2017

Efficient Baseline Utilization In Crossover Clinical Trials Through Linear Combinations Of Baselines: Parametric, Nonparametric, And Model Selection Approaches

Thomas Jemielita
University of Pennsylvania, thomasjemielita@gmail.com

Efficient Baseline Utilization In Crossover Clinical Trials Through Linear Combinations Of Baselines: Parametric, Nonparametric, And Model Selection Approaches

Abstract

In a crossover clinical trial, including period-specific baselines as covariates in a regression model is known to increase the precision of the estimated treatment effect. The potential efficiency gain depends, in part, on the true model, the distribution and covariance matrix of the vector of baselines and outcomes, and the model chosen for analysis. We examine improvements in power that can be achieved by incorporating optimal linear combination of baselines (LCB). For a known distribution, the optimal LCB minimizes the conditional variance corresponding to a treatment effect. The use of a single metric to capture the information in the baseline measurements is appealing for crossover designs. Because of their efficiency, crossover designs tend to have small sample sizes and thus the number of covariates in a model can significantly impact the degrees of freedom in the analysis. We start by examining optimal LCB models under a normality assumption for uniform and incomplete block designs. For uniform designs, such as the AB/BA design, estimation is entirely through within-subject contrasts (and thus ordinary least squares [OLS]) and the optimal LCB minimizes the conditional variance corresponding to the treatment effect. However, since the optimal LCB is a function of the unknown covariance matrix, we propose an adaptive method that uses the LCB covariate corresponding to the most plausible covariance structure guided by the data. For incomplete block designs, data are commonly analyzed using a mixed effects model. Treatment effect estimates from this analysis are complex functions of both within-subject and between-subject treatment contrasts. To improve efficiency, we propose incorporating period-specific optimal LCBs which minimize the conditional variance of the period-specific outcomes. A simpler fixed effects analysis of covariance involving only within-subject contrasts is also described for small sample situations. In the latter, hypothesis tests based on the mixed effects analyses exhibit inflated type I error rates even when using a Kenward and Rogers approach to adjust the degrees of freedom. Lastly, we extend this work to the more general setting where the optimal LCB depends on the distribution of the response vector. In practice, the distribution is unknown and the optimal LCB is estimated under some loss function. To handle both normal and non-normal response data, OLS and a rank-based nonparametric regression model (R-estimation), are considered. A data-driven approach is then proposed which adaptively chooses the best fitting model among a set of models which work well under a range of conditions. Relative to commonly used methods, such as change from baseline analyses without use of covariates, our methods using functions of baselines as period-specific or period-invariant covariates consistently demonstrate improved power across a number of crossover designs, covariance structures, and response distributions.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Epidemiology & Biostatistics

First Advisor

Mary E. Putt

EFFICIENT BASELINE UTILIZATION IN CROSSOVER CLINICAL TRIALS THROUGH LINEAR COMBINATIONS OF BASELINES: PARAMETRIC, NONPARAMETRIC, AND MODEL SELECTION APPROACHES

Thomas O. Jemielita

A DISSERTATION

in

Epidemiology and Biostatistics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2017

Supervisor of Dissertation	Co-Supervisor of Dissertation
Mary E. Putt	Devan V. Mehrotra, Adjunct Associate Professor of Biostatistics
Professor of Biostatistics	Associate Vice President in Biostatistics, Merck & Co.

Graduate Group Chairperson

Nandita Mitra, Professor of Biostatistics

Dissertation Committee

Kathleen J. Propert, Professor of Biostatistics

Justine Shults, Professor of Biostatistics

Michelle Denburg, Assistant Professor of Pediatrics at the Children's Hospital of Philadelphia

Inna Chervoneva, Associate Professor of Biostatistics at Thomas Jefferson University

EFFICIENT BASELINE UTILIZATION IN CROSSOVER CLINICAL TRIALS THROUGH LINEAR COMBINATIONS OF BASELINES: PARAMETRIC, NONPARAMETRIC, AND MODEL SELECTION APPROACHES

ACKNOWLEDGEMENT

My experience at the University of Pennsylvania has been amazing. To start, I would like to thank my two co-advisors Mary Putt and Devan Mehrotra. I really have learned a lot from Mary and Devan and couldn't imagine having better advisors. I appreciate the time, knowledge, and wisdom they put into helping me develop a quality dissertation. Their continual guidance and expertise has truly made the dissertation experience rewarding. I would also like to express my gratitude to the rest of my committee members, Dr. Kathleen Propert, Dr. Justine Shults, Dr. Michelle Denburg, and Dr. Inna Chervoneva, for their feedback and support throughout the dissertation process. Their insights were essential for this dissertation research.

I also want to thank Dr. Jason Roy and Dr. Kathleen Propert for their support during my masters thesis. I would like to again thank Dr. Michelle Denburg and Dr. Justine Shults for their guidance on the renal training grant. I also would like to thank Dr. Richard Landis and Dr. Alisa Stephens for their mentorship through my tenure on the MAPP research group. To the whole MAPP research group, thank you for the valuable experience and fun memories. A special note of recognition goes to all the Upenn biostats students who made my time at Upenn truly memorable, both at school and outside of school.

