \). Therefore,

\[
E \left( -\frac{\partial^2 L}{\partial b_i \partial \theta_k} \right) = \frac{1}{2} \text{tr} \left \{ \bar{p} \frac{\partial \tilde{V}}{\partial \theta_k} \bar{p} \frac{\partial \tilde{V}}{\partial b_i} \right \}.
\]

Also,

\[
\frac{\partial^2 L}{\partial \theta_i \partial \theta_k} = \frac{1}{2} \text{tr} \left \{ \frac{\partial}{\partial \theta_i} \left[ \left( \kappa' \tilde{V} \tilde{K} \right)^{-1} \kappa' \frac{\partial \tilde{V}}{\partial \theta_k} \tilde{K} \right] - \left[ \left( \kappa' \tilde{V} \tilde{K} \right)^{-1} \kappa' \frac{\partial \tilde{V}}{\partial \theta_k} \tilde{K} \right] \frac{\partial}{\partial \theta_i} \left( \kappa' \tilde{V} \tilde{K} \right)^{-1} \right \} \left( \kappa' \tilde{V}' \right) \left( \kappa' \tilde{V}' \right)^t
\]

\[
= \frac{1}{2} \text{tr} \left \{ \left[ \kappa' \tilde{V} \tilde{K} \right]^{-1} \kappa' \tilde{V}_k \tilde{K} \left[ \kappa' \tilde{V} \tilde{K} \right]^{-1} \kappa' \tilde{V}_k \tilde{K} - \left[ \kappa' \tilde{V} \tilde{K} \right]^{-1} \kappa' \tilde{V}_k \tilde{K} \right \}
\]

\[
+ \left \{ \left[ \kappa' \tilde{V} \tilde{K} \right]^{-1} \kappa' \tilde{V}_k \tilde{K} \left[ \kappa' \tilde{V} \tilde{K} \right]^{-1} \kappa' \tilde{V}_k \tilde{K} \right \} \left( \kappa' \tilde{V}' \right) \left( \kappa' \tilde{V}' \right)^t.
\]
Similarly, since \( E \left[ \begin{pmatrix} \bar{r}' \bar{y}' \end{pmatrix} \begin{pmatrix} \bar{r}' \bar{y}' \end{pmatrix} \right] = \bar{r}' \bar{V} \bar{r} \), then

\[
E \left( \frac{\partial^2 \mathcal{L}}{\partial \theta_i \partial \theta_k} \right) = \frac{1}{2} \text{tr} \left\{ \left[ \bar{r}' \bar{V} \bar{r} \right]^{-1} \bar{r}' \bar{V}_i \bar{r} \left[ \bar{r}' \bar{V} \bar{r} \right]^{-1} \bar{r}' \bar{V}_k \bar{r} - \left[ \bar{r}' \bar{V} \bar{r} \right]^{-1} \bar{r}' \bar{V}_{ik} \bar{r} \right\} - \left[ \bar{r}' \bar{V} \bar{r} \right]^{-1} \bar{r}' \bar{V}_i \bar{r} \left[ \bar{r}' \bar{V} \bar{r} \right]^{-1} \bar{r}' \bar{V}_k \bar{r} + \left[ \bar{r}' \bar{V} \bar{r} \right]^{-1} \bar{r}' \bar{V}_k \bar{r} \left[ \bar{r}' \bar{V} \bar{r} \right]^{-1} \bar{r}' \bar{V}_i \bar{r} \right\}
\]

By simplifying the right-hand side,

\[
E \left( \frac{\partial^2 \mathcal{L}}{\partial \theta_i \partial \theta_k} \right) = -\frac{1}{2} \text{tr} \left\{ \bar{r} \left[ \bar{r}' \bar{V} \bar{r} \right]^{-1} \bar{r}' \bar{V}_i \bar{r} \right\}, \text{ which leads to } \]

\[
E \left( \frac{\partial^2 \mathcal{L}}{\partial \theta_i \partial \theta_k} \right) = -\frac{1}{2} \text{tr} \left\{ \bar{P} \bar{V}_k \bar{P} \bar{V}_l \right\} \tag{23}
\]

where \( P = \bar{r} \left[ \bar{r}' \bar{V} \bar{r} \right]^{-1} \bar{r}' = \bar{V}^{-1} - \bar{V}^{-1} \bar{X}' (\bar{X}^* \bar{V}^{-1} \bar{X}^*) \bar{X}^* \bar{V}^{-1} \) and \( \bar{V} = \bar{B}' \left[ \bar{Z}^* \bar{G}^* \bar{Z}'^* + \bar{R}^* \right] \bar{B}' \).

Given Eq. 21, Eq. 22, and Eq. 23, the asymptotic information matrix for the variance components is

\[
M(\hat{\sigma}) = \begin{bmatrix}
\begin{array}{c}
\nu_d \\
\nu_r
\end{array}
\end{bmatrix}
\begin{bmatrix}
\frac{1}{2} \left\{ \sum_{m=1}^{l'} \text{tr} \left( \bar{P}^* \bar{B}' \bar{Z}' \frac{\partial \bar{G}^*}{\partial \theta_{di}} \bar{Z}'' \bar{B}' \bar{Z}' \right) \right\} + \frac{1}{2} \left\{ \sum_{m=1}^{d'} \text{tr} \left( \bar{P}^* \bar{B}' \bar{Z}' \frac{\partial \bar{G}^*}{\partial \theta_{rij}} \bar{Z}'' \bar{B}' \bar{Z}' \right) \right\} & \frac{1}{2} \left\{ \sum_{m=1}^{l'} \text{tr} \left( \bar{P}^* \bar{B}' \bar{Z}' \frac{\partial \bar{R}^*}{\partial \theta_{ij}} \bar{Z}'' \bar{B}' \bar{Z}' \right) \right\} \\
\frac{1}{2} \left\{ \sum_{m=1}^{d'} \text{tr} \left( \bar{P}^* \bar{B}' \bar{Z}' \frac{\partial \bar{R}^*}{\partial \theta_{rij}} \bar{Z}'' \bar{B}' \bar{Z}' \right) \right\} & \frac{1}{2} \left\{ \sum_{m=1}^{l'} \text{tr} \left( \bar{P}^* \bar{B}' \bar{Z}' \frac{\partial \bar{R}^*}{\partial \theta_{ij}} \bar{Z}'' \bar{B}' \bar{Z}' \right) \right\}
\end{bmatrix}
\]

where \( \bar{P}^* = \bar{V}^{-1} - \bar{V}^{-1} \bar{X}' (\bar{X}^* \bar{V}^{-1} \bar{X}^*)^{-1} \bar{X}^* \bar{V}^{-1} \), \( \bar{V} = \bar{B}' \left[ \bar{Z}^* \bar{G}^* \bar{Z}'^* + \bar{R}^* \right] \bar{B}' \), \( \bar{Z}^* = Z_{n \times b} \otimes I_b \), \( \bar{G}^* = I_b \otimes \Sigma_u \), and \( \bar{R}^* = I \otimes \Sigma \).

In the next example, we will use the REML information matrix and the Modified Fedorov algorithm to obtain an optimal design for a causal structure from Example #3.
Example #5: Using the new REML estimators for the variance components, Example #3 will be redone using the new information matrix. Assume there are four blocks with size two each.

Assume that \( \sigma_{12} \neq 0 \) and \( \sigma_{u12} \neq 0 \).

Let \( G^* = I_b \otimes \begin{bmatrix} \sigma_{u11}^2 & \sigma_{u12} \\ \sigma_{u12} & \sigma_{u22}^2 \end{bmatrix} \) and \( R^* = I \otimes \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \), then the first partial derivatives are

\[
\frac{\partial G^*}{\partial \sigma_{u11}^2} = I_b \otimes \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},
\]

\[
\frac{\partial R^*}{\partial \sigma_1^2} = I \otimes \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},
\]

\[
\frac{\partial G^*}{\partial \sigma_{u22}^2} = I_b \otimes \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},
\]

\[
\frac{\partial R^*}{\partial \sigma_2^2} = I \otimes \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},
\]

\[
\frac{\partial G^*}{\partial \sigma_{u12}} = I_b \otimes \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},
\]

\[
\frac{\partial R^*}{\partial \sigma_{12}} = I \otimes \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},
\]

The information matrix is shown in Appendix 4C. The optimal design for the previous example based on the new criteria is shown in Table 4.5.

**Table 4.5:** A REML optimal design for four blocks with block size equal to two.

<table>
<thead>
<tr>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>x_1</td>
<td>x_2</td>
<td>x_1</td>
<td>x_2</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>x_1</td>
<td>x_2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>x_1</td>
<td>x_2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

The determinant of the information matrix is equal to \( 1.3891 \times 10^{12} \).
The REML optimal design is dramatically different than the optimal design for the univariate
case using composite criteria that was obtained by Mylona, Goos, and Jones (2014) as is shown in
Table 4.6.

Table 4.6: The univariate optimal design for four blocks with block size equal to two using
composite criteria.

<table>
<thead>
<tr>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_1$</td>
<td>$x_2$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

The results in Table 4.7 show that the univariate optimal design performs worse than the REML
optimal design.

Table 4.7: Comparison of the determinant of the REML information matrix using different
designs.

<table>
<thead>
<tr>
<th>Design</th>
<th>Determinant of the REML Information Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>REML Optimal Design for a Causal Structure</td>
<td>$1.3891 \times 10^{12}$</td>
</tr>
<tr>
<td>ML Optimal Design for a Causal Structure</td>
<td>$2.6824 \times 10^{11}$</td>
</tr>
<tr>
<td>Univariate Optimal Design</td>
<td>$1.4862 \times 10^{11}$</td>
</tr>
</tbody>
</table>

The REML optimal design is \( \left( \frac{1.3891 \times 10^{12}}{1.4862 \times 10^{11}} \right)^{\frac{1}{5}} \approx 128\% \) as \( D \)-efficient as the univariate optimal
design, which means that it is about 28\% more \( D \)-efficient than the univariate optimal design.
This is expected since the univariate optimal design takes neither the endogenous parameters nor
the covariance structure among the endogenous variables into account.

The REML optimal design is \( \left( \frac{1.3891 \times 10^{12}}{2.6824 \times 10^{11}} \right)^{\frac{1}{5}} \approx 120\% \) as \( D \)-efficient as the ML optimal design,
which means that it is about 20\% more \( D \)-efficient than the ML optimal design. The
improvement comes from the fact that the ML variance components information matrix does not
depend on the design as demonstrated in Section 4.3. The improvement in the REML optimal
design takes into account the endogenous and exogenous parameters, and the variance components, as well as the correlation among the endogenous parameters estimates and the variance components estimates.

4.5 Summary

One of the weaknesses of optimal designs for endogenous and exogenous parameters is that they do not take into account the precision of the variance components estimates. To estimate the endogenous and exogenous parameters, the variance components are needed. Not taking into account the precision of the estimates of the variance components may lead to poor estimates and poor inferential statistics. Similar to univariate optimal design, our objective was to obtain an optimal design for a causal structure for the endogenous and exogenous parameters and the covariance components.

Three criteria for the optimal design for variance components in a random mixed causal structure model were developed. The 3SLS simultaneous optimal design assumes that $W^*$ is fixed and the two FIML simultaneous optimal designs do not. There was a dramatic difference between whether $W^*$ is fixed or not, which was different from the criteria that were developed for optimal design for endogenous and exogenous parameters only.

The 3SLS simultaneous optimal design assumes that $W^*$ is fixed by replacing $W^*$ with $\hat{W}^*$. The endogenous and exogenous parameters estimates are independent from the covariance parameter estimates, which allows for the partition of the information matrix and for composite criteria to be used as in univariate optimal design. This approach requires that an $\alpha$ is chosen where $\alpha$ is a real number between 0 and 1. The choice of $\alpha$ can change the optimal design as was demonstrated in Section 4.2. One of the consequences of the 3SLS approach in assuming that $W^*$ is fixed is that there is no correlation between the endogenous parameters and the covariance parameters estimates.
However, $W^*$ includes endogenous variables and is not fixed. This motivated the need to develop two simultaneous optimal designs using ML and REML estimators for the variance components. These approaches have three advantages over 3SLS simultaneous optimal design. First, the approach is more accurate because it assumes that $W^*$ is not fixed. A second advantage is that $\alpha$ is not included, which makes the optimal design not subjective to change. Third, the ML and REML approaches allow the correlation between the endogenous parameters estimates and the covariance estimates to be included in the information matrix.

One of the advantages of REML estimators for the covariance components and for the correlation between the variance components is that its information matrix depends on the design matrix. However, this is not the case with ML estimators. The ML information matrix of the covariance estimates and the correlation between the endogenous parameters estimates and the covariance estimates do not depend on the design. Therefore, it is recommended to use the REML optimal design (Mylona, Goos, and Jones 2014).

When using the REML optimality criteria to compare the performance of the REML optimal design to the univariate optimal design, results show here that the REML optimal design is about 30% more $D$-efficient than the univariate optimal design. Also, when comparing the performance of the REML optimal design to the ML optimal design, the REML optimal design was about 20% more $D$-efficient. Based on these results, it is recommended to use the REML optimality criteria for an optimal design for a causal structure because this criteria include the information matrix for the endogenous parameters, exogenous parameters, the variance components, and the correlation among these estimates.

4.6 References


Appendix 4A: ML Information Matrix for Endogenous, Exogenous, and Covariance Estimate.
Appendix 4B: REML Covariance Structure.
**Appendix 4C**: REML Information Matrix for Endogenous, Exogenous, and Covariance Estimate.
CHAPTER 5: APPLICATIONS OF OPTIMAL DESIGN FOR A CAUSAL STRUCTURE

5.1 Introduction

The applications of causal structure modeling are widely used in many different fields from epidemiology to agriculture, from education to economics, and many others. Because of the many applications of structural equation modeling, the importance of using optimal designs to produce efficient estimates for the parameters of interest will be demonstrated with two examples from previous studies. The first example is a wheat plant breeding study published in Crop Science (Campbell et al. 2003; Dhungana et al. 2007). The second example comes from a study of durum and bread wheat breeding study conducted in Obregon, Mexico, which was published in Crop Science (Vargas et al. 1998; Vargas et al. 1999) with extension work conducted in dissertation research at the University of Nebraska-Lincoln (Dhungana 2004; Yaseen 2012). The objectives here are to compare the designs that were used in the published papers with the optimal design for causal structure using the techniques that were developed in the previous chapters.

5.2 Optimal Design Application in Relation to Analysis of Genotype-by-Environment Interaction in Wheat Using a Structural Equation Model and Chromosome Substitution Lines

The first application that will be used is an example that was published by Campbell et al (2003) and also reported in Dhungana et al. (2007) where the data are analyzed using a causal structure. The experiment was conducted in Lincoln, NE in 1999, 2000 and 2001 and also in Mead, NE and Sidney, NE in 2000 and 2001 for a total of 7 environments. In 1999, four replicates of randomized complete blocks were used, whereas the other years used incomplete blocks. The genotypes that were used in the study were derived from a cross between cv. Cheyenne (CNN) and the chromosome substitution line CNN (Wichita 3A). Grain yield depends on a number of
supporting traits influenced by many different genes and by a multitude of environmental
conditions at different stages during plant development (Ashikari et al., 2005). The objective of
the study was to understand how gene-environment interactions (GEI) influence complex traits
such as grain yield.

However, the questions that could have been raised before the study was executed include how
the genotypes were chosen and whether or not the genotypes were suitable for the study. More
specifically, how were the genotypes chosen since the objective was to understand the gene-
environment interactions on complex traits? Were these genotypes the most efficient choices to
achieve the objective of the study? These questions would have been important to ask to ensure
that the amount of information obtained from the study were maximized and, ultimately, to meet
the objective of the study.

In order to answer those questions and to obtain the most precise estimates for the endogenous
and exogenous parameters, there are three objectives for this section. First, the genetic
combinations for the ideal genotypes will be obtained which would produce the optimal
estimates. Next, the efficiency of the estimates produced from the optimal genotypes will be
compared to the genotypes that have been used in the paper. Finally, if the ideal genotypes are
unable to be bioengineered or if there are a limited number of genotypes, the combination of the
genotypes that would produce the optimal design from the currently available genotypes will be
obtained.

Since the objective of the study was to model yield gene-environment interaction (GEI) and there
was no interest in the environment or blocks within environments effects, then yield and yield
components residual were used in place of the observed yield and yield components. \( yld_{GEI} \),
\( tkw_{GEI} \), \( kps_{GEI} \), \( spsm_{GEI} \) were used as the measured variables where \( yld \) represents yield, \( tkw \)
represents 1000-kernels weight, \( kps \) represents kernels per spike, and \( spsm \) represents spikes per
square meter. As an example, for the yield, yield\sub{GEI} = yield - \epsilon - b(e) where \epsilon represents the environment effect and b(e) is the block within environment effect. For that reason, the information matrix needs to be adjusted and 3SLS and FIML methodologies will be used to obtain the optimal design. Six DNA marker loci that were closely related to a different quantitative trait loci (QTL) were also selected. The selected markers were Xtam055, Xbarc86, Xbarc67, Xksua6, Xbcd1555, and Xbcd361.

Assume \( y = W_1 \delta_1 + W_2 \delta_2 + \epsilon \). Let \( W_1 \) be exogenous variables. Our objective is to focus on \( \delta_2 \) since it is assumed that \( \delta_1 \) are nuisance parameters. In this example, \( \delta_1 \) are the environment and block effects. Multiply both sides of the equation by \((I - P_{W_1})\) where \( P_{W_1} = W_1(W_1'W_1)^{-1}W_1' \) to obtain

\[
(I - P_{W_1})y = (I - P_{W_1})W_1 \delta_1 + (I - P_{W_1})W_2 \delta_2 + (I - P_{W_1})\epsilon.
\]

Rewriting the previous model, \( Ky = K \delta_2 + \epsilon \) assuming that \( \epsilon \sim N(0, V) \). Thus, \( K \epsilon \sim N(0, KVK') \) which leads to \( (Ky - K \delta_2) \sim N(0, KVK') \). Therefore, the log likelihood is

\[
\ell(\delta_2, V | y, W_2) = -\frac{np}{2} \ln(2\pi) - \frac{1}{2} \ln |KVK'| - \frac{1}{2} \left( Ky - K \delta_2 \right)' [KVK']^{-1} \left( Ky - K \delta_2 \right)
\]

\[
= -\frac{np}{2} \ln(2\pi) - \frac{1}{2} \ln |KVK'| - \frac{1}{2} \left( y - W_2 \delta_2 \right)' K'[KVK']^{-1} K \left( y - W_2 \delta_2 \right)
\]

\[
= -\frac{np}{2} \ln(2\pi) - \frac{1}{2} \ln |KVK'| - \frac{1}{2} \left( y - W_2 \delta_2 \right)' P \left( y - W_2 \delta_2 \right).
\]
Finding the partial derivative and setting it to zero, \( \frac{\partial \ell}{\partial \delta_2} = -W_2'P\frac{\partial y}{\partial \delta_2} + (W_2'PW_2)\tilde{\delta}_2 \). Thus, 
\[-W_2'P\frac{\partial y}{\partial \delta_2} + (W_2'PW_2)\tilde{\delta}_2 = 0.\]
Next, \( \tilde{\delta}_2 = (W_2'PW_2)^{-1}W_2'P\frac{\partial y}{\partial \delta_2} \) where \( P = V^{-1} - V^{-1}W_1(W_1'V^{-1}W_1)^{-1}W_1'V^{-1} \).

\[ y_i = W_{1i}\tilde{\delta}_1 + W_{2i}\tilde{\delta}_2 + \xi_i \] let \( W_{1i} \) be nuisance exogenous variables. Using matrix notation, then

\[
\begin{bmatrix}
Y^{(1)} \\
Y^{(2)} \\
\vdots \\
Y^{(p)}
\end{bmatrix}
= 
\begin{bmatrix}
W_{11} & 0 & \cdots & 0 \\
0 & W_{12} & 0 & \vdots \\
\vdots & \vdots & \ddots & 0 \\
0 & \cdots & 0 & W_{1p}
\end{bmatrix}
\begin{bmatrix}
\delta_{11} \\
\delta_{12} \\
\vdots \\
\delta_{1p}
\end{bmatrix}
+ 
\begin{bmatrix}
W_{21} & 0 & \cdots & 0 \\
0 & W_{22} & 0 & \vdots \\
\vdots & \vdots & \ddots & 0 \\
0 & \cdots & 0 & W_{2p}
\end{bmatrix}
\begin{bmatrix}
\delta_{21} \\
\delta_{22} \\
\vdots \\
\delta_{2p}
\end{bmatrix}
+ 
\begin{bmatrix}
\xi^{(1)} \\
\xi^{(2)} \\
\vdots \\
\xi^{(p)}
\end{bmatrix}
\]

Rewrite the model to obtain

\[ y^* = W_1^*\tilde{\delta}_1^* + W_2^*\tilde{\delta}_2^* + \xi^*. \]

Using the results above, then \( \tilde{\delta}_2^* = (W_2'^*P_{W_2'})^{-1}[W_2'^*]'P_{y^*} \) where \( P = V^{-1} - V^{-1}W_1^*([W_1^*]'V^{-1}W_1^*)^{-1}[W_1^*]'V^{-1} \). Therefore, the information matrix for \( \tilde{\delta}_2^* \) is

\[ M(\tilde{\delta}_2^*) = -E\left( \frac{\partial^2 \ell}{\partial \tilde{\delta}_2^* \partial \tilde{\delta}_2^*} \right). \]

Thus, \( M(\tilde{\delta}_2^*) = E([W_2'^*]'PW_2'^*) \).

Because \( W_2^* \) includes both endogenous variables as well as exogenous variables, the same approach of using 3SLS and FIML methodologies will be used to find the optimal design, as in previous chapters.

5.2.1 Applying 3SLS Methodology

In each of the seven environments there are 32 blocks for a total of 224 blocks where the block size is 13. There are some constraints in the original study that require a simpler example for the purposes of this research. The computational power required to analyze all 224 blocks would be
excessive. A simpler example with a smaller subset of data for the causal structure will be considered in order to demonstrate the improvement in efficiency using 3SLS methodology. The first four blocks in each replicate were selected. Then, the 3SLS optimal design for the genetic sequences of the genotypes will be compared to those which were used in the original experiment from the study. Although this approach is rather arbitrary, it allows for comparison with the ideal genotype combinations from an optimal design to be observed.

Example #6: Considering the structural model shown in Fig. 5.1 (Dhungana et al. 2007), use 3SLS methodology to obtain an optimal design for 3 environments (years) with 4 blocks in each environment and each block with size 8.

**Figure 5.1:** The selected path model for the yield genotype-by-environment interaction in the Dhungana et al. study from 2007.

Table 5.1 shows the genetic combinations for the genotypes used in the study for three replicates where $x_1$ represents Xbarc67, $x_2$ represents Xbcd1555, $x_3$ represents Xtam055, $x_4$ represents Xksua6, and $x_5$ represents Xbarc86.
Table 5.1: Genotype combinations for three replicates that were used in the study.

Replicate 1:

<table>
<thead>
<tr>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_3$</td>
<td>$x_4$</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>•*</td>
<td>•*</td>
<td>•*</td>
<td>•*</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>•*</td>
<td>•*</td>
<td>•*</td>
<td>•*</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

* The symbol ‘•’ indicates that data are missing.
Replicate 2:

<table>
<thead>
<tr>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_3$</td>
<td>$x_4$</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\bullet^*$</td>
<td>$\bullet$</td>
<td>$\bullet$</td>
<td>$\bullet$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>$\bullet$</td>
<td>$\bullet$</td>
<td>$\bullet$</td>
<td>$\bullet$</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>$\bullet$</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>$\bullet$</td>
<td>$\bullet$</td>
<td>$\bullet$</td>
<td>$\bullet$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

* The symbol ‘$\bullet$’ indicates that data are missing.
Replicate 3:

<table>
<thead>
<tr>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_3$</td>
<td>$x_4$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1 1 1 1 1</td>
<td>-1 -1 -1 -1 -1</td>
<td>• • • • •</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>1 -1 -1 -1 -1</td>
<td>1 1 1 1 1</td>
<td>• • • • •</td>
<td>1 -1 -1 -1 -1</td>
</tr>
<tr>
<td>-1 -1 -1 -1 -1</td>
<td>-1 1 1 1 1</td>
<td>1 1 1 1 1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>1 1 1 1 -1</td>
<td>• 1 1 -1 -1</td>
<td>1 -1 -1 -1 -1</td>
<td>1 -1 -1 -1 -1</td>
</tr>
<tr>
<td>1 -1 -1 -1 -1</td>
<td>• -1 -1 -1 -1</td>
<td>1 -1 -1 -1 -1</td>
<td>1 -1 -1 -1 -1</td>
</tr>
<tr>
<td>-1 1 1 1 -1</td>
<td>1 1 1 1 1</td>
<td>• • • • •</td>
<td>1 -1 1 -1 -1</td>
</tr>
<tr>
<td>1 1 1 1 -1</td>
<td>1 1 1 1 -1</td>
<td>1 1 -1 1 1</td>
<td>-1 -1 -1 -1 -1</td>
</tr>
<tr>
<td>1 -1 -1 -1 -1</td>
<td>1 1 1 1 -1</td>
<td>1 -1 -1 1 1</td>
<td>-1 -1 -1 -1 -1</td>
</tr>
<tr>
<td>1 -1 -1 -1 -1</td>
<td>1 1 1 1 -1</td>
<td>• • • • •</td>
<td>-1 -1 -1 -1 -1</td>
</tr>
<tr>
<td>-1 1 1 1 -1</td>
<td>-1 1 1 1 1</td>
<td>• • • • •</td>
<td>-1 -1 1 -1 -1</td>
</tr>
<tr>
<td>1 1 -1 -1 -1</td>
<td>1 1 1 1 -1</td>
<td>-1 1 -1 -1 -1</td>
<td>-1 1 -1 -1 -1</td>
</tr>
<tr>
<td>1 1 1 1 1</td>
<td>1 -1 -1 -1 -1</td>
<td>-1 1 -1 -1 -1</td>
<td>-1 1 -1 -1 -1</td>
</tr>
<tr>
<td>1 1 1 1 -1</td>
<td>-1 -1 -1 -1 -1</td>
<td>1 -1 -1 -1 -1</td>
<td>-1 1 -1 -1 -1</td>
</tr>
</tbody>
</table>

* The symbol ‘•’ indicates that data are missing.