Finally, I need to thank my incredible wife Brittany Oakes Jemielita, who I met at Upenn, for her love and support throughout my PhD studies. I could always count on her for help and motivation and I couldn't have done this without her. I would also like to thank my parents, Kim Severson and Philip Jemielita, for their never-ending support and advice. I would also like to thank the following friends and family: Matthew Jemielita, Isaac Jemielita, Hector Jemielita, Ollie Jemielita, Sarah Morejohn, Tom Oakes, Elizabeth Oakes, and Tommy Oakes. Thank you for everything.

ABSTRACT

EFFICIENT BASELINE UTILIZATION IN CROSSOVER CLINICAL TRIALS THROUGH LINEAR COMBINATIONS OF BASELINES: PARAMETRIC, NONPARAMETRIC, AND MODEL SELECTION APPROACHES

Thomas O. Jemielita

Mary E. Putt

Devan V. Mehrotra

In a crossover clinical trial, including period-specific baselines as covariates in a regression model is known to increase the precision of the estimated treatment effect. The potential efficiency gain depends, in part, on the true model, the distribution and covariance matrix of the vector of baselines and outcomes, and the model chosen for analysis. We examine improvements in power that can be achieved by incorporating optimal linear combination of baselines (LCB). For a known distribution, the optimal LCB minimizes the conditional variance corresponding to a treatment effect. The use of a single metric to capture the information in the baseline measurements is appealing for crossover designs. Because of their efficiency, crossover designs tend to have small sample sizes and thus the number of covariates in a model can significantly impact the degrees of freedom in the analysis. We start by examining optimal LCB models under a normality assumption for uniform and incomplete block designs. For uniform designs, such as the AB/BA design, estimation is entirely through within-subject contrasts (and thus ordinary least squares [OLS]) and the optimal LCB minimizes the conditional variance corresponding to the treatment effect. However, since the optimal LCB is a function of the unknown covariance matrix, we propose an adaptive method that uses the LCB covariate corresponding to the most plausible covariance structure guided by the data. For incomplete block designs, data are commonly analyzed using a mixed effects model. Treatment effect estimates from this analysis are complex functions of both within-subject and between-subject treatment contrasts. To improve efficiency, we propose incorporating period-specific optimal LCBs which minimize the conditional variance of the period-specific outcomes. A simpler fixed effects analysis of covariance involving only within-subject contrasts is also described for small sample situations. In the latter, hypothesis tests based on the mixed effects analyses exhibit inflated type

I error rates even when using a Kenward and Rogers approach to adjust the degrees of freedom. Lastly, we extend this work to the more general setting where the optimal LCB depends on the distribution of the response vector. In practice, the distribution is unknown and the optimal LCB is estimated under some loss function. To handle both normal and non-normal response data, OLS and a rank-based nonparametric regression model (R-estimation), are considered. A data-driven approach is then proposed which adaptively chooses the best fitting model among a set of models which work well under a range of conditions. Relative to commonly used methods, such as change from baseline analyses without use of covariates, our methods using functions of baselines as period-specific or period-invariant covariates consistently demonstrate improved power across a number of crossover designs, covariance structures, and response distributions.

TABLE OF CONTENTS

ACKNOWLEDGEMENT iii

ABSTRACT iv

LIST OF TABLES viii

LIST OF ILLUSTRATIONS ix

CHAPTER 1 : INTRODUCTION 1

CHAPTER 2 : UNIFORM DESIGNS 3

2.1 Motivation and Literature Review 3

2.2 Model and Notation 5

2.3 Choosing the Optimal Linear Combination of Baselines (LCB) 8

2.4 Estimation 10

2.5 Application to Uniform Crossover Designs 12

2.6 Application to Data Analysis 17

2.7 Simulations 19

2.8 Discussion 25

CHAPTER 3 : INCOMPLETE BLOCK DESIGNS 28

3.1 Motivation and Literature Review 28

3.2 Setup and Notation 30

3.3 Baseline Models 32

3.4 Estimation 41

3.5 Application to a Clinical Trial 43

3.6 Simulations 43

3.7 Discussion 46

CHAPTER 4 : EFFICIENT BASELINE UTILIZATION IN CROSSOVER DESIGNS USING BOTH PARAMETRIC AND NON-PARAMETRIC REGRESSIONS 50

4.1 Motivation and Literature Review 50
4.2 Models and Notation 53
4.3 Estimation and Hypothesis Testing 55
4.4 Selecting the LCB 58
4.5 Model Selection and Inference: A Bootstrap Approach 60
4.6 Application of Methods 62
4.7 Simulations 66
4.8 Real Data Analysis 73
4.9 Discussion 77

CHAPTER 5 : CONCLUSION 80

APPENDICES 82

BIBLIOGRAPHY 113

LIST OF TABLES

TABLE 2.1 : Covariance Structure Assumptions: CS, DCS, EP, and AR(1) 14
TABLE 2.2 : Uniform Design: Optimal LCB by Design and Covariance Structure 16
TABLE 2.3 : Uniform Design Baseline Models: 2×2 Real Data Example I (N=20) 18
TABLE 2.4 : Uniform Design Baseline Models: 3×3 Real Data Example (N=24) 19
TABLE 2.5 : 2×2 Simulations: Under the Alternative Hypothesis, Power 22
TABLE 2.6 : 3×3 Simulations: Under the Alternative Hypothesis, Power 23
TABLE 2.7 : 4×4 Simulations: Under the Alternative Hypothesis, Power 24