Table 5.2 shows the optimal genetic combinations for the genotypes for three replicates using 3SLS methodology where $x_1$ represents Xbarc67, $x_2$ represents Xbcd1555, $x_3$ represents Xtam055, $x_4$ represents Xksua6, and $x_5$ represents Xbarc86.
Table 5.2: Optimal genotype combinations for three replicates using 3SLS methodology.

Replicate 1:

<table>
<thead>
<tr>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_3$</td>
<td>$x_4$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>
Replicate 2:

<table>
<thead>
<tr>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_3$</td>
<td>$x_4$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>
Replicate 3:

<table>
<thead>
<tr>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_3$</td>
<td>$x_4$</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

The asymptotic information matrix for the parameters is

$$M(\hat{\delta}_2^*) = [\hat{W}_2^*]' p \hat{W}_2^*. \quad (123)$$

Let $M(\hat{\delta}_2^*)_{3SLS \, optimal}$ represent the information matrix based on the optimal design in Table 5.2 and let $M(\hat{\delta}_2^*)_{\text{Paper}}$ represent the information matrix based on the design in the original study as found in Table 5.1. The log of the determinant for the asymptotic information matrix is

$$\log(|M(\hat{\delta}_2^*)_{3SLS \, optimal}|) = 120.79$$

and then

$$\log(|M(\hat{\delta}_2^*)_{\text{Paper}}|) = 111.18.$$
Thus, the $D$-efficiency is

$$D_{\text{eff}} = \left(\frac{e^{\log\left(|M(\hat{\alpha}^2)_{3SLS\, \text{optimal}}\right)}}{e^{\log\left(|M(\hat{\alpha}^2)_{\text{paper}}\right)}}\right)^{\frac{1}{17}} = \left(\frac{e^{111.18}}{e^{120.79}}\right)^{\frac{1}{17}} = 56.82\%.$$ 

The results show that the genotypic combinations selected for the study from Table 5.1 were about 60% as $D$-efficient as the genotypic combinations from the 3SLS optimal design as shown in Table 5.2. That is to say that the genotype combinations from the 3SLS optimal design are about 40% more $D$-efficient than those selected for the study.

### 5.2.2 Applying FIML Methodology

For the same reasons as mentioned in Section 5.2.1, a simpler example with a smaller subset of data for the causal structure will be considered in order to demonstrate the improvement in efficiency using FIML methodology. Then, the FIML optimal design for the genetic sequences of the genotypes will be compared to those which were used in the original experiment from the study.

Example #7: Considering the structural model shown in Fig. 5.1 (Dhungana et al. 2007), use FIML methodology to obtain an optimal design for 3 replicates with 4 blocks in each environment and each block with size 8.

The information matrix for $\hat{\alpha}_2^*$ is $M(\hat{\alpha}_2^*) = E([W_2^*]'PW_2^*)$. Substituting for $W_2^*$ and multiplying the matrices,

$$[W_2^*]'PW_2^* = \begin{bmatrix} W_{21} & 0 & 0 & 0 & P_{11} & P_{12} & P_{13} & P_{14} \\ 0 & W_{22} & 0 & 0 & P_{21} & P_{22} & P_{23} & P_{24} \\ 0 & 0 & W_{23} & 0 & P_{31} & P_{32} & P_{33} & P_{34} \\ 0 & 0 & 0 & W_{24} & P_{41} & P_{42} & P_{43} & P_{44} \end{bmatrix} \begin{bmatrix} 0 & W_{21} & 0 & 0 \\ W_{22} & 0 & 0 & 0 \\ W_{23} & 0 & 0 & 0 \\ W_{24} & 0 & 0 & 0 \end{bmatrix}$$
\[

c = \begin{bmatrix}
W'_{21}P_{11}W_{21} & W'_{21}P_{12}W_{22} & W'_{21}P_{13}W_{23} & W'_{21}P_{14}W_{24} \\
W'_{22}P_{21}W_{21} & W'_{22}P_{22}W_{22} & W'_{22}P_{23}W_{23} & W'_{22}P_{24}W_{24} \\
W'_{23}P_{31}W_{21} & W'_{23}P_{32}W_{22} & W'_{23}P_{33}W_{23} & W'_{23}P_{34}W_{24} \\
W'_{24}P_{41}W_{21} & W'_{24}P_{42}W_{22} & W'_{24}P_{43}W_{23} & W'_{24}P_{44}W_{24}
\end{bmatrix}.
\]

\[W'_{23}P_{33}W_{23}\] will be used as a demonstration to obtain the information matrix. Substituting for \(W_{23}\) and multiplying the matrices,

\[
W'_{23}P_{33}W_{23} = \begin{bmatrix} \hat{x}_1' \\ \hat{x}_2' \\ \hat{y}_1' \\ \hat{y}_2' \end{bmatrix} P_{33} \begin{bmatrix} x_7 & x_8 & y_1 & y_2 \end{bmatrix}.
\]

Thus,

\[
W'_{23}P_{33}W_{23} = \begin{bmatrix} \hat{x}_1'P_{33}x_7 & \hat{x}_1'P_{33}x_8 & \hat{x}_1'P_{33}y_1 & \hat{x}_1'P_{33}y_2 \\ \hat{x}_2'P_{33}x_7 & \hat{x}_2'P_{33}x_8 & \hat{x}_2'P_{33}y_1 & \hat{x}_2'P_{33}y_2 \\ \hat{y}_1'P_{33}x_7 & \hat{y}_1'P_{33}x_8 & \hat{y}_1'P_{33}y_1 & \hat{y}_1'P_{33}y_2 \\ \hat{y}_2'P_{33}x_7 & \hat{y}_2'P_{33}x_8 & \hat{y}_2'P_{33}y_1 & \hat{y}_2'P_{33}y_2 \end{bmatrix}.
\]

Let \(E \left( y_{(i)} \right) = \hat{y}_{(i)} \), then

\[
E(W'_{23}P_{33}W_{23}) = \begin{bmatrix} \hat{x}_1'P_{33}x_7 & \hat{x}_1'P_{33}x_8 & \hat{x}_1'P_{33}y_1 & \hat{x}_1'P_{33}y_2 \\ \hat{x}_2'P_{33}x_7 & \hat{x}_2'P_{33}x_8 & \hat{x}_2'P_{33}y_1 & \hat{x}_2'P_{33}y_2 \\ \hat{y}_1'P_{33}x_7 & \hat{y}_1'P_{33}x_8 & \hat{y}_1'P_{33}y_1 & \hat{y}_1'P_{33}y_2 \\ \hat{y}_2'P_{33}x_7 & \hat{y}_2'P_{33}x_8 & \hat{y}_2'P_{33}y_1 & \hat{y}_2'P_{33}y_2 \end{bmatrix}.
\]

Table 5.3 shows the optimal genetic combinations for the genotypes for three replicates using

FIML methodology where \(x_1\) represents Xbarc67, \(x_2\) represents Xbcd1555, \(x_3\) represents

Xtam055, \(x_4\) represents Xksua6, and \(x_5\) represents Xbarc86.
Table 5.3: Optimal genotype combinations for three replicates using FIML methodology.

Replicate 1:

<table>
<thead>
<tr>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_3$</td>
<td>$x_4$</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
</tr>
</tbody>
</table>
Replicate 2:

<table>
<thead>
<tr>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_3$</td>
<td>$x_4$</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
</tbody>
</table>
Replicate 3:

<table>
<thead>
<tr>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_3$</td>
<td>$x_4$</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
</tbody>
</table>

Let $M(\tilde{\delta}^2)^{FIML\text{ optimal}}$ represent the information matrix based on the optimal design in Table 5.3.

The log of the determinant of the information matrices are

$\text{Log}(|M(\tilde{\delta}^2)^{FIML\text{ optimal}}|)=120.50$.

and

$\text{Log}(|M(\tilde{\delta}^2)^{\text{Paper}}|)=129.17$.

Thus, the $D$-efficiency is

$$D_{\text{eff}} = \left( \frac{e^{\text{Log}(|M(\tilde{\delta}^2)^{\text{Paper}}|)}}{e^{\text{Log}(|M(\tilde{\delta}^2)^{FIML\text{ optimal}}|)}} \right)^{\frac{1}{17}} = \left( \frac{e^{120.50}}{e^{129.17}} \right)^{\frac{1}{17}} = 62.87\%.$$
The results show that the genotypic combinations selected for the study from Table 5.1 were about 60% as D-efficient as the genotypic combinations from the FIML optimal design as shown in Table 5.3. That is to say that the genotype combinations from the FIML optimal design are about 40% more D-efficient than those selected for the study.

The application presented here represents the ideal genotypic combinations. However, there are some of the challenges that present themselves in this application. First, the genetic sequence of the specific genotypes from the optimal design could be expensive to obtain or are not currently in existence. It may also not be realistic to bioengineer genotypes with those specific genetic sequences in terms of the technology that is currently available. But one way to overcome both of those challenges is to adjust the candidate points and to limit them to genetic sequences of the genotypes which are currently available and to obtain the optimal design based on those available genotypes. In practice, an application like what was presented here could be implemented to improve bioengineering practices such as recombinant DNA technology in order to produce optimal genotypic designs in fields such as agriculture, medicine, and industry.

5.3 Optimal Design Application in Relation to Some Methods for Analyzing Multi-trait Genotype-by-Environment Interactions

The importance of environments can be seen from the previous section. Often an environment is defined broadly as either a different location or different year or both. When considering an environment a location, one of the essential experimental design questions, in a multi-location experiment for a given sample size, is how many replicates each location should have. That is, should all of the locations be treated the same and have the same number of replicates or should some locations have more replicates than other locations. The easy approach in both optimal design for univariate models and for multivariate models is for all locations to be treated the same and to have the same number of replicates, but this is not efficient. There is a considerable history of literature to address the number of replicates in multi-location trials (Sprague and
Federer 1951, Finney 1958, Bos 1983, Pearce 1983, Gauch and Zobel 1996, Endelman et al. 2014). However, there is no current literature on multi-location optimal design for causal structures. Optimal designs for causal structures are more sophisticated than univariate optimal designs because there are multiple endogenous variables with causal intrarelationships among them. For example, grain yield depends on a number of supporting traits influenced by many different genes and by a multitude of environmental conditions at different stages during plant development (Ashikari et al., 2005). This indicates that environments may affect the endogenous variables differently. For that reason, some locations may be more important than others and should be replicated more.

In the next application, 3SLS and FIML optimal design methodologies will be used to obtain the number of replicates for different locations that will produce the most precise estimates for a given causal structure. The data from this second application comes from a study by Vargas et al. (1998, 1999) where multi-trait, multi-environment experiments were conducted in Ciudad Obregon, Mexico over six years from 1990 – 1995. The study analyzed the genotype x environment interaction of seven different durum wheat cultivars and seven different bread wheat cultivars. Extensions of the work were conducted by Dhungana (2004) and Yaseen (2012) in dissertation research at the University of Nebraska-Lincoln. This application applies optimal design methodology to the causal model Yaseen (2012) using the durum wheat data from Vargas et al. (1998; 1999).

The data set for durum wheat cultivars included 6 environments with 7 replicates for a total of 42 replicates. In the original study, an environment was considered as one year. However, in this dissertation research, the data will be used as if the years are different locations with the implicit assumption that the across year variability at this site is comparable to across location variability in the local wheat growing region. The ultimate question for a plant breeder is to determine how many replicates are recommended in each location. The weather data in Table 5.4 are assumed to
be 30-year averages for each “location” where an environment represents a different location.

The variables in the causal structure that were represented included environmental mean daily maximum temperature (MXTD), environmental mean daily minimum temperature (MNTD), mean sun hours per day (SHM), monthly precipitation fall (PRF), spikes per square meter (SPSM), kernels per spike (KPS), and thousand-kernel weight (TKW).

Fig. 5.2 below represents the path model that was developed in order to analyze the genotype-by-environment interactions on yield.

**Figure 5.2:** The path model developed for the yield genotype-by-environment interactions in the Yaseen dissertation from 2012.

The objective of this application is to find the number of replicates for each environment and compare the efficiency of the new optimal design to the efficiency of the design that was used in the study where the environment had the same number of replicates.

### 5.3.1 Applying 3SLS Methodology

Using the 3SLS methodology that was developed in Chapters 3 and 4, the objective would be to maximize the determinant of the asymptotic information matrix of the endogenous and the exogenous parameters. The values of the exogenous variables for each location is given in Table 5.4 below.
Table 5.4: Values of the exogenous variables for each location* for mean daily maximum temperature (MXTD), mean daily minimum temperature (MNTD), mean sun hours per day (SHM), and monthly precipitation fall (PRF) (Vargas et al. 1998 & 1999).

<table>
<thead>
<tr>
<th>Location</th>
<th>MXTD</th>
<th>MNTD</th>
<th>SHM</th>
<th>PRF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24.9</td>
<td>9.2</td>
<td>9.2</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>24.2</td>
<td>10.1</td>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>23.9</td>
<td>10.9</td>
<td>8.2</td>
<td>59</td>
</tr>
<tr>
<td>4</td>
<td>23.9</td>
<td>10.9</td>
<td>10.3</td>
<td>41</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>11</td>
<td>8.4</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>22.9</td>
<td>11.3</td>
<td>9.8</td>
<td>0</td>
</tr>
</tbody>
</table>

* Each location represents a different environment from Vargas et al. (1998, 1999) where the weather variables are assumed to be 30-year averages for the location.

Using the information from Table 5.4 and the 3SLS information matrix, the optimal number of replicates for each location is listed in Table 5.5.

Table 5.5: Optimal number of replicates for each location* using 3SLS methodology.

<table>
<thead>
<tr>
<th>Label</th>
<th>Number of Replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location 1</td>
<td>10</td>
</tr>
<tr>
<td>Location 2</td>
<td>0</td>
</tr>
<tr>
<td>Location 3</td>
<td>5</td>
</tr>
<tr>
<td>Location 4</td>
<td>8</td>
</tr>
<tr>
<td>Location 5</td>
<td>10</td>
</tr>
<tr>
<td>Location 6</td>
<td>9</td>
</tr>
</tbody>
</table>

* Each location represents a different environment from Vargas et al. (1998; 1999) where the weather variables are assumed to be 30 year averages for the location.
The 3SLS information matrix determinant for the optimal number of replicates and the
determinant of the information matrix for the equal number of replicates using durum wheat data
from Vargas (1998, 1999) are calculated as

$$\log(|M(\hat{\delta}_2^2)|_{3SLS\text{ Optimal}}) = 49.4338$$

and

$$\log(|M(\hat{\delta}_2^2)|_{\text{Paper}}) = 48.3851.$$ 

Therefore,

$$e^{\log(|M(\hat{\delta}_2^2)|_{\text{Paper}}) - \log(|M(\hat{\delta}_2^2)|_{3SLS\text{ Optimal}})} \approx 0.35\%.$$ 

Since there are fourteen parameters, the $D$-efficiency is

$$D_{\text{eff}} = \left( \frac{e^{\log(|M(\hat{\delta}_2^2)|_{\text{Paper}}) - \log(|M(\hat{\delta}_2^2)|_{3SLS\text{ Optimal}})}}{e^{48.3851} - e^{49.4338}} \right)^{\frac{1}{14}} = 92.28\%.$$ 

From the results above it can be seen that using equal replicates in each location is not optimal
nor should all locations be used. Using the optimal number of replicates would result in a 65%
increase in the information matrix as compared to the design that was used in the study where all
environments had seven replicates. The design in the original study with equal replicates is
approximately 92% as $D$-efficient as the 3SLS optimal design.

The FIML optimal design methodology is an alternative to 3SLS methodology and will be used
below.

5.3.2 Applying FIML Methodology

Using the FIML methodology that was developed in Chapters 3 and 4, the objective would be to
maximize the determinant of the FIML information matrix of the endogenous and the exogenous
parameters. The values of the exogenous variables for each location from Table 5.4 and the FIML information matrix were used to determine the optimal number of replicates for each location as listed in Table 5.6 below.

Table 5.6: Optimal number of replicates for each location* using FIML methodology.

<table>
<thead>
<tr>
<th>Label</th>
<th>Number of Replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location 1</td>
<td>7</td>
</tr>
<tr>
<td>Location 2</td>
<td>0</td>
</tr>
<tr>
<td>Location 3</td>
<td>8</td>
</tr>
<tr>
<td>Location 4</td>
<td>8</td>
</tr>
<tr>
<td>Location 5</td>
<td>10</td>
</tr>
<tr>
<td>Location 6</td>
<td>9</td>
</tr>
</tbody>
</table>

* Each location represents a different environment from Vargas et al. (1998, 1999) where the weather variables are assumed to be 30 year averages for the location.

The FIML information matrix determinant for the optimal number of replicates and the determinant of the information matrix for the equal number of replicates from the Vargas et al. study (1998, 1999) are calculated as

\[
\log(|M(\hat{\sigma}^2)|_{\text{FIML Optimal}}) = 94.8575
\]

and

\[
\log(|M(\hat{\sigma}^2)|_{\text{Paper}}) = 93.8563.
\]

Therefore,

\[
\frac{e^{\log(|M(\hat{\sigma}^2)|_{\text{Paper}})}}{e^{\log(|M(\hat{\sigma}^2)|_{\text{FIML Optimal}})}} = \frac{e^{93.8563}}{e^{94.8575}} = 0.3674 = 36.74%.
\]
Since there are fourteen parameters, the $D$-efficiency is

$$D_{\text{eff}} = \left( \frac{e^{\log(|M(\hat{\delta}^2)|_{\text{Paper}})}}{e^{\log(|M(\hat{\delta}^2)|_{\text{FIML Optimal}})}} \right)^{\frac{1}{14}} = \left( \frac{e^{93.8563}}{e^{94.8575}} \right)^{\frac{1}{14}} = 93\%$$

Again, the results above demonstrate that using equal replicates in each environment is not optimal. The FIML optimal design increased the determinant of the information matrix by 53% in comparison to the design from the study with equal replicates. The design in the original study with equal replicates is approximately 93% as $D$-efficient as the FIML optimal design.

Using FIML methodology produced similar results as the 3SLS methodology. In both designs, analysis showed that the second location should not be used for testing. This result was surprising, but upon closer inspection the result is logical. The average value of all the environmental factors in Location 2 lie in between the average values from the other locations. Location 2 had neither a maximum nor a minimum value as compared to the other locations in the study and this is unique to Location 2. For example, the mean daily maximum temperature (MXTD) in Location 2 is 24.2. The maximum MXTD was 25 in Location 5 and the minimum MXTD was 22.9 in Location 6. This result was similar with each of the other environmental factors. Location 2 does not give the optimal amount of information because it does not include any extreme values from any of the environmental factors. Locations 4, 5, and 6 have the same number of replicates in both the 3SLS and FIML optimal designs. Both designs also have a similar increase in the determinant of the information matrix and in $D$-efficiency in comparison to the design with equal replicates that was used in the original study.

These results show the importance of taking into account the differences among locations to determine the number of replicates for given locations and can be illustrated using the UNL Wheat Breeding Program, for example. Western Nebraska receives less annual rainfall which is more erratic, versus eastern Nebraska which has higher annual rainfall and higher humidity.
(Nielsen, Virgil, and Benjamin 2009, Shulski et al. 2013). Because factors such as annual rainfall, minimum temperature, and sun hours will affect endogenous variables differently, then the optimal design will produce more precise estimates of the causal structure parameters compared to an equal-replicate design. So, before experiments are run in locations such as Mead, NE, North Platte, NE, or Sidney, NE, it would be advisable to use the 30-year weather variable averages and an optimal design approach to predict the number of replicates in each location, particularly if understanding the complex structure among traits regarding genotype-environment interaction is of interest.

The approach that was used in this application can be used to determine the optimal number of replicates for locations in other research. For example, when studying any complex causal structure, this approach can be used when experiments are performed in different labs or in different locations or with different populations. It can be used in survey research conducted in with different healthcare systems, school districts, congressional districts, etc.

5.4 Summary

The importance of using optimal designs to produce efficient estimates for the parameters of interest was shown through two different applications, one in selecting optimal genotypes and the other in optimizing the number of replicates for different locations. In both applications, it was demonstrated that the optimal design methodologies that were developed in Chapters 3 and 4 resulted in significant increases in $D$-efficiency as compared to the designs that were selected for the original studies. In the application from Section 5.2, the new optimal designs were 40 – 50% more $D$-efficient than the designs chosen for the original study. In the application from Section 5.3, the new optimal designs were 7 – 8% more $D$-efficient than the design chosen for the original study. These results validate the 3SLS and FIML methodologies for producing improved designs in actual applications when causal structure modeling is of interest.
5.5 References


CHAPTER 6: CONCLUSION

6.1 Summary

The need for a design that is optimal for all of the estimated parameters is what has driven the growth of optimal design theory. While the majority of statistical analyses focus on univariate linear models, these models have limitations since with most natural phenomena, multiple endogenous variables are needed to understand the process being studied. This fact ultimately led to the development of the multivariate linear model.

Multivariate linear models give additional insights over the univariate linear model. The simultaneous analysis of multiple dependent variables provides a better understanding of the process with less error and more validity as compared to univariate methods. Cluster analysis, factor analysis, discriminate analysis, principal component analysis, MANOVA, and canonical correlation are the most common types of multivariate methods and incorporate associations among the dependent variables. But in many natural systems, the variables are interrelated in complicated ways, which led to the development of causal structure modeling often termed structural equation modeling.

Prediction and model selection are the main objectives of causal structure modeling. Precise estimates for the parameters are needed to obtain an accurate prediction. In order to select a model, it must both “make sense” as it relates to the research topic and have “good” goodness-of-fit parameters. The challenges for researchers include the number of models to choose from and how to determine between models being “good” based on some goodness-of-fit parameters, but not others. If the model is known and there is no interest in model selection, then it is best to have the most precise estimates to allow for an accurate prediction. This can be achieved by taking previously known information and designing an experiment that will allow for the most precise
estimates for the designated parameters, which will lead to accurate predictions and to better inferential statistics.

Traditional optimal design literature largely focuses on univariate linear and univariate mixed models. While there is a smaller body of work done for multivariate models, there is no current literature including work done for causal structure modeling, which was the motivation for this research. This research covered four objectives. First, we showed how to obtain an optimal design that allowed the most precise estimates of the endogenous and exogenous parameters for a given causal structure. Then, we demonstrated how to obtain an optimal design for the endogenous and exogenous parameters for a given causal structure with random blocks and also for a given causal structure with fixed blocks. Finally, we showed how to obtain a $D$-optimal design that took into account the precision of the endogenous, exogenous, and the variance components estimates for a causal structure with mixed effects using ML and REML information matrices. The efficiency of each of the optimal designs for causal structures were compared with the optimal designs for the univariate case based on simulations. The methods that were developed and the simulations were used to successfully demonstrate the improvement in efficiency for two applications from previous research.

The optimality criteria were developed for both the 3SLS and the FIML estimates for a completely randomized causal structure and for a causal structure with blocks. 3SLS estimators are not based on optimization, therefore the estimate of the asymptotic information matrix was used instead of the asymptotic information matrix. For those reasons, FIML estimators may be more appropriate for small sample sizes because they are based on the information matrix, not its estimate. The optimal designs for the given simulation examples were the same for both 3SLS and FIML estimates. The new optimal design for the causal structure example increased the determinant of the information matrix by at least 20% as compared to the univariate optimal
design. Because of this increase, the new optimal design was 6.6% more $D$-efficient than the univariate optimal design.

The causal structure also had a significant impact on the optimal design based on data from an experiment with random blocks. The optimal design for a univariate mixed model is an orthogonal design which is universally optimal, but it is not optimal for a causal structure model. The new optimal design for the causal structure with random blocks increased the determinant of the information matrix for the simulation by at least 9%, demonstrating that the univariate orthogonal design is less efficient than the optimal design for a causal structure. The new design based on random blocks or fixed blocks was consistently more $D$-efficient than the univariate orthogonal design for both the 3SLS and FIML approaches.