TABLE 3.1 : Baseline Utilization Methods: Incomplete Block Design 39
TABLE 3.2 : Real Data Example: 3×2 Crossover Design 43
TABLE 3.3 : 3×2 Simulations: Under the Null Hypothesis, Type I Error 46
TABLE 3.4 : 3×2 Simulations: Under the Alternative Hypothesis, Power 47

TABLE 4.1 : Crossover Design Optimal Modeling Strategies 60
TABLE 4.2 : Parametric and Nonparametric Crossover Models 64
TABLE 4.3 : 2×2 Simulations: Estimated LCBs 68
TABLE 4.4 : Parametric and Nonparametric Comparisons: 2×2 Real Data Example I (N=20) 75
TABLE 4.5 : Parametric and Nonparametric Comparisons: 2×2 Real Data Example II (N=24) 76
TABLE 4.6 : Parametric and Nonparametric Comparisons: 3×3 Real Data Example (N=24) 77

TABLE A.1 : Notation 82

TABLE B.1 : Treatment Effect Sizes: 2×2, 3×3 and 4×4 Design 92
TABLE B.2 : 2×2 Simulations: Under the Null Hypothesis, Type I Error 94
TABLE B.3 : 3×3 Simulations: Under the Null Hypothesis, Type I Error 95
TABLE B.4 : 4×4 Simulations: Under the Null Hypothesis, Type I Error 96

TABLE D.1 : Parametric and Nonparametric Comparisons, 2×2 Simulations: Type I Error 109
TABLE D.2 : Parametric and Nonparametric Comparisons, 2×2 Simulations: Power . . . 110
TABLE D.3 : Parametric and Nonparametric Comparisons, 3×3 Simulations: Type I Error 111
TABLE D.4 : Parametric and Nonparametric Comparisons, 3×3 Simulations: Power . . . 112

LIST OF ILLUSTRATIONS

FIGURE 4.1 : 2x2 Simulations: Benchmark Comparisons 69
FIGURE 4.2 : 2x2 Simulations: Min-P Comparisons 70
FIGURE 4.3 : 3x3 Simulations: Benchmark Comparisons 71
FIGURE 4.4 : 3x3 Simulations: Min-P Comparisons 72
FIGURE 4.5 : Normal Q-Q Plots: Crossover Design Real Data Examples 74

CHAPTER 1

INTRODUCTION

A crossover trial is a repeated measures design in which patients receive sequences of treatments administered over some number of pre-specified periods. In contrast to a parallel group trial where patients are randomized to specific treatment arms, crossover designs randomize subjects by sequence. One advantage of this design is that the estimate of a treatment effect is obtained wholly or mostly through within-subject contrasts. Relative to a parallel group trial, this leads to large efficiency gains and thus fewer enrolled subjects are required.

The crossover design is defined by the chosen sequences of treatments. A sequence depends on the number of periods and the ordering of the considered treatments. For example, in the AB/BA or 2×2 (2-treatment, 2-period) design, subjects are randomized to either sequence AB or sequence BA. In sequence AB, subjects receive treatment A followed by treatment B. The order is reversed for subjects in sequence BA. The primary disadvantage of a crossover design is the possibility that a treatment effect can linger into the following period. This is called carryover and can complicate the estimation of treatment effects (Jones and Kenward, 2003). A washout period is typically included between periods to minimize the risk of carryover. In a pharmaceutical setting, pharmacokinetics can be used to determine an appropriate length for the washout, such that any carrover is mitigated. For the purposes of this dissertation, carryover is assumed to be null, or where the washout periods are sufficient.

For a design with a continuous outcome of interest (e.g. blood pressure), the outcome is typically measured prior to the period-specific treatment administration. These measurements are called period-specific baselines, or just baselines. For the AB/BA design, each subject has two baseline measurements and two post-treatment or outcome measurements. Often, the baseline and post-treatment measurements are at least moderately correlated (Kenward and Roger, 2010; Mehrotra, 2014). Thus, including the baselines as covariates in a regression model has the potential to reduce the standard error of a treatment estimate and increase the overall power to detect a treatment effect. The overall goal of this research is to efficiently incorporate baselines into the analysis of a general crossover design.

There has been considerable work in baseline utilization in crossover designs, primarily for the AB/BA design. Our work expands on this previous research and builds a general framework for efficient baseline utilization for a variety of designs and regression models. Specifically, we examine improvements in power that can be achieved by incorporating optimal linear combinations of baselines (LCB). For a known distribution, the optimal LCB minimizes the conditional variance corresponding to a treatment effect. Further, given that crosover designs typically have small sample sizes, the number of covariates in a regression model can significantly impact the degrees of freedom and consequently the efficiency of a hypothesis test. Thus, the optimal LCB preserves the limited number of degrees of freedom while explicitly reducing the variance of a treatment effect estimate. Overall, this can greatly increase the efficiency of a hypothesis test. Compared to standard baseline models such as the change from baseline model, our proposed LCB baseline models yield substantial efficiency gains.

Chapters 2-3 respectively discuss efficient baseline utilization under a normality assumption for uniform and incomplete block designs. For the uniform design, estimation of a treatment effect comes entirely from within-subject contrasts and the optimal LCB minimizes the conditional variance related to the contrasts. For the incomplete block design, estimation of a treatment effect does not necessarily need to come from only within-subject contrasts. Consequently, baseline utilization is explored in the framework of both mixed models, where estimation is a weighted summation of within-subject and between-subject information, and also models which only use within-subject information. Chapter 4 extends this work to a more general setting where the optimal LCB depends on some known distribution. A framework is developed for efficient baseline utilization under a general regression model. This naturally extends the research in Chapters 2-3 to non-normal distributions. For practical implementation, data-driven adaptive methods are proposed in all cases. Lastly, Chapter 5 summarizes the overall findings and main points. A summary of most notations, acronyms, and methods for this paper can be found in Appendix A.

CHAPTER 2

UNIFORM DESIGNS

2.1. Motivation and Literature Review

The research in this Chapter appears in Statistics in Medicine (Jemielita, Putt, and Mehrotra, 2016).

One of the most commonly used crossover designs is the uniform design. A uniform design is both uniform within sequence, each treatment appears the same number of times within each sequence, and uniform within period, each treatment appears the same number of times within each period. For example, the AB/BA design is considered uniform since each treatment appears once in each sequence and once in each period. In general, designs that are both uniform within sequence and uniform within period and thus uniform are efficient under commonly used models since all estimation is through within-subject constrasts.

Over the years, a number of publications have considered incorporating baseline measurements into the analysis of crossover designs. One of the recurring themes in this literature involves the importance of the underlying covariance structure in determining how incorporating baselines into the analysis can improve the precision of the treatment estimate. Hills and Armitage first noted that the correlation between the baselines and outcomes was a driving factor in deciding whether or not to use baselines (Armitage and Hills, 1982; Hills and Armitage, 1979). Kenward and Jones explored different estimation techniques when incorporating baselines, including least squares and generalized least squares (GLS) (Kenward and Jones, 1987). Building on earlier work, Kenward and Roger established a clear theoretical framework to illustrate how the underlying covariance structure of the baselines and outcomes influence bias and efficiency when the baselines are incorporated into the analysis using several different methods (Kenward and Roger, 2010). Metcalfe explored using baselines as covariates in the AB/BA design through analysis of covariance (ANCOVA) using the difference between the baselines at period 1 and period 2 as a covariate (Metcalfe, 2010). Metcalfe empirically showed that this covariate yielded improved efficiency across a number of covariance structures, relative to using change scores (post-treatment minus the baselines) or using post-treatment measurements only. Chen, Meng and Zhang examined joint modeling of the baseline and post-treatment outcomes as a way to utilize baseline data to increase efficiency

(Chen, Meng, and Zhang, 2012). This work, through theoretical arguments and empirical simulations, illustrated that including baselines in the analysis could improve the efficiency of a treatment effect estimate. Most recently, Mehrotra in agreement with previous authors, showed that the potential efficiency gained by using baselines as covariates is highly influenced by the covariance structure of the baselines and post-treatment outcomes (Mehrotra, 2014). Using theory and simulations, Mehrotra examined ten different baseline utilization methods for the AB/BA crossover design across a number of underlying baseline and outcome covariance structures and sample sizes. His final recommendation was to use the difference between baselines at period 1 and period 2 as a covariate in ANCOVA.

Our review of the literature thus suggests that the potential gains in efficiency that result from incorporating baselines into an analysis demonstrate a strong model-dependence, both in terms of the structure of the fixed effects model and the covariance structure. In the research reported here we use the simplest possible model for the carryover, and assume that carryover is eliminated by the washout. This is especially reasonable in a pharmaceutical setting, since pharmacokinetics can be used to determine an appropriate length for the washout period. The current work builds on previous findings by Mehrotra and Metcalfe (Mehrotra, 2014; Metcalfe, 2010). For the AB/BA under a model with no carryover and a number of difference covariance structures, these authors showed that using the using the difference in period-specific baselines as a covariate offered increased precision for the estimate of the treatment effect.

The use of a single linear combination to capture the information in the baseline measurements is appealing for crossover designs. Because of their efficiency, crossover studies often use small sample sizes, and thus the number of covariates in the model can significantly impact the degrees of freedom in the analysis. We begin by developing a theoretical framework to determine an optimal linear combination of baselines for uniform designs, first under an unstructured covariance assumption, and then using several different plausible assumptions for the covariance structure. Because the covariance structure in a data analysis is unknown, we develop a data-based 'adaptive' method to choose the optimal covariate. We then apply this work to the AB/BA design and to the three and four-period uniform designs, where the commonly used compound symmetry assumption is increasingly unlikely to realistically represent the covariance structure of data obtained in practice. Overall, we will show that, relative to commonly used methods, using linear combinations of base-

lines can lead to significant efficiency gains.

The model and notation for a general uniform crossover design are defined in Section 2.2. In Section 2.3, we describe the proposed method for choosing a baseline covariate. Section 2.4 covers estimation for the proposed method. Section 2.5 covers the application of the proposed methods for 2×2, 3×3, and 4×4 crossover designs and explores various plausible covariance structures for the post-treatment measurements and baselines. Additionally, we describe the optimal baseline covariates under each covariance structure and crossover design and illustrate how to implement our approach using a data driven 'adaptive' method. In Section 2.6, we evaluate our proposed methods on real data sets. In section 2.7, our proposed methods are evaluated through simulations for 2×2, 3×3, and 4×4 crossover designs. Lastly, in section 2.8, we summarize the overall findings.