Three criteria for the optimal design for the endogenous parameters, exogenous parameters, and variance components in a causal structure model with random blocks were developed. Again, the 3SLS optimal design is not explicitly based on optimization, whereas the FIML simultaneous optimal designs are based on optimization of the log likelihood. The differences between the 3SLS and FIML optimal designs were not large, which was different from the criteria that were developed for the optimal design only considering endogenous and exogenous model parameters.

In the 3SLS simultaneous optimal design, the endogenous and exogenous parameters estimates are independent of the covariance parameter estimates. This allows for the information matrix to be partitioned and for composite criteria to be used, similar to univariate optimal design. The 3SLS approach requires that a weight $\alpha$ be chosen, where $\alpha$ is a real number between 0 and 1. The optimal design can change depending on the choice of $\alpha$. In the univariate case, the fixed parameter estimates and the random parameter estimates are independent. However, such independence is not the case with causal structures where the endogenous parameters estimates are correlated with the covariance parameter estimates.
FIML optimal designs for endogenous and exogenous parameters with ML and REML estimators for the variance components have three advantages over 3SLS simultaneous optimal design. First, ML and REML are based on optimization of the log likelihood. Secondly, the optimal design is not subjective to different $\alpha$ values because $\alpha$ is not included. Lastly, the approaches allow the correlation between the endogenous parameters estimates and the covariance parameter estimates to be included in the information matrix.

Unlike ML estimators, the REML information matrix for the estimators for the covariance components and for the covariance between the variance components depends on the design matrix. The ML information matrix of the covariance estimates and the correlation between the endogenous parameters estimates and the covariance estimates do not depend on the design.

When comparing the performance of the univariate optimal design to the new REML optimal design, the REML design was about 30% more $D$-efficient than the univariate optimal design. The performance of the REML optimal design was also about 20% more $D$-efficient than the performance of the ML optimal design. Based on these results, it is recommended to use the REML optimality criteria for an optimal design for a causal structure because it includes the information matrix for the endogenous parameters, exogenous parameters, the variance components, and the covariances among these estimates.

There are many applications of structural equation modeling in a variety of fields and the importance of using optimal designs to produce efficient estimates for the parameters of interest was demonstrated using two examples from previous studies. The first example came from a study on genotype-by-environment interaction in to wheat published in *Crop Science* (Dhungana et al. 2007). The objective of the original study was to understand how genotype-by-environment interaction influences complex traits such as grain yield. But it failed to take into account whether or not the genotypes were suitable for the study. The results produced here showed that
line genotype combinations based on 3SLS and FIML optimal designs were about 40 – 50% more \(D\)-efficient than the genotype combinations used in the study.

The second example came from a wheat breeding trial conducted in Obregon, Mexico (Vargas et al. 1998) and the causal model described by Yaseen (2012). Yaseen’s objectives were to develop a standard mixed model methodology that takes into account causality among multi-trait, multi-environment trials and to use that method to model the interactions using data sets for genotype plus genotype-by-environment interaction. It was assumed that each year in the study represented a different location from the local wheat production region, using the same number of replicates. Using the equal number of replicates per location failed to consider that some locations should have more replicates than others based on the fact that locations may affect the endogenous variables differently. The results produced here showed that using equal replicates in each location is would not be optimal and further showed the importance of taking into account the differences among locations to determine the number of replicates. The number of replicates from the 3SLS and FIML optimal designs were 7 – 8% more \(D\)-efficient than the design from the study with equal replicates. When assuming the environmental variability in Vargas et al. (1998) was an adequate representation of location variability, the results indicated that, for a given number of plots, some locations should never be used while resources should be oriented toward having more replications in some locations than in others.

### 6.2 Future Work

From the results of this dissertation, it can be concluded that further research on optimal design for causal structures is needed. Future work in this area can be extended to include topics such as optimal design for a generalized causal structure, using a Bayesian approach to allow for the uncertainty of the endogenous and exogenous parameters, optimal design based on goodness-of-fit parameters rather than a function of the covariance of the estimates, nonlinear causal structure optimal design, as well as other topics.
Throughout this current research, normality was assumed for all of the endogenous parameters. However, normality cannot always be assumed for the endogenous variables. For example, endogenous variables can be binary. In this case, it is necessary to use a generalized causal structure that can account for the different distributional assumptions of the variables. There is currently literature that addresses generalized causal structure, but not in relation to optimal design theory. The objective would be to develop the methodology and algorithms to obtain an optimal design for a generalized causal structure where all endogenous variables could be either continuous or discrete. For example, in a study published in *The Annals of Human Genetics* on the subject of coronary heart disease (Mi et al. 2011), the authors studied the complex intrarelationship among the physiological or genetic factors and the environmental factors and their interactions effects on coronary heart disease (CHD), a binary endogenous variable. The complex intrarelationship between these factors and their interactions is given by the model shown in Fig. 6.1. In this example, an optimal design for a generalized structure would be to identify those patients that would be most useful in estimating the causal system.

**Figure 6.1:** Path estimates of SEM of gene-by-environment interaction in the development of CHD as published in *Annals of Human Genetics* by Mi et al. (2011).
In this example, no pre-planned design was used. The data were collected and analysis was performed based on that data. However, a major issue is how might future data be selected to most efficiently estimate the parameters of interest. For example, in a similar hypothetical future experiment, assume that there are enough resources to collect information for 200 people. Some of the questions that should be asked about the 200 subjects should include:

1. How many men and how many women should be included in the study?
2. How many of the subjects smoke?
3. How many of the subjects are obese?
4. If the budget allows for 200 people out of 1,000 volunteers, how do you determine the 200 subjects that should be selected?

Using the optimal design for the generalized causal structure will allow us to obtain the most precise estimates for describing the system.

Throughout this dissertation research, it was assumed that the parameters are known or that there are estimates for the parameters. In planning simple experiments, it is common to base design characteristics on conjectured values of experimental error variance. However, in general, these parameters are unknown. To allow for the uncertainty regarding the endogenous and exogenous parameters, a Bayesian approach may be used. Similar to the approach that was used in the univariate composite criteria that was proposed by Mylona, Goos, and Jones (2014), a prior distribution can be used to allow for the uncertainty of the magnitude of these parameters. Numerical techniques like Gauss-Hermite quadrature can be used to evaluate the integrals.

One important aspect of causal structure modeling is how to measure the goodness-of-fit of the model. In this dissertation research, the optimal designs were based on a function of the covariance of the parameter estimates. Instead, an alternative approach would be to use
goodness-of-fit objective functions to obtain an optimal design. In this case, an optimal design result will make the model best fit the data. This approach has the potential to be a cost-saving measure in that the need to collect additional data can be avoided. The need for redoing an experiment can also be avoided due to the lack of meaningful models that adequately fit the data.

Considering nonlinear structural models is also a needed area of future work. As in the univariate case, nonlinear causal models can be more descriptive and efficient than linear models and the parameters may be more meaningful and easier to interpret. Throughout this dissertation research, the optimal designs were obtained for a linear causal structure. The approach for nonlinear causal structure models would be similar in that a function of the covariance estimates would be used to obtain the optimal design. This will allow for the most efficient estimates to be obtained.

6.3 References


SAS CODE FOR CHAPTER 3

Simulation:

PROC IML; reset;

x={
1 0,
0 1,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0
};
*1++++++++++++++++++++++++++++++++++++;
x1=x[,1]; x2=x[,2];
y1=8*x1+RANNOR(J(20,1,-1));
y2=5*y1+2*x2+RANNOR(J(20,1,-1));
*+++=+++++++++++++++++++++++++++++++;
y=x||y1||y2;
Print y;
quit;
run;
3SLS Optimal Design Completely Randomized Causal Structure:

*1)+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++;
*yhat not y1hat=E(y1)=8x1;
*Fedorov design;
* 3SLS OPTIMAL DESIGN BASED ON THE 1ST SIMULATION AND OPTIMA DESIGN "i";
*This Program Gives An Optimal Design For Causal Structure Equations;
PROC IML;RESET;
sx={(1 0,0 1,1 1,0 1,1 0,1 1,0 1,1 0,1 1,0 1,1 0,1 1,0 1,1 0,1 1,0 1,1 0,1 1,0 1,1 0,1 1,0 1,1 0,1 1,0 1,1 1,0 1,1 1,0 1,1 1,0 1,1 1,0 1,1 1,0 1,1);
s=0;D=-1;
sigma={1 0,0, 1 0};
isigma=inv(sigma);
CanP={0 0,0 1,1 0,1 1};
do i=1 to 17;
do k=1 to 20;
idx = setdif(1:nrow(sx),k);
u = sx[idx, ];
do j=1 to 4;
x=insert(u,CanP[j,],k,0);
y1hat=8*x[ ,1];
r=nrow(x);
z11=x[ ,1];
z12=0*j(r,2);
z21=0*j(r,1);
z22=y1hat|x[ ,2];
z=(z11||z12)/(z21||z22);
pxx=x`*x;
c=det(x`*x);
if c > 0 then Ipxx = Inv(xpxx);
if c > 0 then Ipxxl= x*Ipxx*x`;
if c > 0 then s = det(z`*(Isigma@Ipxxl)*z);
if c = 0 then s = 0;
if s > D then M = x;
if s > D then D = s;
end;
end;
sx=M;
xpx=sx\*sx;
end;
print sx D xpx;
quit;
a)+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

*yhat not y1hat=E(y1)=7.9056905x1;
*Fedorov design;
* 3SLS OPTIMAL DESIGN BASED ON THE 1ST SIMULATION AND OPTIMA DESIGN

"i";
*This Program Gives An Optimal Design For Causal Structure Equations;
PROC IML;RESET;*This to loop and loop;

sx={
1 0,
0 1,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,

1 0
};
s=0;D=-1;
sigma={1 0,0 1};
Isigma=inv(sigma);
CanP={0 0,0 0,1,1 0,0,1 1};
do i=1 to 17;
do k=1 to 20;
idx=setdif(1:nrow(sx),k);
u=sx[idx, ];

do j=1 to 4;
    x=insert(u,CanP[j],k,0);
    ylhat=7.9056905*x[1];
    r=nrow(x);
    z11=x[1, ];
    z12=0*j(r,2);
    z21=0*j(r,1);
    z22=ylhat|x[2, ];

    z=(z11||z12)//(z21||z22);
    xpx=x`*x;
    c=det(x`*x);
    if c > 0 then Ixpx = Inv(xpx);
    if c > 0 then IxpxI= x*Ixpx*x`;
    if c > 0 then s = det(z*(Isigma@IxpxI)*z);
    if c = 0 then s = 0;
    if s > D then M = x;
    if s > D then D = s;
end;
end;
sx=M;
xpx=sx'*sx;
end;
print sx D xpx;
quit;
*1 b)  
+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++;  
yhat not y_1hat=E(y_1)=8.2479385x_1;  
*Fedorov design;  
* 3SLS OPTIMAL DESIGN BASED ON THE 1ST SIMULATION AND OPTIMA DESIGN  
"i";  
*This Program Gives An Optimal Design For Causal Structure Equations;  
PROC IML;  
RESET;  
*This to loop and loop;  
x_1={  
  1 0,  
  0 1,  
  1 1,  
  0 1,  
  1 0,  
  1 1,  
  0 1,  
  1 0,  
  1 1,  
  0 1,  
  1 0,  
  1 1,  
  0 1,  
  1 0,  
  1 1,  
  0 1,  
  1 0,  
  1 1,  
  0 1,  
  1 0,  
  1 1,  
  0 1,  
  1 0,  
  1 1,  
  0 1,  
  1 0,  
  1 1,  
  0 1,  
  1 0,  
  1 1,  
  0 1,  
  1 0,  
  1 1,  
  0 1,  
  1 0,  
  1 1};  
s_
=0;D=-1;  
sigma=(1 0,0 1);  
Isigma=inv(sigma);  
CanP={0 0,0 1,1 0,1 1};  
do i= 1 to 17;  
do k= 1 to 20;  
idx = setdif(1:nrow(sx),k);  
u = sx[idx, ];  
do j=1 to 4;  
x=insert(u,CanP[j,],k,0);  
y_1hat=8.2479385*x[1,];  
r=nrow(x);  
z_11=x[1,];  
z_12=0*j(r,2);  
z_21=0*j(r,1);  
z_22=y_1hat|x[2,];  
z=(z_11||z_12||z_21||z_22);  
xpx=x`*x;  
c=det(x`*x);  
  if c > 0 then Ixpx = Inv(xpx);  
  if c > 0 then IxpxI = x*Ixpx*x`;  
  if c > 0 then s = det(z*(Isigma@IxpxI)*z);  
  if c = 0 then s = 0;  
  if s > D then M = x;  
  if s > D then D = s;  
end;  
end;
sx=M;
xpx=sx'*sx;
end;
print sx D xpx;
quit;
PROC IML;  *This to loop and loop*

sx={
1 0,
0 1,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
0 1,
1 1,
0 1,
1 0,
0 1,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0};  
s=0; D=-1;
sigma={1 0, 0 1};
Isigma=inv(sigma);
CanP={0 0, 1, 1 0, 1 1};
do i=1 to 17;
do k=1 to 20;
idx = setdif(1:nrow(sx), k);
u = sx[idx, ];
do j=1 to 4;
x=insert(u,CanP[j,],k,0);  
ylhat=8.3388438*x[,1];
r=nrow(x);
z11=x[,1];
z12=0*j(r,2);
z21=0*j(r,1);
z22=ylhat| x[,2];

z=(z11||z12)/(z21||z22);

xpx=x`*x;
c=det(x`*x);
if c > 0 then Ixpx = Inv(xpx);
if c > 0 then IxpxI= x*Ixpx*x`;  
if c > 0 then s = det(z`*(Isigma@IxpxI)*z);
if c = 0 then s = 0;
if s > D then M = x;
if s > D then D = s;
end;
end;
sx=M;
xpx=sx'*sx;
end;
print sx D xpx;
quit;
Determinants of the Covariance Matrices for the 3SLS Causal Structure Optimal Design and the 3SLS Univariate Optimal Design:

```
PROC IML; reset; * Simulation Data and estimating the identified Parameters;
*Simulation Parameters;
*xcopt stands for x causal optimal;
div2copt=J(3,3,0);
div2uopt=J(3,3,0);
do k=1 to 100000;
Bhat={8, 2, 5};
xcopt={
   1 0,
   1 0,
   1 1,
   1 1,
   1 0,
   1 1,
   1 1,
   1 0,
   1 1,
   1 0,
   1 1,
   1 0,
   1 0,
   1 0,
   1 1,
   1 0,
   1 1,
   0 1,
   1 0,
   1 1,
   1 0,
   1 0,
   1 1,
   0 1,
   1 0,
   1 1,
   1 0,
   0 1,
   1 0};
*xuopt stands for x univariate optimal;
xuopt={
   0 1,
   1 0,
   1 1,
   0 1,
   1 0,
   1 1,
   0 1,
   1 0,
   1 1,
   0 1,
   1 1,
   0 1,
   1 1,
   0 1,
   1 0,
   1 1,
   0 1,
   1 1,
   0 1,
   1 1,
   0 1,
   1 1,
   0 1,
   1 0,
   1 1,
   0 1,
   1 0};
ylcopt=8*xcopt[,1]+RANNOR(J(20,1,-1));
```
y2copt = 5 * y1copt + 2 * xcopt + RANNOR(J(20, 1, -1));
*++++++++++++++++++++++++++++++;
y1uopt = 8 * xuopt + RANNOR(J(20, 1, -1));
y2uopt = 5 * y1uopt + 2 * xuopt + RANNOR(J(20, 1, -1));
*Print y11 y12 y13 y14 y21 y22 y23 y24;

*Estimating process starts ++++++++++++++++++++++++++++++++;

*Constructing the variance ++++++++++++++++++++++++++++++++;

Sigma = I(2);
SigmaInv = Inv(Sigma);

*vcopt = Sigma @ xpx;
*vinv = Inv(v);

wcopt = (xcopt || J(20, 2, 0)) / (J(20, 1, 0) || xcopt || y1copt);
wcoptp = wcopt;
wuopt = (xuopt || J(20, 2, 0)) / (J(20, 1, 0) || xuopt || y1uopt);
wuoptp = wuopt;

InfMatcopt = wcoptp * (SigmaInv @ xInvxpxxpcopt) * wcopt;
InfMatinvcopt = Inv(InfMatcopt);

InfMatuopt = wuoptp * (SigmaInv @ xInvxpxxpuopt) * wuopt;
InfMatinvuopt = Inv(InfMatuopt);

*estimating ++++++++++++++++++++++++++++++++;

Bhatcopt = InfMatinvcopt * wcoptp * (SigmaInv @ xInvxpxxpcopt) * ycopt;
Bhatcopt = InfMatinvcopt * wcoptp * (SigmaInv @ xInvxpxxpcopt) * ycopt;
div2copt = div2copt + div2copt;
divuopt = Bhat - Bhatcopt;
div2uo = divuopt + divuopt;
end;
detinfcopt = det(div2copt);
detinfuo = det(div2uo);
print detinfcopt detinfuo;
*Print Bhat Bhatcopt Bhatuopt;
quit;
run;
FIML Optimal Design Completely Randomized Causal Structure:

*FIML estimators;

**********;

*This Program Gives An Optimal Design For Causal Structure Equations;
* y1 = 8x1+N(0,1)
  y2 = 2x2 + 5y1 + N(0,1);
*In the dissertation there are 3 simulation and here three designs for these simulations;
*This Program Gives An Optimal Design For Causal Structure Equations;
* This is the 1st design based on the first simulation;

* This give us the optimal design for FIML COVERANCE
  In this trial I considered the TRUE PARAMETERS;

PROC IML; RESET; reset; *This to loop and loop;

sx={
  1 0,
  0 1,
  1 1,
  0 1,
  1 0,
  1 1,
  0 1,
  1 0,
  1 1,
  1 0,
  0 1,
  1 1,
  0 1,
  1 0,
  1 1,
  0 1,
  1 0,
  0 1,
  1 0,
  1 1,
  0 1,
  1 0,
  0 1,
  1 0,
  1 0};
s=0; D=-1; c=-1;
sigma=(1 0, 0 1);
I sigma=inv(sigma);
A =(1 -5, 0 1); IA=inv(A); B=(-8 0, 0 -2);
CanP={0 0, 0 1, 1 0, 1 1};
do i= 1 to 17;
do k= 1 to 20;
idx = setdif(1:nrow(sx),k);
u = sx[idx, ];
do j=1 to 4;
x=insert(u,CanP[j,],k,0);
r=nrow(x);
YHate=x*B*IA;
r=nrow(x);
W=YHate||x;
z11=x[,1];
z12=0*j(r,2);
z21=0*j(r,1);
z22=YHate[,1]||x[,2];
z=(z11||z12)//(z21||z22);
Ide = I(r);
G = Isigma@Ide;
D = det(z\textsuperscript{T}Gz);
\begin{verbatim}
  print i k j D;
  if D > c then M = x;
  if D > c then c = D;
end;
  print c;
end;
sx = M;
print sx D;
end;
quit;
\end{verbatim}
PROC IML;RESET;reset;
sx={
1 0,
0 1,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0};
s=0;D=-1;c=-1;
sigma=(1 0 0 1);
Isigma=inv(sigma);
A = {1 -5.0318 0 1};IA=inv(A);B={-7.9824 0,0 -2.0511};
CanP={0 0,0 1,1 0,1 1};
do i=1 to 10;
do k=1 to 20;
idx=setdif(1:nrow(sx),k);
u = sx[idx, ];
do j=1 to 4;
  x=insert(u,CanP[j,],k,0);
  r=nrow(x);
  YHate=x*B*IA;
  r=nrow(x);
  W=YHate||x;
  z11=x[,1];
  z12=0*j(r,2);
  z21=0*j(r,1);
  z22=YHate[,1]|x[,2];
  z=(z11||z12)/(z21||z22);
  Ide=I(r);
  G=Isigma@Ide;
  D=det(z`*G*z);
  print i k j D;
  if D > c then M = x;
  if D > c then c = D;
end;
print c;
end;
sx=M;
print sx D;
end;
quit;
PROC IML; RESET; reset; *This to loop and loop;
sx={
1 0,
0 1,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0};
s=0; D=-1; c=-1;
sigma=(1 0, 0 1);
Isigma=inv(sigma);
A = {1 -4.9756 0 1}; IA=inv(A); B={-8.1963 0, 0 -1.1748};
CanP={0 0, 1 1 0, 1 1};
do i= 1 to 10;
do k= 1 to 20;
idx = setdif(1:nrow(sx), k);
u = sx[idx, ];
 do j=1 to 4;
  x = insert(u, CanP[j, ], k, 0);
  r = nrow(x);
  YHate = x*B*IA;
  r = nrow(x);
  W = YHate||x;
  z11 = x[, 1];
  z12 = 0*j(r, 2);
  z21 = 0*j(r, 1);
  z22 = YHate[, 1]||x[, 2];
  z = (z11||z12)/(z21||z22);
  Ide = I(r);
  G = Isigma@Ide;
  D = det(z`*G*z);
  print i k j D;
  if D > c then M = x;
  if D > c then c = D;
end;
print c;
end;
sx=M;
print sx D;
end;
quit;
*FIML OPTIMAL DESIGN BASED ON THE 3rd SIMULATION AND OPTIMA DESIGN*

```
PROC IML; RESET; reset; *This to loop and loop;

sX={
0 1,
0 1,
1 1,
1 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 1,
0 1,
1 1,
0 1,
1 0,
1 1,
0 1,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,}

s=0; D=-1; c=-1;
sigma={1 0, 0 1};
Isigma=inv(sigma);
A = {1 -4.9825, 0 1}; IA=inv(A); B={-8.3555 0, 0 -2.0391};
CanP={0 0, 1, 1 0, 1 1, 0 1, 1 0, 1 1, 0 1, 1 1, 0 1, 1 1, 0 1, 1 0, 1 1, 0 1, 1 1, 0 1, 1 0, 1 1, 0 1, 1 0, 1 1, 0 1};

do i= 1 to 10;
do k= 1 to 20;
idx = setdif(1:nrow(sX), k);
u = sX[idx, ];

do j= 1 to 4;
x=insert(u, CanP[j[,], k, 0]);
r=nrow(x);
YHate=x*B*IA;
r=nrow(x);
W=YHate||x;
z11=x[,1];
z12=0*j(r, 2);
z21=0*j(r, 1);
z22=YHate[, 1]||x[, 2];
z=(z11||z12)/(z21||z22);
Ide=I(r);
G=Isigma@Ide;
D=det(z''*G''*z);
    print i k j D;
    if D > c then M = x;
    if D > c then c = D;
end;
print c;
*y0=resp;
*X0=M;
end;
```
sx=M;
print sx D;
end;
quit;
Comparing the FIML, 3SLS, and Univariate Optimal Design for a Completely Randomized Causal Structure:

*1)++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
++;
*yhat not y1hat=E(y1)=8x1;
*Fedorov design;
* 3SLS OPTIMAL DESIGN BASED ON THE 1ST SIMULATION AND OPTIMA DESIGN
"i";
*This Program Gives an Optimal Design for Causal Structure Equations;
PROC IML;RESET;reset;*This to loop and loop;
*3SLS and FIML OPTIMAL DESIGN;
*x={
  1 0,
  1 0,
  1 1,
  1 1,
  1 0,
  1 1,
  1 1,
  1 0,
  1 1,
  0 1,
  1 0,
  1 1,
  1 0,
  1 1,
  0 1,
  1 0,
  0 1,
  1 0,
  1 1,
  1 1,
  0 1,
  1 0,
  1 1,
  1 0,
  1 0};
*Traditional OPTIMAL DESIGN;
  x={
    0 1,
    1 0,
    1 1,
    0 1,
    1 0,
    1 1,
    0 1,
    1 0,
    1 1,
    0 1,
    1 0,
    1 1,
    1 0,
    1 0,
    1 1,
    0 1,
    1 0,
    1 1,
    0 1,
    1 0,
    1 1,
    0 1,
    1 0,
    1 1,
    0 1,
    1 0};
sigma={1 5,5 26};
\begin{verbatim}
Isigma=inv(sigma);
y1hat=8*x[,1];
r=nrow(x);
z11=x[,1];
z12=0*j(r,2);
z21=0*j(r,1);
z22=resp[,1]||x[,2];
z22=y1hat||x[,2];
z=(z11||z12)//(z21||z22);
px=x`*x;
c=det(x`*x);
Ipx = Inv(xpx);
IpxI= x*Ipx*x`;
D = det(z`*(Isigma@IpxI)*z);

print px D;
end;
quit;
\end{verbatim}
**1 a)**
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
+++
**Fedov design;**
**3SLS OPTIMAL DESIGN BASED ON THE 1ST SIMULATION AND OPTIMA DESIGN**
*i*;
**This Program Gives an Optimal Design for Causal Structure Equations:**
PROC IML; *This to loop and loop;
**3SLS and FIML OPTIMAL DESIGN 1.1 and 1.2;**
x={
 1  0,
 1  0,
 1  1,
 1  1,
 1  0,
 1  1,
 1  1,
 1  0,
 1  1,
 0  1,
 1  0,
 1  1,
 1  0,
 1  1,
 1  1,
 1  1,
 1  1,
 0  1,
 1  0,
 1  1,
 1  0,
 1  1,
 1  0,
 1  1};

*yhat not y1hat=E(y1)=8x1*;
**Traditional OPTIMAL DESIGN;**
x={
 0  1,
 1  0,
 1  1,
 0  1,
 1  0,
 1  1,
 0  1,
 1  0,
 1  1,
 0  1,
 1  0,
 1  1,
 1  0,
 1  1,
 0  1,
 1  0,
 1  1,
 0  1,
 1  0,
 1  1,
 0  1,
 1  0,
 1  1,
 0  1,
 1  0,
 1  1,
 0  1,
 1  0,
 1  1,
 0  1,
 1  0};
sigma=(1  0,0  1);
Isigma=inv(sigma);
y1hat=8*x[,1];
r=nrow(x);
z11=x[,1];
z12=0*j(r,2);
z21=0*j(r,1);

z22=y1hat||x[,2];
z=(z11||z12)//(z21||z22);
px=x`*x;
c=det(x`*x);
Ipx = Inv(xpx);
IxpxI= x*Ipx*x`;
D = det(z`*(Isigma@IxpxI)*z);

print xpx D;
end;
quit;
*1 b)

+---------------------------------------------------------------+
*yhat not y1hat=E(y1)=8x1;
+Fedorov design;
* 3SLS OPTIMAL DESIGN BASED ON THE 1ST SIMULATION AND OPTIMA DESIGN
"i":
*This Program Gives an Optimal Design for Causal Structure Equations;
PROC IML; RESET; reset; *This to loop and loop;
*3SLS and FIML OPTIMAL DESIGN 1.1 and 1.2;
*x=
1 0,
1 0,
1 1,
1 1,
1 0,
1 1,
1 1,
1 0,
1 1,
1 1,
1 0,
1 1,
1 1,
1 0,
1 1,
1 1,
1 0,
1 1,
1 1,
...