2.2. Model and Notation

Assume a uniform crossover design. Let:

$$\begin{aligned} \mathbf{X}_{ik} &= (X_{iAk}, ..., X_{iZk})^T \\ \mathbf{Y}_{ik} &= (Y_{iAk}, ..., Y_{iZk})^T \end{aligned} \tag{2.1}$$

be distinct Z-vectors of baseline and outcome measurements respectively, where $i = 1, .., s$ indexes sequence, $d = A, ..., Z$ indexes treatment, and $j = 1, ..., p$ indexes the period. For this Chapter, we focus on uniform crosssover designs where the number of periods equals the number of treatments ($p = Z$). Lastly, $k = 1, ..., n_i$ indexes subject k in sequence i, where subjects are assumed to be independent of each other. Initially, we consider a 'sequence-invariant' approach where we order the outcomes and baselines within each sequence by treatment. Later, as well as in Chapters 3-4, we generalize our approach to the case where each subject retains the vector of outcomes in the order in which they are received. We then assume multivariate normality such that:

$$\begin{pmatrix} \mathbf{X}_{ik} \\ \mathbf{Y}_{ik} \end{pmatrix} \sim N\left(\begin{pmatrix} E(\mathbf{X}_{ik}) \\ E(\mathbf{Y}_{ik}) \end{pmatrix}, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{XY}^T & \Sigma_{YY} \end{pmatrix} \right) \tag{2.2}$$

The elements of $E(\boldsymbol{Y}_{ik})$ are defined by a saturated cell means model (Chinchilli and Esinhart, 1996; Vonesh and Chinchilli, 1997), while placing no restrictions on $E(\boldsymbol{X}_{ik})$. The expectations

under a null carryover assumption are:

$$E(Y_{idk}) = \mu + \tau_d + \gamma_{id} \tag{2.3}$$

$$E(X_{idk}) = \zeta_{id} \tag{2.4}$$

where μ is the overall mean, τ_d is the effect of treatment d with $\sum_d \tau_d = 0$, γ_{id} is a fixed effect for treatment d within sequence i with $\sum_i^s \gamma_{id} = 0$ for all d, and ζ_{id} is fixed effect for the mean of the baseline at sequence i with treatment d. γ_{id} is a nuisance parameter, but ultimately represents sequence by period interactions nested within treatment (Chinchilli and Esinhart, 1996; Vonesh and Chinchilli, 1997). Note that the expectations of the baselines (2.4) could depend on the period and sequence in which treatment d was administered in. This allows for the possibility of period effects.

When $V((\boldsymbol{X}_{ik}, \boldsymbol{Y}_{ik})^T)$ and its sub-matrices are assumed to be sequence invariant, the general form of the covariance matrix for either the baselines (Σ_{XX}) or the outcomes (Σ_{YY}) is written:

$$\Sigma_* = \begin{pmatrix} \sigma_A^2 & \rho_{AB}\sigma_A\sigma_B & \cdots & \cdots & \rho_{AZ}\sigma_A\sigma_Z \\ & \sigma_B^2 & \rho_{BC}\sigma_B\sigma_C & \cdots & \rho_{BZ}\sigma_B\sigma_Z \\ & & \cdots & \cdots & \cdots \\ & & & & \sigma_Z^2 \end{pmatrix} \tag{2.5}$$

where $\Sigma_* = \Sigma_{XX}$ or $\Sigma_* = \Sigma_{YY}$. Within this general framework, the variance components of the baselines and outcomes are denoted as $(\sigma_d^X)^2$ and $(\sigma_d^Y)^2$ and the correlation coefficients by $\rho_{dd'}^X$ and $\rho_{dd'}^Y$. Here the superscripts denote baseline or outcome and the subscripts denote treatment. The covariance matrix between baselines and outcomes, is again sequence invariant, i.e.,

$$\Sigma_{XY} = \begin{pmatrix} \rho_{AA}^{XY}\sigma_A^X\sigma_A^Y & \rho_{AB}^{XY}\sigma_A^X\sigma_B^Y & \cdots & \rho_{AZ}^{XY}\sigma_A^X\sigma_Z^Y \\ \rho_{BA}^{XY}\sigma_B^X\sigma_A^Y & \rho_{BB}^{XY}\sigma_B^X\sigma_B^Y & & \vdots \\ & & \ddots & \vdots \\ & & & \rho_{ZZ}^{XY}\sigma_Z^X\sigma_Z^Y \end{pmatrix} \tag{2.6}$$

The correlations coefficients in (2.6) denote either the correlation between a baseline and an outcome for the same treatment i.e., ρ_{dd}^{XY} or the baseline and outcome for different treatments i.e., $\rho_{dd'}^{XY}$. While we later consider a single linear combination, we begin by considering up to $Z^\star$ linear

combinations of baselines (LCB). In the most general form, where there are potentially $q = 1, ..., Z^*$ LCBs per sequence ($Z^* \leq Z$); let the $Z^* \times Z$ matrix of coefficients be:

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_{Z^*}^T \end{pmatrix} \text{ where } \mathbf{a}_q^T = \begin{pmatrix} a_{q1} & ... & a_{qZ} \end{pmatrix}$$