*Traditional OPTIMAL DESIGN;

x=
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 1,
0 1,
1 0,
1 0,

sigma=(1 0, 0 1); *Var(E);
I.sigma=inv(sigma);
   y1hat=7.9056905*x[,1];
\texttt{r=nrow(x);}
\texttt{z11=x[,1];}
\texttt{z12=0*j(r,2);}
\texttt{z21=0*j(r,1);}

\texttt{z22=y1hat||x[,2];}
\texttt{z=(z11||z12)/(z21||z22);}
\texttt{xpx=x\textasciitilde{x};}
\texttt{c=det(x\textasciitilde{x});}
\texttt{Ixpx = Inv(xpx);}
\texttt{IxpxI= x*Ixpx*x\textasciitilde{};}
\texttt{D = det(z\textasciitilde{}*(I\sigma@IxpxI)*z);}

\texttt{\textbf{print} xpx D;}
\texttt{end;}
\texttt{\textbf{quit};}
*1 c)
+------------------------------------------------------------------+
yhat not y1hat=E(y1)=8x1;
*Fedorov design;
* 3SLS OPTIMAL DESIGN BASED ON THE 1ST SIMULATION AND OPTIMA DESIGN
"i";
*This Program Gives an Optimal Design for Causal Structure Equations;
PROC IML;RESET;reset;*This to loop and loop;
*3SLS and FIML OPTIMAL DESIGN 1.1 and 1.2;
*x=
 1  0,
 1  0,
 1  1,
 1  1,
 1  0,
 1  1,
 1  1,
 1  0,
 1  1,
 0  1,
 1  0,
 1  1,
 1  0,
 1  1,
 1  1,
 1  1,
 0  1,
 1  0,
 1  1,
 1  0,
 1  1,
 0  1,
 1  0,
 1  1,
 1  0,
 1  1,
 0  1;

*y1hat=8.2479385*x[,1];

sigma=(1 0,0 1);*Var(E);
Isigma=inv(sigma);
ylhat=8.2479385*x[,1];
r=nrow(x);  
z11=x[,]1];  
z12=0*j(r,2);  
z21=0*j(r,1);  
z22=y1hat|x[ ,2];  

z=(z11||z12)/(z21|z22);  
px=x`*x;  
c=det(x`*x);  
Ixpx = Inv(xpx);  
IxpxI = x*Ixpx*x`;  
D = det(z`*(Isgma@IxpxI)*z);  

print xpx D;  
end;  
quit;
This Program Gives an Optimal Design for Causal Structure Equations;

PROC IML; RESET; reset; *This to loop and loop;
*3SLS and FIML OPTIMAL DESIGN 1.1 and 1.2;
*x={
  1 0, 1 0, 1 1, 1 1, 1 0, 1 1, 1 1, 1 0, 1 0, 1 1, 1 0, 0 1, 1 0, 1 0, 1 1, 1 0, 1 1, 1 1, 1 1, 1 1, 0 1, 0 1, 0 0, 1 1, 1 0, 1 0, 1 1, 0 1, 1 0, 1 0, 1 1, 0 1, 1 1, 0 1, 1 1, 1 1, 0 1, 1 0, 1 1, 1 1, 1 1, 1 1, 0 1, 1 0, 1 0, 1 1, 1 0, 1 1, 1 1, 0 1, 1 0, 1 1, 1 0, 1 0, 1 1, 1 0, 1 0, 1 0, 1 0};

*Traditional OPTIMAL DESIGN;
{x={
  0 1, 1 0, 1 1, 0 1, 1 0, 1 1, 0 1, 1 0, 1 1, 0 1, 1 0, 1 0, 1 1, 1 0, 1 0, 1 1, 0 1, 1 0, 1 0, 1 1, 0 1, 1 0, 1 0, 1 1, 0 1, 1 0, 1 0, 1 1, 0 1, 1 0, 1 0, 1 0, 1 0, 1 0, 1 0, 1 0, 1 0, 1 0, 1 0, 1 0, 1 0, 1 0, 1 0);

sigma=(1 0, 0 1);
Isigma=inv(sigma);
yhat=8.3388438*x[,1];
r=nrow(x);
z11=x[,1];
z12=0*j(r,2);
z21=0*j(r,1);

z22=y1hat||x[,2];
z=(z11||z12)/(z21||z22);

xpx=x`*x;
c=det(x`*x);
Ixpx = Inv(xpx);
IxpxI = x*Ixpx*x`;
D = det(z`*(Isigma@IxpxI)*z);

print xpx D;
end;
quit;
A 3SLS Optimal Design for Four Blocks with Block Size Equal to Four:

```plaintext
proc IML; RESET;
CP1={-1 -1 -1}; CP2={-1 -1 1}; CP3={-1 1 -1}; CP4={-1 1 1};
CP5={1 -1 -1}; CP6={1 -1 1}; CP7={1 1 -1}; CP8={1 1 1};
CanP={-1 -1 -1, -1 -1 1, -1 1 -1, -1 1 1, 1 -1 -1, 1 -1 1, 1 1 -1, 1 1 1};
s=0; D1=-1; D2=1; D3=-1; D4=-1; D=1; DT=0; Zaher=0;
y21={34.716418, -48.473, 51.057367, -25.94062}; y22={60.810192, 20.172189, 16.761147, -54.70027};
y23={-50.40517, -62.07898, 47.412563, -34.01863}; y24={33.680282, 62.268266, 52.233924, -25.95692};
Alpha=.5;
y1=y11//y12//y13//y14;
y2=y21//y22//y23//y24;
y=y1//y2;
D2=-1;

do l =1 to 5000;
prob = {0.5 0.5 0.5};
p = repeat(prob, 16);
call streaminit(-1);
x = 2*rand("Bernoulli", p)-J(16, 3, 1);
D=-10; D1=0;
diff=10;
z=I(4)@J(4, 1);
diff=10;
do i =1 to 300;
sx=x;
do k =1 to 16;
idx = setdif(1:nrow(sx),k);
u = sx[idx, ];
do n =1 to 8;
x=insert(u,CanP[n,],k,0);

xpz=x`*z;
px=x`*x;
Zstar=I(2)@xpz;Zstarp=Zstar`;

*Constructing the variance ++++++++++++++++++++++++;
G11=.25*I(4); G22=.4*I(4); G12=0*I(4);
G=(G11||G12)//(G12||G22);

Sigma=I(2);
v=Zstar*G*Zstar`+Sigma@px;
if det(v)=0 then
  vinv=J(6, 6, 0);
else
  vinv=Inv(v);

w=(x[,1]|x[,3]|J(16,2,0))//(J(16,2,0)|x[,2]|y1);
Ixpw=(I(2)@x`)*w;
Ixpwp=Ixpw`;
```
\[ \text{InfMat} = \text{Ixpwp}^* \text{inv} \text{Ixpw}; \]
\[
\text{c} = \text{det}(\text{InfMat});
\]
\[
\text{if } \text{c} > \text{D} \text{ then } \text{DM} = \text{x};
\]
\[
\text{if } \text{c} > \text{D} \text{ then } \text{D} = \text{c};
\]
\[
\text{end;}
\]
\[
\text{x} = \text{DM};
\]
\[
\text{diff} = \text{D} - \text{D1};
\]
\[
\text{end;}
\]
\[
\text{if } \text{diff} = 0 \text{ then } \text{i} = 300;
\]
\[
\text{D1} = \text{D};
\]
\[
\text{end;}
\]
\[
\text{if } \text{D1} > \text{D2} \text{ then } \text{xfinal1} = \text{x};
\]
\[
\text{if } \text{D1} > \text{D2} \text{ then } \text{D2} = \text{D1};
\]
\[
\text{end;}
\]
\[
\text{x} = \text{xfinal1};
\]
\[
\text{xp} = \text{x}^* \text{x};
\]
\[
\text{print} \text{ xpx;}
\]
\[
\text{x1} = \text{x}[1, ]; \text{x2} = \text{x}[2, ]; \text{x3} = \text{x}[3, ]; \text{x4} = \text{x}[4, ]; \text{x5} = \text{x}[5, ]; \text{x6} = \text{x}[6, ]; \text{x7} = \text{x}[7, ]; \text{x8} = \text{x}[8, ]; \text{x9} = \text{x}[9, ]; \text{x10} = \text{x}[10, ]; \text{x11} = \text{x}[11, ]; \text{x12} = \text{x}[12, ]; \text{x13} = \text{x}[13, ]; \text{x14} = \text{x}[14, ]; \text{x15} = \text{x}[15, ]; \text{x16} = \text{x}[16, ];
\]
\[
\text{BK1} = \text{x1} /// \text{x2} /// \text{x3} /// \text{x4}; \text{BK2} = \text{x5} /// \text{x6} /// \text{x7} /// \text{x8}; \text{BK3} = \text{x9} /// \text{x10} /// \text{x11} /// \text{x12}; \text{BK4} = \text{x13} /// \text{x14} /// \text{x15} /// \text{x16};
\]
\[
\text{print} \text{ BK1 BK2 BK3 BK4 D2;}
\]
\[
\text{quit};
\]
\[
\text{run;}
\]
A 3SLS Optimal Design for Four Blocks with Block Size Equal to Four:

```plaintext
proc IML; RESET;
CP1={-1 -1 -1}; CP2={-1 1 -1}; CP3={-1 1 -1}; CP4={-1 1 1};
CP5={1 -1 -1}; CP6={1 -1 1}; CP7={1 1 -1}; CP8={1 1 1};
CanP={-1 -1 -1, -1 -1 1, -1 1 -1, -1 1 1, 1 -1 -1, 1 -1 1, 1 1 -1, 1 1 1};
s=0; D1=-1; D2=-1; D3=-1; D4=-1; DT=0; Zahe=0;
y21={34.716418, -48.473, 51.057367, -25.94062}; y22={60.810192, 20.172189, 16.761147, -54.70027};
y23={-50.40517, -62.07898, 47.412563, -34.01863}; y24={33.680282, 62.268266, 52.233924, -25.95692};
Alpha=0.5;
y1=y11//y12//y13//y14;
y2=y21//y22//y23//y24;
y=y1//y2;
D2=-1;

do l =1 to 500;
  prob = (0.5 0.5 0.5);
  p = repeat(prob, 16);
call streaminit(-1);
x = 2*rand("Bernoulli", p)-J(16,3,1);
D=-10; Dl=0;
diff=10;
  z=I(4)@J(4,1);
    diff=10;
do i =1 to 300;
sx=x;
do k =1 to 16;
  idx = setdif(1:nrow(sx),k);
  u = sx[idx, ];
do n =1 to 8;
  x=insert(u,CanP[n,],k,0);
xpz=x`*z;
xpx=x`*x;
Zstar=I(2)@xpz; Zstarp=Zstar`;
*Constructing the variance ++++++++++++++++++++++++;
G11=.25*I(4); G22=.4*I(4); G12=0*I(4);
G=(G11||G12)//(G12||G22);
Sigma=I(2);
v=Zstar*G*Zstar`+Sigma@xpx;
if det(v)=0 then
  vinv=J(6,6,0);
else
  vinv=Inv(v);
y1hat=8*x[,1]+3*x[,3];
w=(x[,1]|x[,3]|J(16,2,0)||(J(16,2,0)||x[,2]|y1hat));
Ixpw=(I(2)@x`)*w;
```

\[ I_{xpwp} = I_{xpw}^\prime; \]
\[ InfMat = I_{xpwp} * \text{inv} * I_{xpw}; \]
\[ c = \text{det}(InfMat); \]
\[ \text{if } c > D \text{ then } DM = x; \]
\[ \text{if } c > D \text{ then } D = c; \]
\[ \text{end}; \]
\[ x = DM; \]
\[ \text{diff} = D - D1; \]
\[ \text{end}; \]
\[ \text{if } \text{diff} = 0 \text{ then } i = 300; \]
\[ D1 = D; \]
\[ \text{end}; \]
\[ \text{if } D1 > D2 \text{ then } x_{\text{final1}} = x; \]
\[ \text{if } D1 > D2 \text{ then } D2 = D1; \]
\[ \text{end}; \]
\[ x = x_{\text{final1}}; \]
\[ x_{px} = x^\prime * x; \]
\[ \text{print } x_{px}; \]
\[ x1 = x[1, ]; x2 = x[2, ]; x3 = x[3, ]; x4 = x[4, ]; x5 = x[5, ]; x6 = x[6, ]; x7 = x[7, ]; x8 = x[8, ]; \]
\[ x9 = x[9, ]; x10 = x[10, ]; x11 = x[11, ]; x12 = x[12, ]; x13 = x[13, ]; x14 = x[14, ]; x15 = x[15, ]; x16 = x[16, ]; \]
\[ BK1 = x1//x2//x3//x4; BK2 = x5//x6//x7//x8; BK3 = x9//x10//x11//x12; BK4 = x13//x14//x15//x16; \]
\[ \text{print } BK1 \ BK2 \ BK3 \ BK4 \ D2; \]
\[ \text{quit}; \]
\[ \text{run}; \]
Comparing Determinates of the Optimal Design and the Orthogonal Design:

```plaintext
proc IML; RESET;
y21={34.716418, -48.473, 51.057367, -25.94062}; y22={60.810192, 20.172189, 16.761147, -54.70027}; y23={-50.40517, -62.07898, 47.412563, -34.01863}; y24={33.680282, 62.268266, 52.233924, -25.95692};
y1=y11//y12//y13//y14;
y2=y21//y22//y23//y24;
y=y1//y2;

*Orthogonal Design;
x = {-1 1 1 -1 1 1 -1,-1 -1 -1, 1 1 1 -1 1 1 -1 -1 -1, -1 1 1 -1 1 1 -1 -1 -1};

*Optimal Design;
x = {-1 -1 -1 -1 -1 1 1 1 -1, 1 -1 -1 1 1 -1 1, 1 1 -1 1 1 1 -1 1 1 -1, 1 -1 1 1 1 1 -1 1 1 1 -1};

z=I(4)@J(4,1);
sx=x;
xpz=x`*z;
xpx=x`*x;
Zstar=I(2)@xpz;Zstarp=Zstar`;

*Constructing the variance ++++++++++++++++++++++++++++++++;
G11=.25*I(4);G22=.4*I(4);G12=0*I(4);
G=(G11||G12)//(G12||G22);
Sigma=I(2);
v=Zstar*G*Zstar`+Sigma@xpx;
vinv=Inv(v);

w=(x[,1]||x[,3]||J(16,2,0))//(J(16,2,0)||x[,2]||y1);
Ixpwp=I(2)@x`)*w;
InfMat=Ixpwp*vinv*Ixpwp;
c=det(InfMat);
print xpx c;
quit;
run;
```
Comparing Determinates of the Optimal Design and the Orthogonal Design for 100 Runs:

*++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
++++;                          *ORTHOGONAL DESIGN V.S OPTIMAL DESIGN We need to generate the Y-
values;
PROC IML; reset;
u11=0.3038357; u12=-0.084872; u13=0.2785379; u14=-0.00894;
u21=0.0651246; u22=-0.786024; u23=-0.548311; u24=-0.266538;
ioth=0;iopt=0;itie=0;
do i=1 to 100;
  x1=({-1 1 1,
       1 -1 1,
       1 1 -1,
       -1 -1 -1});
  x2=({
       1 -1 -1,
       -1 -1 1,
       1 1 1,
       -1 1 -1});
  x3=({
       -1 1 -1,
       1 -1 -1,
       -1 -1 1,
       1 1 1});
  x4=({
       -1 -1 -1,
       1 1 1,
       -1 1 1,
       1 1 -1});
*y1=y11//y12//y13//y14;  y2=y21//y22//y23//y24;
y=x=x1//x2//x3//x4;
*xpz=x`*z;
\[ \text{xpx} = \text{x}^\top \text{x}; \]
\[ \text{Zstar} = \text{I(2)} @ \text{xpz}; \]

*Constructing the variance ++++++++++++++++++++++++++++++++;
\[ \text{G11} = 0.0625 \times \text{I(4)}; \]
\[ \text{G22} = 0 \times \text{I(4)}; \]
\[ \text{G12} = 0.16 \times \text{I(4)}; \]
\[ \text{G} = (\text{G11} || \text{G12}) \left/ (\text{G12} || \text{G22}) \right.; \]
\[ \text{sigma} = \text{I(2)}; \]
\[ \text{v} = \text{Zstar} \times \text{G} \times \text{Zstar}^\top + \text{sigma} @ \text{xpx}; \]
\[ \text{vinv} = \text{Inv(v)}; \]
\[ \text{w} = (\text{x}[;1]) || \text{x}[;3] || \text{J(16,2,0)} \times (\text{J(16,2,0)} || \text{x}[;2] || \text{y1}); \]
\[ \text{Ixpw} = (\text{I(2)} @ \text{x}^\top)^\top \times \text{w}; \]
\[ \text{InfMat} = \text{Ixpwp} \times \text{vinv} \times \text{Ixpw}; \]
\[ \text{InfMatinv} = \text{Inv(InfMat)}; \]
\[ \text{xpx} = \text{x}^\top \times \text{x}; \]
\[ \text{Dorth} = \text{DET(InfMat)}; \]

\[ \text{x1} = (\text{1} -1 -1, \text{1} -1 1, \text{1} 1 -1, \text{1} 1 1); \]
\[ \text{x2} = (\text{1} 1 -1, \text{1} -1 -1, \text{1} -1 1, \text{1} -1 1); \]
\[ \text{x3} = (\text{1} 1 1, \text{1} -1 1, \text{1} -1 -1, \text{1} 1 -1); \]
\[ \text{x4} = (\text{1} -1 1, \text{1} 1 -1, \text{1} -1 -1, \text{1} -1 1); \]

*1+++++++++++++++++++++++++++++++;
\[ \text{y1} = 8 \times \text{x1}^\top + 3 \times \text{x3}^\top + \text{u11} \times \text{J(4,1)} + \text{RANNOR}(\text{J(4,1,-1)}); \]
\[ \text{y2} = 5 \times \text{y1} + 2 \times \text{x21}^\top + \text{u12} \times \text{J(4,1)} + \text{RANNOR}(\text{J(4,1,-1)}); \]
*+++++++++++++++++++++++++++++++;
\[ \text{y1} = 8 \times \text{x1}^\top + 3 \times \text{x3}^\top + \text{u12} \times \text{J(4,1)} + \text{RANNOR}(\text{J(4,1,-1)}); \]
\[ \text{y2} = 5 \times \text{y1} + 2 \times \text{x21}^\top + \text{u22} \times \text{J(4,1)} + \text{RANNOR}(\text{J(4,1,-1)}); \]
*+++++++++++++++++++++++++++++++;
\[ \text{y1} = 8 \times \text{x1}^\top + 3 \times \text{x3}^\top + \text{u13} \times \text{J(4,1)} + \text{RANNOR}(\text{J(4,1,-1)}); \]
\[ \text{y2} = 5 \times \text{y1} + 2 \times \text{x23}^\top + \text{u23} \times \text{J(4,1)} + \text{RANNOR}(\text{J(4,1,-1)}); \]
*+++++++++++++++++++++++++++++++;
\[ \text{y1} = 8 \times \text{x1}^\top + 3 \times \text{x3}^\top + \text{u14} \times \text{J(4,1)} + \text{RANNOR}(\text{J(4,1,-1)}); \]
\[ \text{y2} = 5 \times \text{y1} + 2 \times \text{x24}^\top + \text{u24} \times \text{J(4,1)} + \text{RANNOR}(\text{J(4,1,-1)}); \]
*+++++++++++++++++++++++++++++++;
```plaintext
y1 = y1 // y1 // y1 // y14;
y2 = y2 // y2 // y2 // y24;
y = y1 // y2;

z = I(4) @ J(4, 1);
x = x1 // x2 // x3 // x4;
xpz = x` * z;
xpx = x` * x;
Zstar = I(2) @ xpz;

* Constructing the variance ++++++++++++++++++++++++++++++++;
G11 = .0625 * I(4); G22 = 0 * I(4); G12 = .16 * I(4);
G = (G11 || G12) // (G12 || G22);

Sigma = I(2);
v = Zstar * G * Zstar` + Sigma * xpx;
vinv = Inv(v);
w = (x[ , 1] || x[ , 3] || J(16, 2, 0) // J(16, 2, 0) || x[ , 2] || y1);
IXpw = (I(2) @ x`) * w;
IXwp = IXpw`;
InfMat = IXwp * vinv * IXpw;
InfMatinv = Inv(InfMat);
xpx = x` * x;
Dopt = DET(InfMat);
if Dopt > Dorth then
  iopt = iopt + 1;
if Dorth > Dopt then
  iorth = iorth + 1;
if Dopt = Dorth then
  itie = itie + 1;
end;
Print iorth iopt itie;
quit;
run;
```
Comparing Determinates of the Optimal Design and the Orthogonal Design for 1,000 Runs:

*ORTOGONAL DESIGN V.S OPTIMAL DESIGN We need to generate the Y-values;
PROC IML; reset;
u11=0.3038357; u12=-0.084872; u13=0.2785379; u14=-0.00894;
u21=0.0651246; u22=-0.786024; u23=-0.548311; u24=-0.266538;
iorth=0; iopt=0; itie=0;
do i=1 to 1000;
x11=x1[,1]; x21=x1[,2]; x31=x1[,3];
x12=x2[,1]; x22=x2[,2]; x32=x2[,3];
x13=x3[,1]; x23=x3[,2]; x33=x3[,3];
x14=x4[,1]; x24=x4[,2]; x34=x4[,3];
y11=8*x11+3*x31+u11*J(4,1)+RANNOR(J(4,1,-1));
y21=5*y11+2*x21+u12*J(4,1)+RANNOR(J(4,1,-1));
y12=8*x12+3*x32+u12*J(4,1)+RANNOR(J(4,1,-1));
y22=5*y12+2*x22+u22*J(4,1)+RANNOR(J(4,1,-1));
y13=8*x13+3*x33+u13*J(4,1)+RANNOR(J(4,1,-1));
y23=5*y13+2*x23+u23*J(4,1)+RANNOR(J(4,1,-1));
y14=8*x14+3*x34+u14*J(4,1)+RANNOR(J(4,1,-1));
y24=5*y14+2*x24+u24*J(4,1)+RANNOR(J(4,1,-1));
y1=y11//y12//y13//y14;
y2=y21//y22//y23//y24;
y=y1//y2;
z=I(4)@J(4,1);
x=x1//x2//x3//x4;
kpz=x`*z;
xpx=x`*x;
Zstar=I(2)@xpz;
*Constructing the variance +++++++++++++++++++++++++++++++++++++++;
G11=.0625*I(4);G22=0*I(4);G12=.16*I(4);
G=(G11||G12)//(G12||G22);

Sigma=I(2);

v=Zstar*G*Zstar`+Sigma@xpx;
vinv=Inv(v);

w=(x[,1]||x[,3]||J(16,2,0))//(J(16,2,0)||x[,2]||y1);
Ixpw=(I(2)@x`)*w;
Ixpwp=Ixpw`;
InfMat=Ixpwp*vinv*Ixpw;
InfMatinv=Inv(InfMat);

ixpw=(I(2)@x`)*w;

Dorth=DET(InfMat);

*estimating ++++++++++++++++++++++++++++++++++++;
*Bhat=InfMatinv*Ixpwp*vinv*(I(2)@x`)*y;

x1={
  1 1 1,
  -1 -1 -1,
  1 -1 -1,
  -1 1 1};
x2={
  1 1 1,
  -1 -1 -1,
  -1 1 -1,
  1 -1 1};
x3={
  -1 1 -1,
  -1 -1 1,
  1 -1 1,
  1 1 -1};
x4={
  -1 -1 -1,
  -1 1 1,
  1 -1 1,
  1 1 -1};

*y1++++++++++++++++++++++++++++++++++++++;
y11=8*x11+3*x31+u11*J(4,1)+RANNOR(J(4,1,-1));
y21=5*y11+2*x21+u12*J(4,1)+RANNOR(J(4,1,-1));
*y2++++++++++++++++++++++++++++++++++++++;
y12=8*x12+3*x32+u12*J(4,1)+RANNOR(J(4,1,-1));
y22=5*y12+2*x22+u22*J(4,1)+RANNOR(J(4,1,-1));
*y3++++++++++++++++++++++++++++++++++++++;
y13=8*x13+3*x33+u13*J(4,1)+RANNOR(J(4,1,-1));
y23=5*y13+2*x23+u23*J(4,1)+RANNOR(J(4,1,-1));
*y4++++++++++++++++++++++++++++++++++++++;
y14=8*x14+3*x34+u14*J(4,1)+RANNOR(J(4,1,-1));
y24=5*y14+2*x24+u24*J(4,1)+RANNOR(J(4,1,-1));
\[ y_1 = y_{11} / y_{12} / y_{13} / y_{14}; \]
\[ y_2 = y_{21} / y_{22} / y_{23} / y_{24}; \]
\[ y = y_1 / y_2; \]
\[ z = I(4) @ J(4, 1); \]
\[ x_1 = x_1 / x_2 / x_3 / x_4; \]
\[ xpz = x` * z; \]
\[ xpx = x` * x; \]
\[ Z_{\text{star}} = I(2) @ xpz; \]

*Constructing the variance ++++++++++++++++++++++++++++++++*

\[ G_{11} = 0.0625 * I(4); G_{12} = 0 * I(4); G_{12} = 0.16 * I(4); \]
\[ G = (G_{11} || G_{12}) / (G_{12} || G_{22}); \]

\[ \Sigma = I(2); \]

\[ v = Z_{\text{star}} * G * Z_{\text{star}}` + \Sigma @ xpx; \]
\[ vinv = \text{Inv}(v); \]
\[ w = (x[ ,1] || x[ ,3] || J(16, 2, 0)) / (J(16, 2, 0) || x[ ,2] || y_1); \]
\[ Ixpw = (I(2) @ x`) * w; \]
\[ Ixpwp = Ixpw`; \]
\[ \text{InfMat} = Ixpwp * vinv * Ixpw; \]
\[ \text{InfMatinv} = \text{Inv}(\text{InfMat}); \]
\[ xpx = x` * x; \]
\[ D_{\text{opt}} = \text{DET}(\text{InfMat}); \]

*estimating ++++++++++++++++++++++++++++++++++++++++++++++++++++

\[ \text{if} \ D_{\text{opt}} > D_{\text{orth}} \text{ then} \]
\[ \text{iort} = \text{iort} + 1; \]

\[ \text{if} \ D_{\text{orth}} > D_{\text{opt}} \text{ then} \]
\[ \text{iort} = \text{iort} + 1; \]

\[ \text{if} \ D_{\text{opt}} = D_{\text{orth}} \text{ then} \]
\[ \text{itie} = \text{itie} + 1; \]
end;

Print iort iopt itie;
quit;
run;
FIML Optimal Design for Four Blocks with Block Size Equal to Four:

```plaintext
proc IML; reset;
CP1={-1 -1 -1}; CP2={-1 -1 1}; CP3={-1 1 -1}; CP4={-1 1 1};
CP5={1 -1 -1}; CP6={1 -1 1}; CP7={1 1 -1}; CP8={1 1 1};
CanP={-1 -1 -1, -1 1 1, -1 1 1, -1 1 1}; s=0; D1=-1; D2=-1; D3=-1; D4=-1; DT=0; Zaher=0;
y21={34.716418, -48.473, 51.057367, -25.94062}; y22={60.810192, 20.172189, 16.761147, -54.70027};
y23={-50.40517, -62.07898, 47.412563, -34.01863}; y24={33.680282, 62.268266, 52.233924, -25.95692};
Alpha=0;
y1=y11//y12//y13//y14;
y2=y21//y22//y23//y24;
y2=y2-5*y1;
y=y1//y2;
D2=-1;
B={1 -5, 0 1};
BPRIME=B`;
BprINV=Inv(BPRIME);
Bstar=I(16)@BprINV;
BstarP=Bstar`;
do l =1 to 50;
prob = {0.5 0.5 0.5};
p = repeat(prob, 16);
call streaminit(-1);
x = 2*rand("Bernoulli", p)-J(16,3,1);
D=-10; D1=0;
diff=10;
z=I(4)@J(4,1,1);
Zstar=z@I(2);
ZstarP=Zstar`;
do i =1 to 300;
sx=x;
do k =1 to 16;
idx = setdif(1:nrow(sx),k);
u = sx[idx, ];
do n =1 to 8;
x=insert(u,CanP[n,],k,0);
ylhat=8*x[,1]+3*x[,3];
Xstar=x@I(2);
XstarP=Xstar`;
*for fixed parameters;
zf=I(4)@J(4,1,1);
Zstarf=z@zf;
ZstarfP=Zstarf`;
SigmaUf={.25 0 0 .4};
```

Gf=SigmaUf@I(4);
Rf=I(2)@I(16);

vf=(Zstarf*Gf*Zstarf`+Rf);

vfI=Inv(vf);
v11=vf[1:16,1:16];
Iv11=vfI[1:16,1:16];
Iv21=vfI[17:32,1:16];
Iv22=vfI[17:32,17:32];
x1=x[ ,1];x2=x[ ,2];x3=x[ ,3];
v11Iv22=v11*Iv22;

*Information Matrix for fixed parameters"endog and exog";
M11=x1`*Iv11*x1;M12=x1`*Iv11*x3;M13=x1`*Iv12*x2;M14=x1`*Iv12*y1hat;
M21=x3`*Iv11*x1;M22=x3`*Iv11*x3;M23=x3`*Iv12*x2;M24=x3`*Iv12*y1hat;
M31=x2`*Iv21*x1;M32=x2`*Iv21*x3;M33=x2`*Iv22*x2;M34=x2`*Iv22*y1hat;
M41=y1hat`*Iv21*x1;M42=y1hat`*Iv21*x3;M43=y1hat`*Iv22*x2;M44=TRACE(v11Iv22)+y1hat`*Iv22*y1hat;
MF=(M11||M12||M13||M14)///<
(M21||M22||M23||M24)///<
(M31||M32||M33||M34)///<
(M41||M42||M43||M44);
H1=DET(MF);
InfMat=H1;

if InfMat > D then DM = x;
if InfMat > D then D = InfMat;
end;
x=DM;
diff=D-D1;
end;
if diff=0 then i=300;
D1=D;
end;
if D1 > D2 then xfinal1 = x;
if D1 > D2 then D2 = D1;
end;

x=xfinal1;
xpx=x`*x;
print xpx;
x1=x[1, ];x2=x[2, ];x3=x[3, ];x4=x[4, ];x5=x[5, ];x6=x[6, ];x7=x[7, ];x8=x[8, ];
x9=x[9, ];x10=x[10, ];x11=x[11, ];x12=x[12, ];x13=x[13, ];x14=x[14, ];x15=x[15, ];x16=x[16, ];
BK1 = x1//x2//x3//x4;BK2= x5//x6//x7//x8;BK3 = x9//x10//x11//x12;BK4 = x13//x14//x15//x16;
print BK1 BK2 BK3 BK4 D2;
quit;
run;
**3SLS Optimal Design for a Causal Structure with Fixed Blocks:**

```r
proc IML ;
RESET;
CP1={-1 -1 -1}; CP2={-1 -1 1}; CP3={-1 1 -1}; CP4={-1 1 1};
CP5={1 -1 -1}; CP6={1 -1 1}; CP7={1 1 -1}; CP8={1 1 1};
CanP={-1 -1 -1, -1 -1 1, -1 1 -1, -1 1 1, 1 -1 -1, 1 -1 1, 1 1 -1, 1 1 1};
s=0; D1=-1; D2=-1; D3=-1; D4=-1; D=0; D^T=0; Zaheer=0;
D2=-1;

do l =1 to 50;
prob = {0.5 0.5 0.5};
p = repeat(prob, 16);
call streaminit(-1);
x = 2*rand("Bernoulli", p)-J(16,3,1);
D=-10; D1=0;
diff=10;
z=I(4)@J(4,1,1);
Zstar=z@I(2);
ZstarP=Zstar^`;
do i =1 to 300;
sx=x;
do k =1 to 16;
idx = setdif(1:nrow(sx),k);
u = sx[idx, ];
do n =1 to 8;
x=insert(u,CanP[n,],k,0);
y1hat=8*x[,1]+3*x[,3];
what=(x[,1]|x[,3]|J(16,2,0))/(J(16,2,0)|x[,2]|y1hat);
whatP=what^`;
Xstar=x@I(2);
XstarP=Xstar^`;

*for fixed parameters;
zf=I(4)@J(4,1,1);
Zstarf=I(2)@zf;
ZstarfP=Zstarf^`;
Sigmaf={1 0,0 1};
Rf=Sigmaf@I(16);
vf=(Rf);
vfI=Inv(vf);
M11=whatP*vfI*what;
M12=whatP*vfI*Zstarf;
M21=M12^`;
M22=ZstarfP*vfI*Zstarf;
M22Inv=Inv(M22);
INF=M11-M12*M22Inv*M21;
H1=DET(INF);
InfMat=H1;
if InfMat > D then DM = x;
if InfMat > D then D = InfMat;
```
x=DM;
diff=D-D1;
end;
if diff=0 then i=300;
D1=D;
end;
if D1 > D2 then xfinal1 = x;
if D1 > D2 then D2 = D1;
end;

x=xfinal1;

xpx=x`*x;
print xpx;

x1=x[1, ]; x2=x[2, ]; x3=x[3, ]; x4=x[4, ]; x5=x[5, ]; x6=x[6, ]; x7=x[7, ]; x8=x[8, ];
x9=x[9, ]; x10=x[10, ]; x11=x[11, ]; x12=x[12, ]; x13=x[13, ]; x14=x[14, ];
x15=x[15, ]; x16=x[16, ];
BK1 = x1//x2//x3//x4; BK2 = x5//x6//x7//x8; BK3 = x9//x10//x11//x12; BK4 = x13//x14//x15//x16;
print BK1 BK2 BK3 BK4 D2;
quit;
run;
FIML Optimal Design for a Causal Structure with Fixed Blocks:

```plaintext
proc IML ;
RESET;
CP1={-1 -1 -1};CP2={-1 -1 1};CP3={-1 1 -1};CP4={-1 1 1};
CP5={1 -1 -1};CP6={1 -1 1};CP7={1 1 -1};CP8={1 1 1};
CanP={-1 -1 -1, -1 1 1, -1 1 -1, -1 1 1, 1 -1 -1, 1 -1 1, 1 1 -1, 1 1 1};
s=0;D1=-1;D2=-1;D3=-1;D4=-1;D=-1;DT=0;Zaher=0;
D2=-1;

do l =1 to 50;
prob = {0.5 0.5 0.5};
p = repeat(prob, 16);
call streaminit(-1);
x = 2*rand("Bernoulli", p)-J(16,3,1);
D=-10; D1=0;
diff=10;
z=I(4)@J(4,1,1);
Zstar=z@I(2);
ZstarP=Zstar`;
do i =1 to 300;
sx=

do k =1 to 16;
idx = setdif(1:nrow(sx),k);
u = sx[idx, ];
do n =1 to 8;
x=insert(u,CanP[n],k,0);
y1hat=8*x[,1]+3*x[,3];
what=(x[,1]||x[,3]||J(16,2,0)||J(16,2,0)||x[,2]||y1hat);
whatP=what``;
Xstar=x@I(2);
XstarP=Xstar``;

*for fixed parameters;
zf=I(4)@J(4,1,1);
Zstarf=I(2)@zf;
ZstarfP=Zstarf`;
Sigmaf={1 0,0 1};
Rf=Sigmaf@I(16);

vf=(Rf);

vfI=Inv(vf);
vll=vf[1:16,1:16];
Iv11=vfI[1:16,1:16];
Iv12=vfI[1:16,17:32];
Iv21=vfI[17:32,1:16];
Iv22=vfI[17:32,17:32];
x1=x[,1];x2=x[,2];x3=x[,3];
vll1v22=vll*Iv22;
*Information Matrix for fixed parameters"endog and exog";
FM11=x1`*Iv11*x1;FM12=x1`*Iv12*x2;FM13=x1`*Iv13*x3;FM14=x1`*Iv12*y1hat;
FM21=x3`*Iv11*x1;FM22=x3`*Iv12*x2;FM23=x3`*Iv13*x3;FM24=x3`*Iv12*y1hat;
FM31=x2`*Iv21*x1;FM32=x2`*Iv22*x2;FM33=x2`*Iv22*x2;FM34=x2`*Iv22*y1hat;
```
FM41=yihat*Iv21*x1; FM42=yihat*Iv21*x3; FM43=yihat*Iv22*x2; FM44=TRACE(v11Iv22)+yihat*Iv22*y1hat;
M11=(FM11||FM12||FM13||FM14)//
(FM21||FM22||FM23||FM24)//
(FM31||FM32||FM33||FM34)//
(FM41||FM42||FM43||FM44);
M21=M12';
M22=ZstarfP*vfI*Zstarf);
M22Inv=Inv(M22);
INF=M11-M12*M22Inv*M21;
H1=DET(INF);
InfMat=H1;
if InfMat > D then DM = x;
   if InfMat > D then D = InfMat;
end;
x=DM;
diff=D-D1;
end;
if diff=0 then i=300;
   D1=D;
end;
if D1 > D2 then xfinal1 = x;
if D1 > D2 then D2 = D1;
end;
x=xfinal1;
xpx=x'*x;
print xpx;
x1=x[1, ];x2=x[2, ];x3=x[3, ];x4=x[4, ];x5=x[5, ];x6=x[6, ];x7=x[7, ];x8=x[8, ];x9=x[9, ];x10=x[10, ];x11=x[11, ];x12=x[12, ];x13=x[13, ];x14=x[14, ];x15=x[15, ];x16=x[16, ];
BK1 = x1//x2//x3//x4; BK2= x5//x6//x7//x8; BK3 = x9//x10//x11//x12; BK4 = x13//x14//x15//x16;
print BK1 BK2 BK3 BK4 D2;
quit;
run;
SAS CODES FOR CHAPTER 4

A 3SLS Optimal Design for Four Blocks with Block Size Equal to Two with $\alpha = 0.5$ with No Contemporaneous Correlation Among the Endogenous Parameters or the Observation Within Block:

```sas
proc IML; RESET;
CanP={-1 -1, 1 0, 0, -1 0, 1 1, 0 -1, 0 1};
s=0; D1=1; D2=1; D3=1; D4=1; DT=0; Zaher=0;
Alpha=.5;
D2=1;
do l =1 to 10;
SEED=-1;
c=j(8, 2, seed);
x=3*uniform(c);
x=ceil(x);
x=x-2*j(8, 2);
D=-10; D1=0;
diff=10;
z=I(2)@J(2,1,1);
Zstar=I(4)@z;
ZstarP=Zstar``;
do i =1 to 300;
sx=x;
do k =1 to 8;
idx = setdif(1:nrow(sx),k);
u = sx[idx, ];
do n =1 to 9;
    x=insert(u,CanP[n,],k,0);
SigmaU={.25 .8 .8 .16};
Sigma=.10 .5 .5 .10;
G=Sigma@I(4);
R=Sigma@I(8);
v=Zstar*G*ZstarP+R;
    vinv=Inv(v);
ylhat=8*x[ ,1];
w=(x[ ,1]||J(8,2,0))/(J(8,1,0)||x[ ,2]||ylhat);
wp=w``;
InfMatFixed=wp*vinv*w;
Xstar=w;
XstarP=Xstar``;
M=wp*vinv*w;
    Minv=Inv(M);
*Constructing N the Information Matrix for the covariance Parameters;
dD1dSu1=(I(4)||J(4,4,0))/(J(4,8,0));
dD1dSu2=(J(4,8,0))/(J(4,4,0)||I(4));
dR1dSl1=(I(8)||J(8,8,0))/(J(8,8,0));
dR1dS2=(J(8,16,0))/(J(8,8,0)||I(8));
P=vinv-vinv*w*Minv*wp*vinv;
```
Q11 = (P*Zstar*dD1dS1*ZstarP) * (P*Zstar*dD1dS1*ZstarP); N11 = 0.5*TRACE(Q11);
Q12 = (P*Zstar*dD1dS1*ZstarP) * (P*Zstar*dD1dS2*ZstarP); N12 = 0.5*TRACE(Q12);
Q13 = P*Zstar*dD1dS1*ZstarP*P*dR1dS1; N13 = 0.5*TRACE(Q13);
Q14 = P*Zstar*dD1dS1*ZstarP*P*dR1dS2; N14 = 0.5*TRACE(Q14);
Q22 = (P*Zstar*dD1dS2*ZstarP) * (P*Zstar*dD1dS2*ZstarP); N22 = 0.5*TRACE(Q22);
Q23 = P*Zstar*dD1dS2*ZstarP*P*dR1dS1; N23 = 0.5*TRACE(Q23);
Q24 = P*Zstar*dD1dS2*ZstarP*P*dR1dS2; N24 = 0.5*TRACE(Q24);
Q33 = (P*dR1dS1)*(P*dR1dS1); N33 = 0.5*TRACE(Q33);
Q34 = (P*dR1dS1)*(P*dR1dS2); N34 = 0.5*TRACE(Q34);
Q44 = (P*dR1dS2)*(P*dR1dS2); N44 = 0.5*TRACE(Q44);
  (N14 || N24 || N34 || N44);
zk = NRAN;
H2 = DET(zk);
H1 = DET(InfMatFixed);
InfMat = (1 - Alpha) / 3 * (LOG(H1)) + Alpha / 4 * (LOG(H2));
  if InfMat > D then DM = x;
  if InfMat > D then D = InfMat;
end;
x = DM;
diff = D - D1;
end;
if diff = 0 then i = 300;
D1 = D;
end;
if D1 > D2 then xfinal1 = x;
if D1 > D2 then D2 = D1;
end;
x = xfinal1;
xpux = x' * x;
print xpux;
BK1 = x[1:2, 1:2]; BK2 = x[3:4, 1:2]; BK3 = x[5:6, 1:2]; BK4 = x[7:8, 1:2];
print BK1 BK2 BK3 BK4 D2;
quit;
run;
A 3SLS Optimal Design for Four Blocks with Block Size Equal to Two with $\alpha = 0.5$ with Contemporaneous Correlation Among the Endogenous Parameters and the Observation Within Block:

```plaintext
proc IML; reset;
Alpha=.5;
CanP={-1 -1 0, -1 1, 1 -1, 1 0, 1 1, 1 0, 1 1};
s=0; D1=-1; D2=-1; D3=-1; D4=-1; D=-1; DT=0; Zaher=0;
Alpha=.5;
D2=-1;
do l =1 to 5;
SEED=-1;
c=j(8, 2, seed);
x=3*uniform(c);
x=ceil(x);
x=x-2*j(8, 2);
D=-10; D1=0;
diff=10;
z=I(2)@J(2, 1, 1);
Zstar=I(4)@z;
ZstarP=Zstar`;
do i =1 to 300;
sx=x;
do k =1 to 8;
idx = setdif(1:nrow(sx), k);
u = sx[idx, ];
do n =1 to 9;
   x=insert(u, CanP[n, ], k, 0);
SigmaU={.25 .8 .8 .16};
Sigma={.10 .5 .5 .10};
G=SigmaU@I(4);
R=Sigma@I(8);
v=Zstar*G*ZstarP+R;

if det(v)=0 then
   winv=J(16,16,0);
else
   winv=Inv(v);
yhat=8*x[, 1];
w=(x[, 1]|J(8, 2, 0))/(J(8, 1, 0)|x[, 2]|yhat);
wp=w`;
Xstar=w;
XstarP=Xstar`;
M=wp*winv*w;
   Minv=Inv(M);
InfMatFixed=wp*winv*w;
dD1dSu1=(I(4)|J(4, 4, 0))/(J(4, 8, 0));
dD1dSu2=(J(4, 8, 0))/(J(4, 4, 0)|I(4));
dD1dSu12=(J(4, 4, 0)|I(4))/(I(4)|J(4, 4, 0));
dR1dS1=(I(8)|J(8, 8, 0))/(J(8, 16, 0));
```