Initially, $\boldsymbol{A}$ is constant for all sequences. Later we generalize this to allow different $\boldsymbol{A}_i$'s for different sequences. It follows that $\boldsymbol{A}\boldsymbol{X}_{ik}$ is a $Z^\star$-length vector, with each element representing a unique LCB. For example, if we consider the 2×2 crossover design, with $q = 2$:

$$\mathbf{A}\mathbf{X}_{ik} = \begin{pmatrix} \boldsymbol{a}_1^T \boldsymbol{X}_{ik} \\ \boldsymbol{a}_2^T \boldsymbol{X}_{ik} \end{pmatrix} = \begin{pmatrix} a_{11} X_{iAk} + a_{12} X_{iBk} \\ a_{21} X_{iAk} + a_{22} X_{iBk} \end{pmatrix}$$

This example would then yield different baseline covariates. Next, given (2.2), the distribution of the outcomes conditional on the vector of LCBs is:

$$\mathbf{Y}_{ik} | \mathbf{A}\mathbf{X}_{ik} \sim N\Big(E(\mathbf{Y}_{ik}) - \Sigma_{XY}^T \mathbf{A}^T (\mathbf{A}\Sigma_{XX}\mathbf{A}^T)^{-1} (\mathbf{A}E(\mathbf{X}_{ik}) - \mathbf{A}\mathbf{X}_{ik}), \quad \Sigma_{YY} - \Sigma_{XY}^T \mathbf{A}^T (\mathbf{A}\Sigma_{XX}\mathbf{A}^T)^{-1} \mathbf{A}\Sigma_{XY} \Big) \tag{2.7}$$

Notably, this general framework also allows for additional pre-treatment or baseline covariates to be considered. For example, say we wanted to include a covariate for age in the analysis. $\boldsymbol{X}_{ik}$ would then include the pre-treatment baselines and age, while $\boldsymbol{A}$ would be be a $Z^\star \times (Z + 1)$ matrix of coefficients. Moreover, while a covariate for age is likely constant across the study, time-varying covariates could also be considered in this framework. For example, there could be additional lab tests done prior to treatment administration in each period. Lastly, make the simplification that $q = 1$, and condition on a single LCB for each sequence so that $\boldsymbol{A} = \boldsymbol{a}^T = (a_1, ..., a_Z)$. We choose $\boldsymbol{a}^T$ such that the inclusion of $\boldsymbol{a}^T \boldsymbol{X}_{ik}$ as a covariate in a regression model minimizes the variance of the estimate of a pair-wise treatment difference.

2.3. Choosing the Optimal Linear Combination of Baselines (LCB)

2.3.1. Treatment-Ordered Approach

Define $\boldsymbol{b}$ as a Z-length vector such that $\boldsymbol{bY}_{ik}$ yields a contrast of interest. In general, we could consider any linear contrast of interest. Here we focus on the pairwise differences such that $\boldsymbol{b} = (1, -1, 0, ...0)$ and $\boldsymbol{bY}_{ik} = Y_{iAk} - Y_{iBk}$. For a uniform crossover design, in the absence of baselines, the unbiased estimate of the treatment difference ($\widehat{\tau_A - \tau_B}$), is simply the mean of the within-subject contrasts. For example, given (2.3):

$$E(\widehat{\tau_A - \tau_B}) = E\big(\frac{1}{s}\sum_{i=1}^{s}\frac{1}{n_i}\sum_{k=1}^{n_i}(Y_{iAk} - Y_{iBk})\big) = \tau_A - \tau_B$$

We now condition on an LCB, $\boldsymbol{a}^T\boldsymbol{X}_{ik}$. To show that the estimate remains unbiased after conditioning on $\boldsymbol{a}^T\boldsymbol{X}_{ik}$, note that from (2.7) with $\boldsymbol{A} = \boldsymbol{a}^T$ and letting $\boldsymbol{\beta} = (\beta_A, ..., \beta_Z)^T = \Sigma_{XY}^T\boldsymbol{a}(\boldsymbol{a}^T\Sigma_{XX}\boldsymbol{a})^{-1}$, it follows that the conditional means within the ith sequence are:

$$E(Y_{idk}|\boldsymbol{a}^T\boldsymbol{X}_{ik}) = \mu + \tau_d + \gamma_{id} + \beta_d(\boldsymbol{a}^T\boldsymbol{X}_{ik} - E(\boldsymbol{a}^T\boldsymbol{X}_{ik}))$$
$$E(Y_{iAk} - Y_{iBk}|\boldsymbol{a}^T\boldsymbol{X}_{ik}) = \tau_A - \tau_B + \gamma_{iA} - \gamma_{iB} + (\beta_A - \beta_B)(\boldsymbol{a}^T\boldsymbol{X}_{ik} - E(\boldsymbol{a}^T\boldsymbol{X}_{ik}))$$

Thus, the unconditional expectation is:

$$E_X\Big(E\big(\frac{1}{s}\sum_{i=1}^{s}\frac{1}{n_i}\sum_{k=1}^{n_i}(Y_{iAk} - Y_{iBk}|\boldsymbol{a}^T\boldsymbol{X}_{ik})\big)\Big) = \tau_A - \tau_B \tag{2.8}$$

Next, given (2.7), the general form of the variance of the some linear contrast conditional on an LCB is:

$$V(\boldsymbol{bY}_{ik}|\boldsymbol{a}^T\boldsymbol{X}_{ik}) = V(\boldsymbol{bY}_{ik}) - \frac{(\boldsymbol{b}\Sigma_{XY}^T\boldsymbol{a})^2}{\boldsymbol{a}^T\Sigma_{XX}\boldsymbol{a}} \tag{2.9}$$