\[ dR_1 dS_2 = (J(8, 16, 0) || I(8)) / (J(8, 8, 0) || I(8)) / (I(8) || J(8, 8, 0)); \]
\[ dR_1 dS_12 = (J(8, 8, 0) || I(8)) / (I(8) || J(8, 8, 0)); \]
\[ P = vinv - vinv * w * Minv * wp * vinv; \]
\[ Q_{11} = (P * Zstar * dD_{1dSu1} * ZstarP) * (P * Zstar * dD_{1dSu1} * ZstarP); N_{11} = 0.5 * \text{TRACE}(Q_{11}); \]
\[ Q_{12} = (P * Zstar * dD_{1dSu1} * ZstarP) * (P * Zstar * dD_{1dSu2} * ZstarP); N_{12} = 0.5 * \text{TRACE}(Q_{12}); \]
\[ Q_{13} = (P * Zstar * dD_{1dSu1} * ZstarP) * (P * Zstar * dD_{1dSu12} * ZstarP); N_{13} = 0.5 * \text{TRACE}(Q_{13}); \]
\[ Q_{14} = (P * Zstar * dD_{1dSu1} * ZstarP) * (P * Zstar * dD_{1dSu12} * ZstarP) * (P * Zstar * dD_{1dSu1} * ZstarP); N_{14} = 0.5 * \text{TRACE}(Q_{14}); \]
\[ Q_{15} = (P * Zstar * dD_{1dSu1} * ZstarP) * (P * Zstar * dD_{1dSu12} * ZstarP) * (P * Zstar * dD_{1dSu2} * ZstarP); N_{15} = 0.5 * \text{TRACE}(Q_{15}); \]
\[ Q_{16} = (P * Zstar * dD_{1dSu1} * ZstarP) * (P * Zstar * dD_{1dSu2} * ZstarP); N_{16} = 0.5 * \text{TRACE}(Q_{16}); \]
\[ Q_{22} = (P * Zstar * dD_{1dSu2} * ZstarP) * (P * Zstar * dD_{1dSu2} * ZstarP); N_{22} = 0.5 * \text{TRACE}(Q_{22}); \]
\[ Q_{23} = (P * Zstar * dD_{1dSu2} * ZstarP) * (P * Zstar * dD_{1dSu12} * ZstarP); N_{23} = 0.5 * \text{TRACE}(Q_{23}); \]
\[ Q_{24} = (P * Zstar * dD_{1dSu2} * ZstarP) * (P * Zstar * dD_{1dSu1} * ZstarP); N_{24} = 0.5 * \text{TRACE}(Q_{24}); \]
\[ Q_{25} = (P * Zstar * dD_{1dSu2} * ZstarP) * (P * Zstar * dD_{1dSu2} * ZstarP); N_{25} = 0.5 * \text{TRACE}(Q_{25}); \]
\[ Q_{26} = (P * Zstar * dD_{1dSu2} * ZstarP) * (P * Zstar * dD_{1dSu1} * ZstarP); N_{26} = 0.5 * \text{TRACE}(Q_{26}); \]
\[ Q_{33} = (P * Zstar * dD_{1dSu12} * ZstarP) * (P * Zstar * dD_{1dSu12} * ZstarP); N_{33} = 0.5 * \text{TRACE}(Q_{33}); \]
\[ Q_{34} = (P * Zstar * dD_{1dSu12} * ZstarP) * (P * Zstar * dD_{1dSu1} * ZstarP); N_{34} = 0.5 * \text{TRACE}(Q_{34}); \]
\[ Q_{35} = (P * Zstar * dD_{1dSu12} * ZstarP) * (P * Zstar * dD_{1dSu2} * ZstarP); N_{35} = 0.5 * \text{TRACE}(Q_{35}); \]
\[ Q_{36} = (P * Zstar * dD_{1dSu12} * ZstarP) * (P * Zstar * dD_{1dSu12} * ZstarP); N_{36} = 0.5 * \text{TRACE}(Q_{36}); \]
\[ Q_{44} = (P * dR_{1dS1}) * (P * dR_{1dS1}); N_{44} = 0.5 * \text{TRACE}(Q_{44}); \]
\[ Q_{45} = (P * dR_{1dS1}) * (P * dR_{1dS2}); N_{45} = 0.5 * \text{TRACE}(Q_{45}); \]
\[ Q_{46} = (P * dR_{1dS1}) * (P * dR_{1dS12}); N_{46} = 0.5 * \text{TRACE}(Q_{46}); \]
\[ Q_{55} = (P * dR_{1dS2}) * (P * dR_{1dS2}); N_{55} = 0.5 * \text{TRACE}(Q_{55}); \]
\[ Q_{56} = (P * dR_{1dS2}) * (P * dR_{1dS12}); N_{56} = 0.5 * \text{TRACE}(Q_{56}); \]
\[ Q_{66} = (P * dR_{1dS12}) * (P * dR_{1dS12}); N_{66} = 0.5 * \text{TRACE}(Q_{66}); \]
\[ NRAN = (N_{11} || N_{12} || N_{13} || N_{14} || N_{15} || N_{16}) / (N_{12} || N_{22} || N_{23} || N_{24} || N_{25} || N_{26}) / (N_{13} || N_{23} || N_{33} || N_{34} || N_{35} || N_{36}) / (N_{14} || N_{24} || N_{34} || N_{44} || N_{45} || N_{46}) / (N_{15} || N_{25} || N_{35} || N_{45} || N_{55} || N_{56}) / (N_{16} || N_{26} || N_{36} || N_{46} || N_{56} || N_{66}); \]
\[ zk = NRAN; \]
\[ H_2 = \text{DET}(zk); \]
\[ H_1 = \text{DET}(\text{InfMatFixed}); \]
\[ H_2 = \text{DET}(\text{NRAN}); \]
\[ \text{InfMat} = (1 - \text{Alpha}) / 3 * (\text{LOG}(H_1)) + \text{Alpha} / 6 * (\text{LOG}(H_2)); \]
\[ \text{if } \text{InfMat} > D \text{ then } DM = x; \]
\[ \text{if } \text{InfMat} > D \text{ then } D = \text{InfMat}; \]
\[ \text{end}; \]
\[ x = DM; \]
\[ \text{if } \text{diff} = 0 \text{ then } i = 300; \]
D1=D;
end;
if D1 > D2 then xfinal1 = x;
if D1 > D2 then D2 = D1;
end;
x=xfinal1;
px=x'*x;
print xpx;
BK1 =x[1:2,1:2 ];BK2 =x[3:4,1:2 ];BK3 =x[5:6,1:2 ];BK4 =x[7:8,1:2 ];
print BK1 BK2 BK3 BK4 D2;
quit;
run;
An ML Optimal Design for Four Blocks with Block Size Equal to Two:

```plaintext
proc IML; reset;
CanP={-1 -1,-1 0,-1 1,1 -1,1 0,1 1,0 -1,0 0,1 1};
s=0;Dl=-1;D2=-1;D3=-1;D4=-1;Dx=0;Zaher=0;
Alpha=.5;
D2=-1;
do l =1 to 5;
SEED=-1;
c=j(8, 2, seed);
x=3*uniform(c);
x=ceil(x);
x=x-2*j(8, 2);
D=-10;D1=0;
diff=10;
z=I(4)@J(2,1,1);
zp=z`; 
zzp=z*zp;
Zstar=I(2)@z;
ZstarP=Zstar`;
do i =1 to 300;
sx=x;
do k =1 to 8;
idx = setdif(1:nrow(sx),k);
u = sx[idx, ];
do n =1 to 9;
    x=insert(u,CanP[n,],k,0);
y1hat=8*x[ ,1];
y2hat=5*y1hat+2*x[ ,2];
y1hatp=y1hat`
x=X0I(2);
x=Xstar`;
SigmaU={.25 .8 .8 .16};
Sigma={.10 .5 .5 .10};
G=SigmaU@I(4);
R=Sigma@I(8);
v=Zstar*G*ZstarP+R;
    vinv=Inv(v);
v11=vinv[1:8,1:8];v12=vinv[1:8,9:16];
v21=vinv[9:16,1:8];v22=vinv[9:16,9:16];
v1=v[1:8,1:8];v12=v[1:8,9:16];
v2=v[9:16,1:8];v22=v[9:16,9:16];
x1=x[ ,1];
x2=x[ ,2];
v11*v22=v11*vi22;
M1=x1`*v11*x1;M2=x1`*v12*x2;M3=x1`*v12*y1hat;
M21=x2`*v21*x1;M22=x2`*v22*x2;M23=x2`*v22*y1hat;
```
\[ M_{31} = y_1 \hat{v}_1 x_1; M_{32} = y_1 \hat{v}_2 x_2; M_{33} = \text{TRACE}(v_1, v_2) + y_1 \hat{v}_2 y_1 \]

\[ \text{InfMatFixed} = (M_{11} \| M_{12} \| M_{13}) / (M_{21} \| M_{22} \| M_{23}) / (M_{31} \| M_{32} \| M_{33}) \]

*Construct The Variance components N;

\[ dD_{1dSu1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \oplus I(4); \]
\[ dD_{1dSu2} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \oplus I(4); \]
\[ dD_{1dSu12} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \oplus I(4); \]
\[ dR_{1dS1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \oplus I(8); \]
\[ dR_{1dS2} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \oplus I(8); \]
\[ dR_{1dS12} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \oplus I(8); \]

*Construct The Correlation components;

\[ G_1 = v \text{inv} \cdot dR_{1dS1} \cdot v \text{inv}; \]
\[ G_{11} = G_1[1:8,1:8]; G_{12} = G_1[1:8,9:16]; \]
\[ G_{21} = G_1[9:16,1:8]; G_{22} = G_1[9:16,9:16] \]

\[ V_{11}G_{112} = V_{11} \cdot G_{112}; \]
\[ V_{12}G_{122} = V_{12} \cdot G_{122}; \]
\[ V_{11}G_{122} = V_{11} \cdot G_{111}; \]

\[ \text{Cor}_1 = \text{TRACE}(V_{11}G_{112}) + y_1 \hat{v}_1 p \cdot G_{112} \cdot y_1 \hat{v}_1 - 8 \cdot y_1 \hat{v}_1 p \cdot G_{112} \cdot x_1 + \text{TRACE}(V_{12}G_{122}) + y_1 \hat{v}_1 p \cdot G_{122} \cdot y_2 \hat{v}_2 - 2 \cdot y_1 \hat{v}_1 p \cdot G_{122} \cdot x_2 - 5 \cdot \text{TRACE}(V_{11}G_{122}) - 5 \cdot y_1 \hat{v}_1 p \cdot G_{122} \cdot y_1 \hat{v}_1; \]

\[ G_2 = v \text{inv} \cdot dR_{1dS2} \cdot v \text{inv}; \]
\[ G_{21} = G_2[1:8,1:8]; G_{22} = G_2[1:8,9:16]; \]
\[ G_{22} = G_2[9:16,1:8]; G_{222} = G_2[9:16,9:16] \]

\[ V_{11}G_{212} = V_{11} \cdot G_{212}; \]
\[ V_{12}G_{222} = V_{12} \cdot G_{222}; \]
\[ V_{11}G_{222} = V_{11} \cdot G_{211}; \]

\[ \text{Cor}_2 = \text{TRACE}(V_{11}G_{212}) + y_1 \hat{v}_1 p \cdot G_{212} \cdot y_1 \hat{v}_1 - 8 \cdot y_1 \hat{v}_1 p \cdot G_{212} \cdot x_1 + \text{TRACE}(V_{12}G_{222}) + y_1 \hat{v}_1 p \cdot G_{222} \cdot y_2 \hat{v}_2 - 2 \cdot y_1 \hat{v}_1 p \cdot G_{222} \cdot x_2 - 5 \cdot \text{TRACE}(V_{11}G_{222}) - 5 \cdot y_1 \hat{v}_1 p \cdot G_{222} \cdot y_1 \hat{v}_1; \]

\[ G_1 = v \text{inv} \cdot dD_{1dSu1} \cdot v \text{inv} \cdot Z^* \cdot dD_{1dSu1} \cdot Z^* \cdot v \text{inv}; \]
\[ G_{11} = G_1[1:8,1:8]; G_{12} = G_1[1:8,9:16]; \]
\[ G_{21} = G_1[9:16,1:8]; G_{22} = G_1[9:16,9:16]; \]

\[ V_{11}G_{112} = V_{11} \cdot G_{112}; \]
\[ V_{12}G_{122} = V_{12} \cdot G_{122}; \]
\[ V_{11}G_{122} = V_{11} \cdot G_{111}; \]

\[ \text{Cor}_{12} = \text{TRACE}(V_{11}G_{112}) + y_1 \hat{v}_1 p \cdot G_{112} \cdot y_1 \hat{v}_1 - 8 \cdot y_1 \hat{v}_1 p \cdot G_{112} \cdot x_1 + \text{TRACE}(V_{12}G_{122}) + y_1 \hat{v}_1 p \cdot G_{122} \cdot y_2 \hat{v}_2 - 2 \cdot y_1 \hat{v}_1 p \cdot G_{122} \cdot x_2 - 5 \cdot \text{TRACE}(V_{11}G_{122}) - 5 \cdot y_1 \hat{v}_1 p \cdot G_{122} \cdot y_1 \hat{v}_1; \]

\[ G_{11} = G_1[1:8,1:8]; G_{112} = G_1[1:8,9:16]; \]
\[ G_{12} = G_1[9:16,1:8]; G_{122} = G_1[9:16,9:16]; \]

\[ V_{11}G_{112} = V_{11} \cdot G_{112}; \]
\[ V_{12}G_{122} = V_{12} \cdot G_{122}; \]
\[ V_{11}G_{122} = V_{11} \cdot G_{111}; \]

\[ \text{Cor}_{12} = \text{TRACE}(V_{11}G_{112}) + y_1 \hat{v}_1 p \cdot G_{112} \cdot y_1 \hat{v}_1 - 8 \cdot y_1 \hat{v}_1 p \cdot G_{112} \cdot x_1 + \text{TRACE}(V_{12}G_{122}) + y_1 \hat{v}_1 p \cdot G_{122} \cdot y_2 \hat{v}_2 - 2 \cdot y_1 \hat{v}_1 p \cdot G_{122} \cdot x_2 - 5 \cdot \text{TRACE}(V_{11}G_{122}) - 5 \cdot y_1 \hat{v}_1 p \cdot G_{122} \cdot y_1 \hat{v}_1; \]

\[ G_{11} = G_1[1:8,1:8]; G_{112} = G_1[1:8,9:16]; \]
\[ G_{12} = G_1[9:16,1:8]; G_{122} = G_1[9:16,9:16]; \]

\[ V_{11}G_{112} = V_{11} \cdot G_{112}; \]
\[ V_{12}G_{122} = V_{12} \cdot G_{122}; \]
\[ V_{11}G_{u12} = V_{11}G_{u11}; \]

\[ \text{Cor}_{u1} = \text{TRACE}(V_{11}G_{u12}) + y_1\hat{p}G_{u12}y_1\hat{h} - 8y_1\hat{p}G_{u12}x_1 + \text{TRACE}(V_{12}G_{u22}) + y_1\hat{p}G_{u12}y_2\hat{h} - 2y_1\hat{p}G_{u22}x_2 - 5\text{TRACE}(V_{11}G_{u22}) - 5y_1\hat{p}G_{u22}y_1\hat{h}; \]

\[ G_{u2} = v_{11}Z_{star}dD_{1d}S_{u2}Z_{starP}v_{11}; \]
\[ G_{u211} = G_{u2}[1:8, 1:8]; G_{u212} = G_{u2}[1:8, 9:16]; \]
\[ G_{u221} = G_{u2}[9:16, 1:8]; G_{u222} = G_{u2}[9:16, 9:16]; \]

\[ V_{11}G_{u12} = V_{11}G_{u21}; \]
\[ V_{12}G_{u22} = V_{12}G_{u22}; \]
\[ V_{11}G_{u22} = V_{11}G_{u21}; \]

\[ \text{Cor}_{u2} = \text{TRACE}(V_{11}G_{u12}) + y_1\hat{p}G_{u12}y_1\hat{h} - 8y_1\hat{p}G_{u12}x_1 + \text{TRACE}(V_{12}G_{u22}) + y_1\hat{p}G_{u22}y_2\hat{h} - 2y_1\hat{p}G_{u22}x_2 - 5\text{TRACE}(V_{11}G_{u22}) - 5y_1\hat{p}G_{u22}y_1\hat{h}; \]

\[ G_{u12} = v_{11}Z_{star}dD_{1d}S_{u12}Z_{starP}v_{11}; \]
\[ G_{u1211} = G_{u12}[1:8, 1:8]; G_{u1212} = G_{u12}[1:8, 9:16]; \]
\[ G_{u1221} = G_{u12}[9:16, 1:8]; G_{u1222} = G_{u12}[9:16, 9:16]; \]

\[ V_{11}G_{u12} = V_{11}G_{u12}; \]
\[ V_{12}G_{u12} = V_{12}G_{u22}; \]
\[ V_{11}G_{u22} = V_{11}G_{u21}; \]

\[ \text{Cor}_{u2} = \text{TRACE}(V_{11}G_{u12}) + y_1\hat{p}G_{u12}y_1\hat{h} - 8y_1\hat{p}G_{u12}x_1 + \text{TRACE}(V_{12}G_{u22}) + y_1\hat{p}G_{u22}y_2\hat{h} - 2y_1\hat{p}G_{u22}x_2 - 5\text{TRACE}(V_{11}G_{u22}) - 5y_1\hat{p}G_{u22}y_1\hat{h}; \]

\[ \text{corrb} = \text{Cor}_{u1}||\text{Cor}_{u2}||\text{Cor}_{u12}||\text{Cor}_{1}||\text{Cor}_{2}||\text{Cor}_{12}; \]
\[ \text{corrfixed} = \text{corr} = \text{corr} = \text{corr} \text{fixed} // \text{corrb}; \]

\[ Q_{11} = (v_{11}Z_{star}dD_{1d}S_{u1}Z_{starP})*(v_{11}Z_{star}dD_{1d}S_{u1}Z_{starP}); N_{11} = 0.5*\text{TRACE}(Q_{11}); \]
\[ Q_{12} = (v_{11}Z_{star}dD_{1d}S_{u1}Z_{starP})*(v_{11}Z_{star}dD_{1d}S_{u2}Z_{starP}); N_{12} = 0.5*\text{TRACE}(Q_{12}); \]
\[ Q_{13} = (v_{11}Z_{star}dD_{1d}S_{u1}Z_{starP})*(v_{11}Z_{star}dD_{1d}S_{u12}Z_{starP}); N_{13} = 0.5*\text{TRACE}(Q_{13}); \]
\[ Q_{14} = (v_{11}Z_{star}dD_{1d}S_{u1}Z_{starP})*(v_{11}dR_{1d}S_{1}); N_{14} = 0.5*\text{TRACE}(Q_{14}); \]
\[ Q_{15} = (v_{11}Z_{star}dD_{1d}S_{u1}Z_{starP})*(v_{11}dR_{1d}S_{2}); N_{15} = 0.5*\text{TRACE}(Q_{15}); \]
\[ Q_{16} = (v_{11}Z_{star}dD_{1d}S_{u1}Z_{starP})*(v_{11}dR_{1d}S_{1}); N_{16} = 0.5*\text{TRACE}(Q_{16}); \]
\[ Q_{22} = (v_{11}Z_{star}dD_{1d}S_{u2}Z_{starP})*(v_{11}Z_{star}dD_{1d}S_{u2}Z_{starP}); N_{22} = 0.5*\text{TRACE}(Q_{22}); \]
\[ Q_{23} = (v_{11}Z_{star}dD_{1d}S_{u2}Z_{starP})*(v_{11}Z_{star}dD_{1d}S_{u2}Z_{starP}); N_{23} = 0.5*\text{TRACE}(Q_{23}); \]
\[ Q_{24} = (v_{11}Z_{star}dD_{1d}S_{u2}Z_{starP})*(v_{11}dR_{1d}S_{1}); N_{24} = 0.5*\text{TRACE}(Q_{24}); \]
\[ Q_{25} = (v_{11}Z_{star}dD_{1d}S_{u2}Z_{starP})*(v_{11}dR_{1d}S_{2}); N_{25} = 0.5*\text{TRACE}(Q_{25}); \]
\[ Q_{26} = (v_{11}Z_{star}dD_{1d}S_{u2}Z_{starP})*(v_{11}dR_{1d}S_{12}); N_{26} = 0.5*\text{TRACE}(Q_{26}); \]
\[ Q_{33} = (v_{11}Z_{star}dD_{1d}S_{u12}Z_{starP})*(v_{11}Z_{star}dD_{1d}S_{u12}Z_{starP}); N_{33} = 0.5*\text{TRACE}(Q_{33}); \]
Q34 = (\text{vin}^* \text{Zstar} \cdot d\text{D1} \cdot \text{su12} \cdot \text{ZstarP}) \cdot (\text{vin} \cdot d\text{R1dS1}) ; \text{N34} = .5 \cdot \text{TRACE} (Q34);
Q35 = (\text{vin}^* \text{Zstar} \cdot d\text{D1} \cdot \text{su12} \cdot \text{ZstarP}) \cdot (\text{vin} \cdot d\text{R1dS2}) ; \text{N35} = .5 \cdot \text{TRACE} (Q35);
Q36 = (\text{vin}^* \text{Zstar} \cdot d\text{D1} \cdot \text{su12} \cdot \text{ZstarP}) \cdot (\text{vin} \cdot d\text{R1dS12}) ; \text{N36} = .5 \cdot \text{TRACE} (Q36);
Q44 = (\text{vin} \cdot d\text{R1dS1}) \cdot (\text{vin} \cdot d\text{R1dS1}) ; \text{N44} = .5 \cdot \text{TRACE} (Q44);
Q45 = (\text{vin} \cdot d\text{R1dS1}) \cdot (\text{vin} \cdot d\text{R1dS2}) ; \text{N45} = .5 \cdot \text{TRACE} (Q45);
Q46 = (\text{vin} \cdot d\text{R1dS1}) \cdot (\text{vin} \cdot d\text{R1dS12}) ; \text{N46} = .5 \cdot \text{TRACE} (Q46);
Q55 = (\text{vin} \cdot d\text{R1dS2}) \cdot (\text{vin} \cdot d\text{R1dS2}) ; \text{N55} = .5 \cdot \text{TRACE} (Q55);
Q56 = (\text{vin} \cdot d\text{R1dS2}) \cdot (\text{vin} \cdot d\text{R1dS12}) ; \text{N56} = .5 \cdot \text{TRACE} (Q56);
Q66 = (\text{vin} \cdot d\text{R1dS12}) \cdot (\text{vin} \cdot d\text{R1dS12}) ; \text{N66} = .5 \cdot \text{TRACE} (Q66);

\text{NRAN} = (\text{N11} || \text{N12} || \text{N13} || \text{N14} || \text{N15} || \text{N16})
   \quad /\ /(\text{N12} || \text{N22} || \text{N23} || \text{N24} || \text{N25} || \text{N26})
   \quad /\ /(\text{N13} || \text{N23} || \text{N33} || \text{N34} || \text{N35} || \text{N36})
   \quad /\ /(\text{N14} || \text{N24} || \text{N34} || \text{N44} || \text{N45} || \text{N46})
   \quad /\ /(\text{N15} || \text{N25} || \text{N35} || \text{N45} || \text{N55} || \text{N56})
   \quad /\ /(\text{N16} || \text{N26} || \text{N36} || \text{N46} || \text{N56} || \text{N66})

z_k = \text{NRAN};
H_2 = \text{DET}(z_k);

H_1 = (\text{InfMatFixed} || \text{corr}) /\ (\text{corrp} || \text{NRAN});
H_2 = \text{DET}(H_1);
\text{InfMat} = H_2;

\text{if} \ \text{InfMat} > D \ \text{then} \ DM = x;
\text{if} \ \text{InfMat} > D \ \text{then} \ D = \text{InfMat};
\text{end};
x = DM;
diff = D - D_1;
\text{if} \ \text{diff} = 0 \ \text{then} \ i = 300;
D_1 = D;
\text{end};
\text{if} \ D_1 > D_2 \ \text{then} \ x_{\text{final1}} = x;
\text{if} \ D_1 > D_2 \ \text{then} \ D_2 = D_1;
\text{end};
x = x_{\text{final1}};
xpx = x^x;
\text{print} \ xpx;
\text{BK1} = x[[1:2,1:2]]; \text{BK2} = x[[3:4,1:2]]; \text{BK3} = x[[5:6,1:2]]; \text{BK4} = x[[7:8,1:2]];
\text{print} \ \text{BK1} \ \text{BK2} \ \text{BK3} \ \text{BK4} \ \text{D2};
\text{quit};
\text{run};
A REML Optimal Design for Four Blocks with Block Size Equal to Two Assuming:

**proc IML ;**
**RESET;**
CanP={-1 -1,-1 0,-1 1,1 -1,1 0,1 1,0 -1,0 0,1 1};

s=0;D1=-1;D2=-1;D3=-1;D4=-1;D9=0;Zaher=0;

Alpha=.5;

D2=-1;
B=(1 -5, 0 1);
BPRIME=B`;
BprINV=Inv(BPRIME);
Bstar=I(8)@BprINV;
BstarP=Bstar`;

do l =1 to 10;
SEED=-1;
c=j(8, 2, seed);
x=3*uniform(c);
x=ceil(x);
x=x-2*j(8, 2);
D=-10; D1=0;
diff=10;
z=I(2)@J(2,1,1);
Zstar=z@I(4);
ZstarP=Zstar`;
do i =1 to 300;
sx=x;
do k =1 to 8;
idx = setdif(1:nrow(sx),k);
u = sx[idx, ];
do n =1 to 9;
   x=insert(u,CanP[n,],k,0);

yhat=8*x[ ,1];
Xstar=x@I(2);
XstarP=Xstar`;

SigmaU={.25 0.1,0.1 .16};
RU={1 0.5,0.5 1};
G=I(4)@SigmaU;
R=I(8)@RU;
v=Bstar*(Zstar*G*Zstar`+R)*BstarP;
vinv=Inv(v);
v11=vinv[1:8,1:8];v12=vinv[1:8,9:16];
v21=vinv[9:16,1:8];v22=vinv[9:16,9:16];
v1=v[1:8,1:8];

M=XstarP*vinv*Xstar;
**if** det(M)=0 **then**
   Minv=J(2,2,0);
**else**
   Minv=Inv(M);
   xl=x[ ,1];
x2=x[ , 2];

*Construct M the Information Matrix for the endogenous and exogenous parameters;

$v11v22=v11\cdot v12;
M11=x1\cdot v11\cdot x1;M12=x1\cdot v11\cdot x2;M13=x1\cdot v12\cdot y1hat;
M21=x2\cdot v11\cdot x1;M22=x2\cdot v11\cdot x2;M23=x2\cdot v12\cdot y1hat;
M31=y1hat\cdot v21\cdot x1;M32=y1hat\cdot v22\cdot x2;M33=\text{TRACE}(v11v22)+y1hat\cdot v12\cdot y1hat;

\text{InfMatFixed}=(M11||M12||M13)/(M21||M22||M23)/(M31||M32||M33);

dG1dSu1={1 0 , 0 0};
dG1dSu2={0 0 , 0 1};
dG1dSu12={0 1 , 1 0};

dD1dSu1=I(4)@dG1dSu1;
dD1dSu2=I(4)@dG1dSu2;
dD1dSu12=I(4)@dG1dSu12;
dR1dS1=I(8)@dG1dSu1;
dR1dS2=I(8)@dG1dSu2;
dR1dS12=I(8)@dG1dSu12;

P=vinv-\text{vinv}\cdot Xstar\cdot \text{Minv}\cdot XstarP\cdot \text{vinv};

*Construct the Correlation between the Exogenous parameter(s) and the random parameters;

dB1db1={0 -1 , 0 0};
BdB1db1=I(8)@dB1db1;
dv1db=-Bstar\cdot Bd1db1\cdot v-v\cdot Bd1db1\cdot Bstar;

*Cor1;
O11=(P\cdot Bstar\cdot Z1star\cdot dD1dSu1\cdot Z1starP\cdot BstarP\cdot P\cdot dv1db);Cor1=0.5*\text{TRACE}(O11);

*Cor2;
O22=(P\cdot Bstar\cdot Z1star\cdot dD1dSu2\cdot Z1starP\cdot BstarP\cdot P\cdot dv1db);Cor2=0.5*\text{TRACE}(O22);

*Cor12;
O33=(P\cdot Bstar\cdot Z1star\cdot dD1dSu12\cdot Z1starP\cdot BstarP\cdot P\cdot dv1db);Cor12=0.5*\text{TRACE}(O33);

*Cor1;
O44=(P\cdot Bstar\cdot dR1dS1\cdot BstarP\cdot P\cdot dv1db);Cor1=0.5*\text{TRACE}(O44);

*Cor2;
O55=(P\cdot Bstar\cdot dR1dS2\cdot BstarP\cdot P\cdot dv1db);Cor2=0.5*\text{TRACE}(O55);

*Cor12;
O66=(P\cdot Bstar\cdot dR1dS12\cdot BstarP\cdot P\cdot dv1db);Cor12=0.5*\text{TRACE}(O66);

corrb=Cor1||Coru2||Coru12||Cor1||Cor2||Cor12;
corrfixed={0 0 0 0 0 0,0 0 0 0 0 0};
corr=corrfixed//corrb;
corrp=corr;

*Construct N the Information Matrix for the covariance Parameters;

Q11=(P\cdot Bstar\cdot Z1star\cdot dD1dSu1\cdot Z1starP\cdot BstarP)\cdot (P\cdot Bstar\cdot Z1star\cdot dD1dSu1\cdot Z1starP\cdot BstarP);N11=0.5*\text{TRACE}(Q11);

Q12=(P\cdot Bstar\cdot Z1star\cdot dD1dSu2\cdot Z1starP\cdot BstarP)\cdot (P\cdot Bstar\cdot Z1star\cdot dD1dSu2\cdot Z1starP\cdot BstarP);N12=0.5*\text{TRACE}(Q12);
Q13 = (P*Bstar*Zstar*dD1dSu1*ZstarP*BstarP)*(P*Bstar*Zstar*dD1dSu12*ZstarP*BstarP); N13 = 0.5*TRACE(Q13);
Q14 = P*Bstar*Zstar*dD1dSu1*ZstarP*BstarP*P*Bstar*dR1dS1*BstarP; N14 = 0.5*TRACE(Q14);
Q15 = P*Bstar*Zstar*dD1dSu1*ZstarP*BstarP*P*Bstar*dR1dS2*BstarP; N15 = 0.5*TRACE(Q15);
Q16 = P*Bstar*Zstar*dD1dSu1*ZstarP*BstarP*P*Bstar*dR1dS12*BstarP; N16 = 0.5*TRACE(Q16);
Q22 = (P*Bstar*Zstar*dD1dSu2*ZstarP*BstarP)*(P*Bstar*Zstar*dD1dSu2*ZstarP*BstarP); N22 = 0.5*TRACE(Q22);
Q23 = (P*Bstar*Zstar*dD1dSu2*ZstarP*BstarP)*(P*Bstar*Zstar*dD1dSu12*ZstarP*BstarP); N23 = 0.5*TRACE(Q23);
Q24 = P*Bstar*Zstar*dD1dSu2*ZstarP*BstarP*P*Bstar*dR1dS1*BstarP; N24 = 0.5*TRACE(Q24);
Q25 = P*Bstar*Zstar*dD1dSu2*ZstarP*BstarP*P*Bstar*dR1dS2*BstarP; N25 = 0.5*TRACE(Q25);
Q26 = P*Bstar*Zstar*dD1dSu2*ZstarP*BstarP*P*Bstar*dR1dS12*BstarP; N26 = 0.5*TRACE(Q26);
Q33 = (P*Bstar*Zstar*dD1dSu12*ZstarP*BstarP)*(P*Bstar*Zstar*dD1dSu12*ZstarP*BstarP); N33 = 0.5*TRACE(Q33);
Q34 = P*Bstar*Zstar*dD1dSu12*ZstarP*BstarP*P*Bstar*dR1dS1*BstarP; N34 = 0.5*TRACE(Q34);
Q35 = P*Bstar*Zstar*dD1dSu12*ZstarP*BstarP*P*Bstar*dR1dS2*BstarP; N35 = 0.5*TRACE(Q35);
Q36 = P*Bstar*Zstar*dD1dSu12*ZstarP*BstarP*P*Bstar*dR1dS12*BstarP; N36 = 0.5*TRACE(Q36);
Q44 = (P*Bstar*dR1dS1*BstarP)*(P*Bstar*dR1dS1*BstarP); N44 = 0.5*TRACE(Q44);
Q45 = (P*Bstar*dR1dS1*BstarP)*(P*Bstar*dR1dS2*BstarP); N45 = 0.5*TRACE(Q45);
Q46 = (P*Bstar*dR1dS1*BstarP)*(P*Bstar*dR1dS12*BstarP); N46 = 0.5*TRACE(Q46);
Q55 = (P*Bstar*dR1dS2*BstarP)*(P*Bstar*dR1dS2*BstarP); N55 = 0.5*TRACE(Q55);
Q56 = (P*Bstar*dR1dS2*BstarP)*(P*Bstar*dR1dS12*BstarP); N56 = 0.5*TRACE(Q56);
Q66 = (P*Bstar*dR1dS12*BstarP)*(P*Bstar*dR1dS12*BstarP); N66 = 0.5*TRACE(Q66);
NRAN = (N11 || N12 || N13 || N14 || N15 || N16) //
(N12 || N22 || N23 || N24 || N25 || N26) //
(N13 || N23 || N33 || N34 || N35 || N36) //
(N14 || N24 || N34 || N44 || N45 || N46) //
(N15 || N25 || N35 || N45 || N55 || N56) //
(N16 || N26 || N36 || N46 || N56 || N66);
zk = NRAN;
H2 = DET(zk);

H1 = (InfMatFixed || corr) // (corrp || NRAN);
H2 = DET(H1);
InfMat = H2;

if InfMat > D then DM = x;
if InfMat > D then D = InfMat;
end;
x = DM;
diff=D-D1;
end;
if diff=0 then i=300;
D1=D;
end;
if D1 > D2 then xfinal1 = x;
if D1 > D2 then D2 = D1;
end;
x=xfinal1;
xpx=x`*x;
print xpx;
BK1 =x[1:2,1:2 ];BK2 =x[3:4,1:2 ];BK3 =x[5:6,1:2 ];BK4 =x[7:8,1:2 ];
print BK1 BK2 BK3 BK4 D2;
quit;
run;
SAS CODES FOR CHAPTER 5

Optimal Hybrid Genetic Combinations for Three Replicates Using 3SLS Methodology

SLURM Code:

#!/bin/sh
#SBATCH --time=72:00:00
#SBATCH --ntasks-per-node=12
#SBATCH --nodes=1
#SBATCH --mem-per-cpu=4096
#SBATCH --job-name=OptimalDesign11OCT18
#SBATCH --mail-type=END
#SBATCH --mail-user=zlkmail2003@yahoo.com
#SBATCH --error=OptimalDesign11OCT18.stderr
#SBATCH --output=OptimalDesign11OCT18.stdout
module load sas/9.4
sas CMD BATCH MLE.sas

proc IML;  \*RESET;
c=(-1 1);
CanP={0 0 0 0 0};
do j =1 to 2;
do k =1 to 2;
do l =1 to 2;
do m =1 to 2;
do n =1 to 2;
cl=c[1, j]|c[1, k]|c[1, l]|c[1, m]|c[1, n];
CanP=CanP//cl;
end;
end;
end;
end;
CanP=CanP[2:33,1:5];
s=0;D1=-1;D2=-1;D3=-1;D4=-1;D=-1;DT=0;Zaher=0;
D2=-1;
do l =1 to 100;
prob = {0.5 0.5 0.5 0.5 0.5};
p = repeat(prob, 96);
call streaminit(-1);
x = 2*rand("Bernoulli", p) - J(96, 5, 1);
D=-10; D1=0;
diff=10;
ZEnv=I(3) @ J(32, 1, 1);
ZBk=I(12) @ J(8, 1, 1);

do i = 1 to 3000;
sx=x;
do k = 1 to 96;
idx = setdif(1:nrow(sx), k);
u = sx[idx, ];
do n = 1 to 32;
  x = insert(u, CanP[n,], k, 0);
T1=6*ZEnv[,1]+4.2*ZEnv[,2]+4.5*ZEnv[,3];
T2=14.3*ZEnv[,1]+14.3*ZEnv[,2]+13.2*ZEnv[,3];
T3=21.6*ZEnv[,1]+21.2*ZEnv[,2]+20.2*ZEnv[,3];
P0=109.2*ZEnv[,1]+159.5*ZEnv[,2]+132.1*ZEnv[,3];
P1=92.5*ZEnv[,1]+101.1*ZEnv[,2]+90.2*ZEnv[,3];
P2=55.9*ZEnv[,1]+84.1*ZEnv[,2]+116.1*ZEnv[,3];
SR2=774.7*ZEnv[,1]+734.7*ZEnv[,2]+1035.1*ZEnv[,3];
w1=ZBk||x;
w1Star=I(4) @ w1;
w1StarP=w1Star`;
x1=x[1]; T2;
x2=x[2]; P2;
x3=x[3]; T1;
x4=x[3]; P0;
x5=x[2]; P1;
x6=x[4]; T1;
x7=x[5]; T2;
x8=x[3]; SR2;
y1hat=0.11*x1-0.10*x2+0.12*x3;
y2hat=-0.07*x4-0.08*x5+0.08*x6-0.62*y1hat;
y3hat=0.11*x7+0.09*x8-0.35*y2hat-0.63*y1hat;
y4hat=0.07*x1+0.07*x4+0.05*x5+1.25*y1hat+0.76*y2hat+0.47*y3hat;
w1=x1||x2||x3||J(96, 14, 0);
w2=J(96, 3, 0)||x4||x5||x6||y1hat||J(96, 10, 0);
w3=J(96, 7, 0)||x7||x8||y1hat||y2hat||J(96, 6, 0);
w4=J(96, 11, 0)||x1||x4||x5||y1hat||y2hat||y3hat;
X2Star=w1//w2//w3//w4;
X2StarP=X2Star`;
Sigma=(1 0 0 0, 0 1 0 0, 0 0 1 0, 0 0 0 1);
SigmaR=Sigma@I(96);
V=SigmaR;
VI=Inv(V);
M=w11StarP*VI*w11Star;
Minv= ginv(M);
P=VI-VI*w11Star*Minv*w11StarP*VI;
INF=X2StarP*P*X2Star;
H1=DET(INF);
InfMat=LOG(H1);
    if InfMat > D then DM = x;
    if InfMat > D then D = InfMat;
end;
x=DM;
diff=D-D1;
end;
if diff=0 then i=3000;
D1=D;
end;
if D1 > D2 then xfinal1 = x;
if D1 > D2 then D2 = D1;
end;

x=xfinal1;
xx=x"*x;
print xx;
BK1 =x[1:8,1:5];BK2 =x[9:16,1:5];BK3 =x[17:24,1:5];BK4 =x[25:32,1:5];
BK5 =x[33:40,1:5];BK6 =x[41:48,1:5];BK7 =x[49:56,1:5];BK8
=x[57:64,1:5];
BK9 =x[65:72,1:5];BK10 =x[73:80,1:5];BK11 =x[81:88,1:5];BK12
=x[89:96,1:5];
print BK1 BK2 BK3 BK4 BK5 BK6 BK7 BK8 BK9 BK10 BK11 BK12 D2;
quit;
run;
Optimal Hybrid Genetic Combinations for Three Replicates Using FIML Methodology

SLURM Code:

```bash
#!/bin/sh
#SBATCH --time=72:00:00
#SBATCH --ntasks-per-node=12
#SBATCH --nodes=1
#SBATCH --mem-per-cpu=4096
#SBATCH --job-name=OptimalDesign24OCT18
#SBATCH --mail-type=END
#SBATCH --mail-user=zlkmail2003@yahoo.com
#SBATCH --error=OptimalDesign24OCT18.stderr
#SBATCH --output=OptimalDesign24OCT18.stdout
module load sas/9.4
sas CMD BATCH MLEWNotFixed.sas
```

```sas
proc IML ;RESET;
c={-1 1};
CanP={0 0 0 0};
do j =1 to 2;
do k =1 to 2;
do l =1 to 2;
do m =1 to 2;
do n =1 to 2;
c1=c[1, j]||c[1, k]||c[1, l]||c[1, m]||c[1, n];
CanP=CanP//c1;
end;
end;
end;
end;
CanP=CanP[2:33,1:5];

s=0;D1=-1;D2=-1;D3=-1;D4=-1;D=-1;DT=0;Zaher=0;
D2=-1;

do l =1 to 50;
prob = {0.5 0.5 0.5 0.5 0.5};
p = repeat(prob, 96);
call streaminit(-1);
x = 2*rand("Bernoulli", p)-J(96,5,1);

D=-10; D1=0;
diff=10;
```
ZEnv = I(3) @ J(32, 1, 1);
ZBk = I(12) @ J(8, 1, 1);

do i = 1 to 3000;
sx = x;
do k = 1 to 96;
idx = setdif(1:nrow(sx), k);
u = sx[idx, ];
do n = 1 to 32;
x = insert(u, CanP[n, ], k, 0);
T1 = 6 * ZEnv[, 1] + 4.2 * ZEnv[, 2] + 4.5 * ZEnv[, 3];
T2 = 14.3 * ZEnv[, 1] + 14.3 * ZEnv[, 2] + 13.2 * ZEnv[, 3];
T3 = 21.6 * ZEnv[, 1] + 21.2 * ZEnv[, 2] + 20.2 * ZEnv[, 3];
P0 = 109.2 * ZEnv[, 1] + 159.5 * ZEnv[, 2] + 132.1 * ZEnv[, 3];
P1 = 92.5 * ZEnv[, 1] + 101.1 * ZEnv[, 2] + 90.2 * ZEnv[, 3];
P2 = 55.9 * ZEnv[, 1] + 84.1 * ZEnv[, 2] + 116.1 * ZEnv[, 3];
SR2 = 774.7 * ZEnv[, 1] + 734.7 * ZEnv[, 2] + 1035.1 * ZEnv[, 3];
w11 = ZBk || x;
w11Star = I(4) @ w11;
w11StarP = w11Star;
x1 = x[, 1] # T2;
x2 = x[, 2] # P2;
x3 = x[, 3] # T1;
x4 = x[, 2] # P0;
x5 = x[, 2] # P1;
x6 = x[, 4] # T1;
x7 = x[, 5] # T2;
x8 = x[, 3] # SR2;
y1hat = 0.11 * x1 - 0.10 * x2 + 0.12 * x3;
y2hat = -0.07 * x4 - 0.08 * x5 + 0.08 * x6 - 0.62 * y1hat;
y3hat = 0.11 * x7 + 0.09 * x8 - 0.35 * y2hat - 0.63 * y1hat;
y4hat = 0.07 * x1 + 0.07 * x4 + 0.05 * x5 + 1.25 * y1hat + 0.71 * y2hat + 0.47 * y3hat;

w1 = x1 || x2 || x3 || J(96, 14, 0);
w2 = J(96, 3, 0) || x4 || x5 || x6 || y1hat || J(96, 10, 0);
w3 = J(96, 7, 0) || x7 || x8 || y1hat || y2hat || J(96, 6, 0);
w4 = J(96, 11, 0) || x1 || x4 || x5 || y1hat || y2hat || y3hat;

X2Star = w1 // w2 // w3 // w4;
X2StarP = X2Star;

Sigma = {1 0 0 0 0 0 1 0 0, 0 0 1 0 0 0 0 1};
SigmaR = Sigma @ I(96);
V = SigmaR;
VI = Inv(V);
M = w11StarP * VI * w11Star;
Minv = ginv(M);
P = VI - VI * w11Star * Minv * w11StarP * VI;


\(M_{11} = x_1^*P_{11}x_1; M_{12} = x_1^*P_{12}x_2; M_{13} = x_1^*P_{13}x_3; \)
\(M_{14} = x_1^*P_{14}x_4; \)
\(M_{11} = (M_{11_{11}} | M_{11_{12}} | M_{11_{13}}) // (M_{11_{21}} | M_{11_{22}} | M_{11_{23}}) // (M_{11_{31}} | M_{11_{32}} | M_{11_{33}}); \)
\(M_{12} = (M_{12_{11}} | M_{12_{12}} | M_{12_{13}} | M_{12_{14}}) // (M_{12_{21}} | M_{12_{22}} | M_{12_{23}} | M_{12_{24}}) // (M_{12_{31}} | M_{12_{32}} | M_{12_{33}} | M_{12_{34}}); \)
\(M_{13} = (M_{13_{11}} | M_{13_{12}} | M_{13_{13}} | M_{13_{14}}) // (M_{13_{21}} | M_{13_{22}} | M_{13_{23}} | M_{13_{24}}) // (M_{13_{31}} | M_{13_{32}} | M_{13_{33}} | M_{13_{34}}); \)
\(M_{14} = (M_{14_{11}} | M_{14_{12}} | M_{14_{13}} | M_{14_{14}} | M_{14_{15}} | M_{14_{16}} // (M_{14_{21}} | M_{14_{22}} | M_{14_{23}} | M_{14_{24}}) // (M_{14_{31}} | M_{14_{32}} | M_{14_{33}} | M_{14_{34}}) // (M_{14_{35}} | M_{14_{36}} | M_{14_{37}} | M_{14_{38}}); \)
\(M_{21} = M_{12}; \)

\(v_{11P22} = v_{11}P_{22}; \)
\(v_{21P22} = v_{12}P_{22}; \)

M22_{11} = x_4^*P_{22}x_5; M22_{12} = x_4^*P_{22}x_5; M22_{13} = x_4^*P_{22}x_6; M22_{14} = x_4^*P_{22}y_1; 
M22_{21} = x_5^*P_{22}x_5; M22_{22} = x_5^*P_{22}x_5; M22_{23} = x_5^*P_{22}x_6; M22_{24} = x_5^*P_{22}y_1; \)
M22_31=x6`*P22*x4; M22_32=x6`*P22*x5; M22_33=x6`*P22*x6; M22_34=x6`*P22*y1\hat;
M22_41=y1\hat`*P22*x4; M22_42=y1\hat`*P22*x5; M22_43=y1\hat`*P22*x6; M22_44=TRA\CE(v11P22)+y1\hat`*P22*y1\hat;
M22=(M22_11||M22_12||M22_13||M22_14)//(M22_21||M22_22||M22_23||M22_24)
//(M22_31||M22_32||M22_33||M22_34)//(M22_41||M22_42||M22_43||M22_44);
v11P23=V11*P23;
v12P23=V12*P23;
M23_11=x4`*P23*X7; M23_12=x4`*P23*X8; M23_13=x4`*P23*y1\hat; M23_14=x4`*P23*y2\hat;
M23_21=x5`*P23*X7; M23_22=x5`*P23*X8; M23_23=x5`*P23*y1\hat; M23_24=x5`*P23*y2\hat;
M23_31=x6`*P23*X7; M23_32=x6`*P23*X8; M23_33=x6`*P23*y1\hat; M23_34=x6`*P23*y2\hat;
M23_41=y1\hat`*P23*X7; M23_42=y1\hat`*P23*X8; M23_43=TRA\CE(v11P23)+y1\hat`*P23*y1\hat;
M23_44=TRA\CE(v12P23)+y1\hat`*P23*y2\hat;
M23=(M23_11||M23_12||M23_13||M23_14)//(M23_21||M23_22||M23_23||M23_24)//(M23_31||M23_32||M23_33||M23_34)//(M23_41||M23_42||M23_43||M23_44);
v11P24=V11*P24;
v12P24=V12*P24;
v13P24=V13*P24;
M24_11=x4`*P24*X1; M24_12=x4`*P24*X4; M24_13=x4`*P24*x5; M24_14=x4`*P24*y1\hat;
M24_15=x4`*P24*y2\hat; M24_16=x4`*P24*y3\hat;
M24_21=x5`*P24*X1; M24_22=x5`*P24*X4; M24_23=x5`*P24*x5; M24_24=x5`*P24*y1\hat;
M24_25=x5`*P24*y2\hat; M24_26=x5`*P24*y3\hat;
M24_31=x6`*P24*X1; M24_32=x6`*P24*X4; M24_33=x6`*P24*x5; M24_34=x6`*P24*y1\hat;
M24_35=x6`*P24*y2\hat; M24_36=x6`*P24*y3\hat;
M24_41=y1\hat`*P24*X1; M24_42=y1\hat`*P24*x4; M24_43=y1\hat`*P24*x5; M24_44=TRA\CE(v11P24)+y1\hat`*P24*y1\hat;
M24_45=TRA\CE(v12P24)+y1\hat`*P24*y2\hat; M24_46=TRA\CE(v13P24)+y1\hat`*P24*y3\hat;
M31=M13`; M32=M23`;
M33_41=y2hat`*P33*x7;M33_42=y2hat`*P33*x8;M33_43=TRACE(v21P33)+y2hat`*P33*y1hat;M33_44=TRACE(v22P33)+y2hat`*P33*y2hat;

M33=(M33_11||M33_12||M33_13||M33_14)//(M33_21||M33_22||M33_23||M33_24)//(M33_31||M33_32||M33_33||M33_34)//(M33_41||M33_42||M33_43||M33_44);

v11P34=V11*P34;
v12P34=V12*P34;
v13P34=V13*P34;
v21P34=V21*P34;
v22P34=V22*P34;
v23P34=V23*P34;

M34_11=x7`*P34*x1;M34_12=x7`*P34*x4;M34_13=x7`*P34*x5;M34_14=x7`*P34*y1hat;M34_15=x7`*P34*y2hat;M34_16=x7`*P34*y3hat;
M34_21=x8`*P34*x1;M34_22=x8`*P34*x4;M34_23=x8`*P34*x5;M34_24=x8`*P34*y1hat;M34_25=x8`*P34*y2hat;M34_26=x8`*P34*y3hat;
M34_31=y1hat`*P34*x1;M34_32=y1hat`*P34*x4;M34_33=y1hat`*P34*x5;M34_34=TRACE(v11P34)+y1hat`*P34*y1hat;M34_35=TRACE(v12P34)+y1hat`*P34*y2hat;M34_36=TRACE(v13P34)+y1hat`*P34*y3hat;
M34_41=y2hat`*P34*x1;M34_42=y2hat`*P34*x4;M34_43=y2hat`*P34*x5;M34_44=TRACE(v21P34)+y2hat`*P34*y1hat;M34_45=TRACE(v22P34)+y2hat`*P34*y2hat;M34_46=TRACE(v23P34)+y2hat`*P34*y3hat;

M34=(M34_11||M34_12||M34_13||M34_14||M34_15||M34_16)//(M34_21||M34_22||M34_23||M34_24||M34_25||M34_26)//(M34_31||M34_32||M34_33||M34_34||M34_35||M34_36)//(M34_41||M34_42||M34_43||M34_44||M34_45||M34_46);

M41=M14`;M42=M24`;M43=M34`;
v11P44=V11*P44;
v12P44=V12*P44;
v13P44=V13*P44;
v21P44=V21*P44;
v22P44=V22*P44;
v23P44=V23*P44;
v31P44=V31*P44;
v32P44=V32*P44;
v33P44=V33*P44;

M44_11=x1`*P44*x1;M44_12=x1`*P44*x4;M44_13=x1`*P44*x5;M44_14=x1`*P44*y1hat;M44_15=x1`*P44*y2hat;M44_16=x1`*P44*y3hat;
M44_21=x4`*P44*x1;M44_22=x4`*P44*x4;M44_23=x4`*P44*x5;M44_24=x4`*P44*y1hat;M44_25=x4`*P44*y2hat;M44_26=x4`*P44*y3hat;
M44_31=x5`*P44*x1;M44_32=x5`*P44*x4;M44_33=x5`*P44*x5;M44_34=x5`*P44*y1hat;M44_35=x5`*P44*y2hat;M44_36=x5`*P44*y3hat;
M44_41=y1hat`*P44*x1;M44_42=y1hat`*P44*x4;M44_43=y1hat`*P44*x5;M44_44=TRACE(v11P44)+y1hat`*P44*y1hat;M44_45=TRACE(v12P44)+y1hat`*P44*y2hat;M44_46=TRACE(v13P44)+y1hat`*P44*y3hat;
M44_51=y2hat`*P44*x1;M44_52=y2hat`*P44*x4;M44_53=y2hat`*P44*x5;M44_54=TRACE(v21P44)+y2hat`*P44*y1hat;M44_55=TRACE(v22P44)+y2hat`*P44*y2hat;M44_56=TRACE(v23P44)+y2hat`*P44*y3hat;
M44_61=y3hat`*P44*x1;M44_62=y3hat`*P44*x4;M44_63=y3hat`*P44*x5;M44_64=TRACE(v31P44)+y3hat`*P44*y1hat;M44_65=TRACE(v32P44)+y3hat`*P44*y2hat;M44_66=TRACE(v33P44)+y3hat`*P44*y3hat;