Then for our specific pairwise treatment contrast of interest:

$$V(Y_{iAk} - Y_{iBk}|\boldsymbol{a}^T\boldsymbol{X}_{ik}) = V(Y_{iAk} - Y_{iBk}) - \frac{\text{cov}(Y_{iAk} - Y_{iBk}, \boldsymbol{a}\boldsymbol{X}_{ik})^2}{\boldsymbol{a}^T\Sigma_{XX}\boldsymbol{a}} \tag{2.10}$$

The second terms in (2.9) and (2.10) are non-negative; thus the variance of the treatment effect will never be increased by conditioning on an LCB. The magnitude of any reduction in variance will depend on the structure of the covariance matrices as well as the linear combination ($\boldsymbol{a}^T$). Notably, (2.8-2.10) hold true even if there are additional baseline covariates in $\boldsymbol{X}_{ik}$ (ex: age). We now chose the linear combination, $\boldsymbol{a}_\star^T$, to minimize (2.10). To do this, we solve:

$$\boldsymbol{a}_\star^T = (a_1^\star, ..., a_Z^\star) = \underset{\boldsymbol{a}^T}{\text{Argmin}} V(Y_{iAk} - Y_{iBk} | \boldsymbol{a}^T \boldsymbol{X}_{ik}) \tag{2.11}$$

When the variance and covariance terms are known, we solve (2.11) either analytically using the partial derivatives of (2.10) with respect to $\boldsymbol{a}^T$, or iteratively using an optimization algorithm. The LCB chosen in this fashion is optimized for a specific design and covariance structure.

2.3.2. Period-Ordered Approach

Up until now, the methodology was developed using outcomes and baselines ordered by treatment. While this simplifies the notation, it is natural to think of the baselines/outcomes in terms of a temporal ordering defined by the periods in which the treatments are administered. Ordering by periods also allows us to consider an autoregressive (AR) covariance structure and explicitly model a decay in the correlation between successive measurements over time. Because we use a saturated model, the transition to a period-ordered model simply involves a re-parameterization of the fixed effects, and the interpretation of the fixed effects remains the same. Despite this, for sequences with more than two periods, the period-ordered model also yields sequence-specific LCBs, and from this perspective is somewhat more complicated.

Let $\boldsymbol{X}_{ik} = (X_{i1k}, ..., X_{ijk}, ..., X_{ipk})^T$ be the period ordered baselines, $\boldsymbol{Y_{ik}} = (Y_{i1k}, ..., Y_{ijk}, ..., Y_{ipk})^T$ be the period-ordered outcomes, and $\boldsymbol{W}_{ik} = (X_{i1k}, Y_{i1k}, ..., X_{ijk}, Y_{ijk}, ..., X_{ipk}, Y_{ipk})^T$ be the temporally ordered baseline-outcome pairs. We continue to use saturated models for the outcomes and baselines but with slightly different notation to match the period ordering:

$$\begin{aligned} E(Y_{ijk}) &= \mu + \tau_{d[i,j]} + \gamma_{id[i,j]} \\ E(X_{ijk}) &= \zeta_{ij} \end{aligned} \tag{2.12}$$

Under a period ordering, the treatment effect, $\tau_{d[i,j]}$ is defined with respect to sequence i and period

j. Further, $E(Y_{ijk})$ still depends on treatment and sequence, as in (2.3). Similarly, the expectation of the baselines is now defined with respect to sequence i and period j, allowing for period effects. Lastly, we again define the variance of the outcomes/baselines in terms of block matrices as in (2.5) and (2.6), but with the treatment designations (A through Z) replaced with period designations (1 through p).

In general, define $\boldsymbol{b}_i$ as a p-length vector such that $\boldsymbol{b}_i\boldsymbol{Y}_{ik}$ denotes the within-subject contrast for a pair of treatments in sequence i. With outcomes ordered by period, $\boldsymbol{b}_i$ depends on sequence i. For example, if we were comparing treatments A and B in an AB/BA design, in sequence AB, $\boldsymbol{b}_i\boldsymbol{Y}_{ik} = (1,-1)(Y_{i1k}, Y_{i2k})^T = Y_{i1k} - Y_{i2k}$, and in sequence BA, $\boldsymbol{b}_i\boldsymbol{Y}_{ik} = (-1,1)(Y_{i1k}, Y_{i2k})^T = Y_{i2k} - Y_{i1k}$. Next, since the within-subject contrast ($\boldsymbol{b}_i\boldsymbol{Y}_{ik}$) varies by sequence, the LCB may also vary by sequence. Consequently, let $\boldsymbol{a}_i^T = (a_{i1}, ..., a_{ip})^T$. Then, by replacing $\boldsymbol{A}$ with $\boldsymbol{a}_i^T$ in (2.7), it is then straightforward to show that:

$$V(\boldsymbol{b}_i\boldsymbol{Y}_{ik}|\boldsymbol{a}_i^T\boldsymbol{X}_{ik}) = V(\boldsymbol{b}_i\boldsymbol{Y}_{ik}) - \frac{(\boldsymbol{b}_i\Sigma_{XY}^T\boldsymbol{a}_i)^2}{\boldsymbol{a}_i^T\Sigma_{XX}\boldsymbol{a}_i} \tag{2.13}$$