M44=(M44_11||M44_12||M44_13||M44_14||M44_15||M44_16)//(M44_21||M44_22||M44_23||M44_24||M44_25||M44_26)//
INF=M;
H1=DET(INF);
InfMat=LOG(H1);

if InfMat > D then DM = x;
if InfMat > D then D = InfMat;
end;
x=DM;
diff=D-D1;
end;
if diff=0 then i=3000;
  D1=D;
end;
if D1 > D2 then xfinal1 = x;
if D1 > D2 then D2 = D1;
end;

x=xfinal1;
xpx=x`*x;

print xpx;
BK1 =x[1:8,1:5];BK2 =x[9:16,1:5];BK3 =x[17:24,1:5];BK4 =x[25:32,1:5];
BK5 =x[33:40,1:5];BK6 =x[41:48,1:5];BK7 =x[49:56,1:5];BK8 =x[57:64,1:5];
BK9 =x[65:72,1:5];BK10 =x[73:80,1:5];BK11 =x[81:88,1:5];BK12 =x[89:96,1:5];

print BK1 BK2 BK3 BK4 BK5 BK6 BK7 BK8 BK9 BK10 BK11 BK12 D2; quit;
run;
Optimal Number of Replicates Using 3SLS Methodology

```r
proc IML; reset;
CP1={24.9 9.2 9.2 15}; CP2={24.2 10.1 9 12}; CP3={23.9 10.9 8.2 59};
CP4={23.9 10.9 10.3 41}; CP5={25 11 8.4 0}; CP6={22.9 11.3 9.8 0};
CanP={1 24.9 9.2 9.2 15, 2 24.2 10.1 9 12, 3 23.9 10.9 8.2 59, 4 23.9 10.9 10.3 41, 5 25 11 8.4 0, 6 22.9 11.3 9.8 0};
s=0; D1=-1; D2=-1; D3=-1; D4=-1; D=-1; DT=0; Zaiher=0;
z={1 0 0 0 0 0};
Alpha=.5;
y1=y11//y12//y13//y14;
y2=y21//y22//y23//y24;
ZI=i(6);
y=y1//y2;
D2=-1;
max=6;
do l =1 to 1;
x=CanP[l, ];
call randseed(123);
u=j(42,1);
call randgen(u,"uniform");
x=a+(b-a)*u; to generate numbers between -1 and 1;
k=ceil(max*u); to generate numbers between 1 and 10;
do n=1 to 42;
m=k[n, ];
x=x//CanP[m, ];
end;
x=x[2:43, ];
D=-10; D1=0;
diff=10;
z=I(6)@J(7,1);
diff=10;
do i =1 to 300;
sx=x;
do k =1 to 42;
idx = setdif(1:nrow(sx),k);
u = sx[idx, ];
do n =1 to 6;
x=insert(u,CanP[n,],k,0);
end;
end;
do n=1 to 42;
m=x[n, ];
zc=ZI[m, ];
z=z//zc;
end;="/"
print z;
z=z[2:43, ];
xl=x[2:5];
xpz=xl^*z;
px=xl^*xl;
Zstar=I(4)@xpz; Zstarp=Zstar^; 
* Constructing the variance ++++++++++++++++++++++++++++;
* G11=.25*I(4); *G22=.4*I(4); **G12=0*I(4);
* G=(G11 || G12) // (G12 || G22);
```
G = .4*I(24); *for the simplicity assume that they have the same affect on the variables;
Sigma = I(4);
*v = Zstar * G * Zstar' + Sigma@xpx;
v = Sigma@xpx;

if det(v) = 0 then
  vinv = J(6, 6, 0);
else
  vinv = Inv(v);
y1hat = -0.144*x1[, 1] + 0.517*x1[, 3];
y2hat = -0.435*y1hat - 0.171*x1[, 1] - 0.259*x1[, 2] - 0.007*x1[, 4];
y3hat = -0.639*y1hat - 1.415*y2hat + 0.305*x1[, 1] - 0.525*x1[, 3] - 0.022*x1[, 4];
y4hat = -0.639*y1hat - 1.415*y2hat + 0.305*x[, 1] - 0.525*x[, 3];

y3hat = -0.639*y1hat - 1.415*y2hat + 0.305*y3hat;

w1 = x1[, 1] || x1[, 3];
w2 = x1[, 1] || x1[, 2] || x1[, 4] || y1hat;
w3 = x1[, 1] || x1[, 3] || x1[, 4] || y1hat || y2hat;
w4 = y1hat || y2hat || y3hat;

w = (w1 || J(42, 12, 0)) // (J(42, 2, 0) || w2 || J(42, 8, 0)) // (J(42, 6, 0) || w3 || J(42, 3, 0)) // (J(42, 11, 0) || w4);

Ixpw = (I(4) @ x1') * w;
Ixwpw = Ixpw';
InfMat = Ixwpw * vinv * Ixpw;
InfMat1 = InfMat[1:2, 1:2];
InfMat2 = InfMat[3:6, 3:6];
InfMat3 = InfMat[7:11, 7:11];
InfMat4 = InfMat[12:14, 12:14];

H1 = DET(InfMat1) + 1;
H2 = DET(InfMat2) + 1;
H3 = DET(InfMat3) + 1;
H4 = DET(InfMat4) + 1;
c = LOG(H1) + LOG(H2) + LOG(H3) + LOG(H4);
*print MF, H1;

*InfMat = H1;
*c = LOG(H1); *print i k n c;
  if c > D then DM = x;
  if c > D then D = c;
end;
x = DM;
diff = D - D1;
end;
if diff = 0 then i = 300;
D1 = D;
end;
if D1 > D2 then xfinal1 = x;
if D1 > D2 then D2 = D1;
end;
x=xfinal1;
*xpx=x`*x;
*print xpx;
/*x1=x[1, ];x2=x[2, ];x3=x[3, ];x4=x[4, ];x5=x[5, ];x6=x[6, ];x7=x[7, ];x8=x[8, ];
x9=x[9, ];x10=x[10, ];x11=x[11, ];x12=x[12, ];x13=x[13, ];x14=x[14, ];x15=x[15, ];x16=x[16, ];
BK1 = x1//x2//x3//x4;BK2 = x5//x6//x7//x8;BK3 = x9//x10//x11//x12;BK4 = x13//x14//x15//x16;*/
print InfMat;
print x D2 H1 H2 H3 H4;
quit;
run;
Optimal Number of Replicates Using FIML Methodology

```plaintext
proc IML; RESET;
CP1={24.9 9.2 9.2 15 59}; CP2={24.2 10.1 9 12}; CP3={23.9 10.9 8.2 10.3 41};
CP5={25 11 8.4 0}; CP6={22.9 11.3 9.8 0};
CanP={1 24.9 9.2 9.2 15, 24.2 10.1 9 12, 23.9 10.9 8.2 59, 23.9 10.9 10.3 41, 22.9 11.3 9.8 0};
s=0; D1=-1; D2=-1; D3=-1; D4=-1; DT=0; Zaher=0;
y13={-10.21369, -11.58492, 10.476823, -6.035327}; y14={6.3500968, 12.236637, 11.071422, -0.1879};
y21={34.716418, -48.473, 51.057367, -25.94062}; y22={60.810192, 20.172189, 16.761147, -54.70027};
y23={-50.40517, -62.07898, 47.412563, -34.01863}; y24={33.680282, 62.268266, 52.233924, -25.95692};
z=(1 0 0 0 0 0);
Alpha=.5;
y1=y11//y12//y13//y14;
y2=y21//y22//y23//y24;
ZI=i(6);
y=y1//y2;
D2=-1;
max=6;
do l =1 to 1;
x=CanP[1, ];
call randseed(123);
u=j(42,1);
call randgen(u,"uniform");
*x=a+(b-a)*u; /*to generate numbers between -1 and 1;
k=ceil(max*u); /*to generate numbers between 1 and 10;
do n=1 to 42;
m=k[n, ];
x=x//CanP[m, ];
end;
x=x[2:43, ];
D=-10; D1=0;
diff=10;
z=I(6)@J(7,1);
    diff=10;
do i =1 to 300;
sx=x;
do k =1 to 42;
idx = setdif(1:nrow(sx),k);
u = sx[idx, ];
    do n =1 to 6;
    x=insert(u,CanP[n,],k,0);
    /do n=1 to 42;
    m=x[n,1 ]; /*print m;
    zc=ZI[m, ]; /*print zc;
    z=z//zc; /*print z;
end; /*
*print z;
```
*\(z = z[2:43, ]\); *print \(z\); 
\(x_4 = \text{cal}[\text{ }2:5];\) 
\(x_{pz} = x_4\text{cal}^\ast z;\) 
\(x_{px} = x_4\text{cal}^\ast x_4\text{cal};\) 
\(Z_{star} = I(4)@x_{pz}; Z_{star'} = Z_{star}^\ast;\) 
*Constructing the variance ++++++++++++++++++++++++++++++++; 
\(G_{11} = 0.25*I(4); \) 
\(G_{22} = 0.4*I(4); \) 
\(G_{12} = 0*I(4); \) 
\(G = (G_{11} || G_{12})/(G_{12} || G_{22});\) 
\(G = 0.4*I(24);\) *for the simpliity assume that they have the same affect on the variables; 
\(\Sigma = I(4); \) 
*\(v = Z_{star} \ast G \ast Z_{star}^\ast + \Sigma @ x_{px};\) 
\(v = \Sigma @ x_{px};\) 
if \(\det(v) = 0\) then 
\(v_{inv} = J(6, 6, 0);\) 
else 
\(v_{inv} = \text{Inv}(v);\) 
x_1 = x_4\text{cal}[\text{ }, 1]; 
x_2 = x_4\text{cal}[\text{ }, 2]; 
x_3 = x_4\text{cal}[\text{ }, 3]; 
x_4 = x_4\text{cal}[\text{ }, 4]; 
\(y_{1hat} = -0.144\ast x_1 + 0.517\ast x_3;\) 
\(y_{2hat} = -0.435\ast y_{1hat} + 0.171\ast x_1 - 0.259\ast x_2 - 0.007\ast x_4;\) 
\(y_{3hat} = 0.639\ast y_{1hat} - 1.415\ast y_{2hat} + 0.305\ast x_1 - 0.525\ast x_3 - 0.022\ast x_4;\) 
\(y_{4hat} = 0.561\ast y_{1hat} + 0.988\ast y_{2hat} + 0.289\ast y_{3hat};\) 
\(w_1 = x[\text{ }1] || x[\text{ }3];\) 
\(w_2 = x[\text{ }1] || x[\text{ }2] || x[\text{ }4];\) 
\(w_3 = x[\text{ }1] || x[\text{ }3] || x[\text{ }4];\) 
\(w_4 = y_{1hat} || y_{2hat} || y_{3hat};\) 
\(\Sigma = 0.4*I(4);\) 
\(\Sigma_{R} = \Sigma@I(42);\) 
\(V = I_{\text{Pw11}} \ast \Sigma_{R} \ast I_{\text{Pw11}}^\ast + 0.06*I(384);\) 
\(V = \Sigma_{R};\) 
\(V_{I} = \text{Inv}(V);\) 
\(V_{11} = V_{I}[1:42, 1:42]; V_{12} = V_{I}[1:42, 43:84]; V_{13} = V_{I}[1:42, 85:126]; V_{14} = V_{I}[1:42, 127:168];\) 
\(V_{31} = V[85:126, 1:42]; V_{32} = V[85:126, 43:84]; V_{33} = V[85:126, 85:126]; V_{34} = V[85:126, 127:168];\) 
\[ M_{11} = \frac{x_1 \cdot V_{11} \cdot x_1}{x_1 \cdot V_{11} \cdot x_3}; M_{11} = (M_{11_11} \| M_{11_12}) / (M_{11_21} \| M_{11_22}); \]

\[ M_{12} = \frac{x_1 \cdot V_{12} \cdot x_1}{x_1 \cdot V_{12} \cdot x_2}; M_{12} = (M_{12_11} \| M_{12_12} \| M_{12_13} \| M_{12_14}) / (M_{12_21} \| M_{12_22} \| M_{12_23} \| M_{12_24}); \]

\[ M_{13} = \frac{x_1 \cdot V_{13} \cdot x_1}{x_1 \cdot V_{13} \cdot x_3}; M_{13} = (M_{13_11} \| M_{13_12} \| M_{13_13} \| M_{13_14} \| M_{13_15}) / (M_{13_21} \| M_{13_22} \| M_{13_23} \| M_{13_24} \| M_{13_25}); \]

\[ M_{14} = \frac{x_1 \cdot V_{14} \cdot y_1{\hat{h}}}{x_1 \cdot V_{14} \cdot y_2{\hat{h}}}; M_{14} = (M_{14_11} \| M_{14_12} \| M_{14_13} \| M_{14_14} \| M_{14_15}) / (M_{14_21} \| M_{14_22} \| M_{14_23} \| M_{14_24} \| M_{14_25}); \]

\[ v_{11} = V_{11} \cdot V_{22}; \]

\[ M_{21} = M_{12}; \]

\[ v_{11} = V_{11} \cdot V_{22}; \]

\[ v_{12} = V_{12} \cdot V_{22}; \]

\[ M_{22} = \frac{x_1 \cdot V_{22} \cdot x_1}{x_1 \cdot V_{22} \cdot x_2}; M_{22} = (M_{22_11} \| M_{22_12} \| M_{22_13} \| M_{22_14} \| M_{22_15}) / (M_{22_21} \| M_{22_22} \| M_{22_23} \| M_{22_24} \| M_{22_25}); \]

\[ v_{11} = V_{11} \cdot V_{23}; \]

\[ v_{12} = V_{12} \cdot V_{23}; \]

\[ M_{23} = \frac{x_1 \cdot V_{23} \cdot x_1}{x_1 \cdot V_{23} \cdot x_3}; M_{23} = (M_{23_11} \| M_{23_12} \| M_{23_13} \| M_{23_14} \| M_{23_15}) / (M_{23_21} \| M_{23_22} \| M_{23_23} \| M_{23_24} \| M_{23_25}); \]

\[ v_{11} = V_{11} \cdot V_{24}; \]
\[ v_{12}v_{24} = V_{12}V_{24} ; \]
\[ v_{13}v_{24} = V_{13}V_{24} ; \]
\[ M_{24 \_11} = x_1^*V_{12}^*y_{1 \hat{h} a t} ; M_{24 \_12} = x_1^*V_{12}^*y_{2 \hat{h} a t} ; M_{24 \_13} = x_1^*V_{24}^*y_{3 \hat{h} a t} ; \]
\[ M_{24 \_21} = x_2^*V_{12}^*y_{1 \hat{h} a t} ; M_{24 \_22} = x_2^*V_{12}^*y_{2 \hat{h} a t} ; M_{24 \_23} = x_2^*V_{24}^*y_{3 \hat{h} a t} ; \]
\[ M_{24 \_31} = x_4^*V_{12}^*y_{1 \hat{h} a t} ; M_{24 \_32} = x_4^*V_{12}^*y_{2 \hat{h} a t} ; M_{24 \_33} = x_4^*V_{24}^*y_{3 \hat{h} a t} ; \]
\[ M_{24 \_41} = \text{TRACE}(v_{11}v_{24}) + y_{1 \hat{h} a t}^*V_{12}^*y_{1 \hat{h} a t} ; M_{24 \_42} = \text{TRACE}(v_{12}v_{24}) + y_{1 \hat{h} a t}^*V_{12}^*y_{2 \hat{h} a t} ; M_{24 \_43} = \text{TRACE}(v_{13}v_{24}) + y_{1 \hat{h} a t}^*V_{12}^*y_{3 \hat{h} a t} ; \]
\[ M_{24} = (M_{24 \_11} || M_{24 \_12} || M_{24 \_13} ) / (M_{24 \_21} || M_{24 \_22} || M_{24 \_23} ) / (M_{24 \_31} || M_{24 \_32} || M_{24 \_33} ) / (M_{24 \_41} || M_{24 \_42} || M_{24 \_43} ) ; \]
\[ M_{31} = M_{13}^* ; \]
\[ M_{32} = M_{23}^* ; \]
\[ v_{11}v_{33} = V_{11}V_{33} ; \]
\[ v_{12}v_{33} = V_{12}V_{33} ; \]
\[ v_{13}v_{33} = V_{21}V_{33} ; \]
\[ v_{22}v_{33} = V_{22}V_{33} ; \]
\[ M_{33 \_11} = x_1^*V_{33}^*x_1 ; M_{33 \_12} = x_1^*V_{33}^*x_3 ; M_{33 \_13} = x_1^*V_{33}^*x_4 ; M_{33 \_14} = x_1^*V_{33}^*y_{1 \hat{h} a t} ; M_{33 \_15} = x_1^*V_{33}^*y_{2 \hat{h} a t} ; \]
\[ M_{33 \_21} = x_3^*V_{33}^*x_1 ; M_{33 \_22} = x_3^*V_{33}^*x_3 ; M_{33 \_23} = x_3^*V_{33}^*x_4 ; M_{33 \_24} = x_3^*V_{33}^*y_{1 \hat{h} a t} ; M_{33 \_25} = x_3^*V_{33}^*y_{2 \hat{h} a t} ; \]
\[ M_{33 \_31} = x_4^*V_{33}^*x_1 ; M_{33 \_32} = x_4^*V_{33}^*x_3 ; M_{33 \_33} = x_4^*V_{33}^*x_4 ; M_{33 \_34} = x_4^*V_{33}^*y_{1 \hat{h} a t} ; M_{33 \_35} = x_4^*V_{33}^*y_{2 \hat{h} a t} ; \]
\[ M_{33 \_41} = y_{1 \hat{h} a t}^*V_{33}^*x_1 ; M_{33 \_42} = y_{1 \hat{h} a t}^*V_{33}^*x_3 ; M_{33 \_43} = y_{1 \hat{h} a t}^*V_{33}^*x_4 ; M_{33 \_44} = \text{TRACE}(v_{11}v_{33}) + y_{1 \hat{h} a t}^*V_{33}^*y_{1 \hat{h} a t} ; M_{33 \_45} = \text{TRACE}(v_{12}v_{33}) + y_{1 \hat{h} a t}^*V_{33}^*y_{2 \hat{h} a t} ; \]
\[ M_{33 \_51} = y_{2 \hat{h} a t}^*V_{33}^*x_1 ; M_{33 \_52} = y_{2 \hat{h} a t}^*V_{33}^*x_3 ; M_{33 \_53} = y_{2 \hat{h} a t}^*V_{33}^*x_4 ; M_{33 \_54} = \text{TRACE}(v_{21}v_{33}) + y_{2 \hat{h} a t}^*V_{33}^*y_{1 \hat{h} a t} ; M_{33 \_55} = \text{TRACE}(v_{22}v_{33}) + y_{2 \hat{h} a t}^*V_{33}^*y_{2 \hat{h} a t} ; \]
\[ M_{33} = (M_{33 \_11} || M_{33 \_12} || M_{33 \_13} || M_{33 \_14} || M_{33 \_15} ) / (M_{33 \_21} || M_{33 \_22} || M_{33 \_23} || M_{33 \_24} || M_{33 \_25} ) / (M_{33 \_31} || M_{33 \_32} || M_{33 \_33} || M_{33 \_34} || M_{33 \_35} ) / (M_{33 \_41} || M_{33 \_42} || M_{33 \_43} || M_{33 \_44} || M_{33 \_45} ) / (M_{33 \_51} || M_{33 \_52} || M_{33 \_53} || M_{33 \_54} || M_{33 \_55} ) ; \]
\[ v_{11}v_{34} = V_{11}V_{34} ; \]
\[ v_{12}v_{34} = V_{12}V_{34} ; \]
\[ v_{13}v_{34} = V_{13}V_{34} ; \]
\[ v_{21}v_{34} = V_{21}V_{34} ; \]
\[ v_{22}v_{34} = V_{22}V_{34} ; \]
\[ v_{23}v_{34} = V_{23}V_{34} ; \]
\[ M_{34 \_11} = x_1^*V_{34}^*y_{1 \hat{h} a t} ; M_{34 \_12} = x_1^*V_{34}^*y_{2 \hat{h} a t} ; M_{34 \_13} = x_1^*V_{34}^*y_{3 \hat{h} a t} ; \]
\[ M_{34 \_21} = x_3^*V_{34}^*y_{1 \hat{h} a t} ; M_{34 \_22} = x_3^*V_{34}^*y_{2 \hat{h} a t} ; M_{34 \_23} = x_3^*V_{34}^*y_{3 \hat{h} a t} ; \]
\[ M_{34 \_31} = x_4^*V_{34}^*y_{1 \hat{h} a t} ; M_{34 \_32} = x_4^*V_{34}^*y_{2 \hat{h} a t} ; M_{34 \_33} = x_4^*V_{34}^*y_{3 \hat{h} a t} ; \]
\[ M_{34 \_41} = \text{TRACE}(v_{11}v_{34}) + y_{1 \hat{h} a t}^*V_{34}^*y_{1 \hat{h} a t} ; M_{34 \_42} = \text{TRACE}(v_{12}v_{34}) + y_{1 \hat{h} a t}^*V_{34}^*y_{2 \hat{h} a t} ; M_{34 \_43} = \text{TRACE}(v_{13}v_{34}) + y_{1 \hat{h} a t}^*V_{34}^*y_{3 \hat{h} a t} ; \]
\[ M_{34 \_51} = \text{TRACE}(v_{21}v_{34}) + y_{2 \hat{h} a t}^*V_{34}^*y_{1 \hat{h} a t} ; M_{34 \_52} = \text{TRACE}(v_{22}v_{34}) + y_{2 \hat{h} a t}^*V_{34}^*y_{2 \hat{h} a t} ; M_{34 \_53} = \text{TRACE}(v_{23}v_{34}) + y_{2 \hat{h} a t}^*V_{34}^*y_{3 \hat{h} a t} ; \]
\[
M_{34} = (M_{34\_11} \| M_{34\_12} \| M_{34\_13}) / (M_{34\_21} \| M_{34\_22} \| M_{34\_23} \| M_{34\_24}) / (M_{34\_31} \| M_{34\_32} \| M_{34\_33}) / (M_{34\_41} \| M_{34\_42} \| M_{34\_43} \| M_{34\_44})
\]

\[
M_{41} = M_{14}^\prime; M_{42} = M_{24}^\prime; M_{43} = M_{34}^\prime;
\]

\[
v_{11} = v_{11}^* vi_{44}; v_{12} = v_{12}^* vi_{44}; v_{13} = v_{13}^* vi_{44};
\]

\[
v_{21} = v_{21}^* vi_{44}; v_{22} = v_{22}^* vi_{44}; v_{23} = v_{23}^* vi_{44};
\]

\[
v_{31} = v_{31}^* vi_{44}; v_{32} = v_{32}^* vi_{44}; v_{33} = v_{33}^* vi_{44};
\]

\[
M_{44\_11} = \text{TRACE}(v_{11} vi_{44}) + \gamma_{1}^\prime * v_{11}^* vi_{44}; M_{44\_12} = \text{TRACE}(v_{12} vi_{44}) + \gamma_{1}^\prime * v_{12}^* vi_{44};
\]

\[
M_{44\_21} = \text{TRACE}(v_{21} vi_{44}) + \gamma_{2}^\prime * v_{21}^* vi_{44}; M_{44\_22} = \text{TRACE}(v_{22} vi_{44}) + \gamma_{2}^\prime * v_{22}^* vi_{44};
\]

\[
M_{44\_31} = \text{TRACE}(v_{31} vi_{44}) + \gamma_{3}^\prime * v_{31}^* vi_{44}; M_{44\_32} = \text{TRACE}(v_{32} vi_{44}) + \gamma_{3}^\prime * v_{32}^* vi_{44};
\]

\[
M_{44} = (M_{44\_11} \| M_{44\_12} \| M_{44\_13}) / (M_{44\_21} \| M_{44\_22} \| M_{44\_23} \| M_{44\_24}) / (M_{44\_31} \| M_{44\_32} \| M_{44\_33});
\]

\[
M = (M_{11} \| M_{12} \| M_{13} \| M_{14}) / (M_{21} \| M_{22} \| M_{23} \| M_{24}) / (M_{31} \| M_{32} \| M_{33} \| M_{34}) / (M_{41} \| M_{42} \| M_{43} \| M_{44});
\]

\[
\text{INF} = M; \quad H_1 = \text{DET}(\text{INF}); \quad c = \text{LOG}(H_1);
\]

\[
*\text{InfMat} = H_1; \quad *c = \text{LOG}(H_1); \quad *\text{print i k n c};
\]

\[
\text{if } c > D \text{ then } DM = x;
\]

\[
\text{if } c > D \text{ then } D = c;
\]

\[
x = DM; \quad \text{diff} = D - D_1;
\]

\[
\text{if } \text{diff} = 0 \text{ then } i = 300;
\]

\[
D_1 = D;
\]

\[
\text{end}; \quad \text{if } D_1 > D_2 \text{ then } x_{\text{final1}} = x;
\]

\[
\text{if } D_1 > D_2 \text{ then } D_2 = D_1;
\]

\[
\text{end}; \quad x = x_{\text{final1}};
\]

\[
*xpx = x^\top x;
\]

\[
*\text{print xpx}; \quad /*x1=x[1, ];x2=x[2, ];x3=x[3, ];x4=x[4, ];x5=x[5, ];x6=x[6, ];x7=x[7, ];x8=x[8, ];x9=x[9, ];x10=x[10, ];x11=x[11, ];x12=x[12, ];x13=x[13, ];x14=x[14, ];x15=x[15, ];x16=x[16, ];BK1 = x1/x2/x3/x4;BK2 = x5/x6/x7/x8;BK3 = x9/x10/x11/x12;BK4 = x13/x14/x15/x16;*/
\]

\[
*\text{print InfMat}; \quad \text{print } x \ c \ H_1;
\]

\[
\text{quit}; \quad \text{run};