As before, choose an LCB, $\boldsymbol{a}_{i\star}^T$, such that:

$$\boldsymbol{a}_{i\star}^T = (a_{i1}^\star, ..., a_{ip}^\star) = \underset{\boldsymbol{a}_i^T}{\text{Argmin}} V(\boldsymbol{b}_i\boldsymbol{Y}_{ik}|\boldsymbol{a}_i^T\boldsymbol{X}_{ik}) \tag{2.14}$$

2.4. Estimation

Under the treatment-ordered approach (Section 2.3.1), re-defining $\boldsymbol{\beta} = \Sigma_{XY}^T\boldsymbol{a}_\star(\boldsymbol{a}_\star^T\Sigma_{XX}\boldsymbol{a}_\star)^{-1}$, we fit the following linear mixed model:

$$Y_{idk} = \mu^\star + \tau_d + \gamma_{id} + \beta_d\boldsymbol{a}_\star^T\boldsymbol{X}_{ik} + \epsilon_{idk}^\star \tag{2.15}$$

where $\epsilon_{idk}^\star$ corresponds to the appropriate covariance term from (2.7) with $\boldsymbol{A} = \boldsymbol{a}_\star^T$ and $\mu^\star$ is the intercept for the conditional model. This linear mixed model can be estimated using generalized least squares where the variance parameters are estimated through restricted maximum likelihood. Alternatively, we could simply fit an ordinary least squares (OLS) model where the outcomes are

the appropriate within-subject contrasts. With the goal of estimating $\tau_A - \tau_B$, the OLS model is:

$$Y_{iAk} - Y_{iBK} = (\tau_A - \tau_B) + (\gamma_{iA} - \gamma_{iB}) + (\beta_A - \beta_B)\boldsymbol{a}_\star^T \boldsymbol{X}_{ik} + (\epsilon^\star_{iAk} - \epsilon^\star_{iBk}) \tag{2.16}$$

In this formulation, each subject has a single derived outcome. Furthermore, (2.15) and (2.16) will yield equivalent inference for $\widehat{\tau_A - \tau_B}$ if we assume that the conditional outcomes in (2.15) have an unstructured covariance structure. Under this assumption, the test for a pairwise treatment difference in the mixed model is exact under the assumption of normality (Chinchilli and Esinhart, 1996). Given the equivalent inference on $\widehat{\tau_A - \tau_B}$ between the mixed model and the OLS model, it follows that the OLS model also makes no assumptions about the underlying covariance structure. Furthermore, because misspecifying the covariance structure of the regression model can cause type I error inflation (Gurka, Edwards, and Muller, 2011), we assume an unstructured covariance structure as a robust approach.

Under the period-ordered approach (Section 2.3.2), there may be sequence-specific optimal LCBs (2.14) and by implication, sequence-specific optimal LCBs as covariates. However, for any two sequences with treatment A and B in the same two periods, the solution to (2.14) is sequence invariant. Additionally, the solution to (2.14) is not unique. For example, if $\boldsymbol{a}_{i\star}$ is a solution to (2.14), it is also true that $-\boldsymbol{a}_{i\star}$ is a solution to (2.14). Regardless, it may be inefficient to condition on multiple LCBs at small sample sizes. We thus assumed a common regression coefficient for all of the LCBs. This simplification yields unbiased estimates as long as we condition at the overall mean of the LCBs, or $E(\sum_i^s \boldsymbol{a}_{i\star}^T \boldsymbol{X}_{ik})$. Moreover, from simulation results (Section 2.7), assuming a common regression coefficient still results in efficiency gains. In other words, using the framework from the OLS model in (2.16), we model:

$$\boldsymbol{b}_i \boldsymbol{Y}_{ik} = (\tau_A - \tau_B) + (\gamma_{iA} - \gamma_{iB}) + (\beta_A - \beta_B)\boldsymbol{a}_{i\star}^T \boldsymbol{X}_{ik} + (\epsilon^\star_{iAk} - \epsilon^\star_{iBk}) \tag{2.17}$$

The hypotheses of interest are:

$$\begin{aligned} H_0 &: \tau_A - \tau_B = 0 \\ H_A &: \tau_A - \tau_B \neq 0 \end{aligned}$$

Depending on the LCB covariate decided on, we fit (2.16) or (2.17) to obtain OLS point estimates, $\hat{\boldsymbol{\theta}} = (\widehat{\tau_A - \tau_B}, \widehat{\gamma_{1A} - \gamma_{1B}}, ..., \widehat{\gamma_{sA} - \gamma_{sB}}, \hat{\beta})^T$, along with the corresponding estimated covariance matrix $\widehat{V(\hat{\theta})}$. Next, assume that the LCB covariate is centered at 0. Then, letting $\boldsymbol{L}^T = (1, \frac{1}{s}, ..., \frac{1}{s}, 0)$, it follows that:

$$E(\boldsymbol{L}^T\hat{\boldsymbol{\theta}}) = \tau_A - \tau_B$$

The test statistic is then:

$$t = \frac{\boldsymbol{L}^T\hat{\boldsymbol{\theta}}}{\sqrt{\boldsymbol{L}^T V(\hat{\boldsymbol{\theta}})\boldsymbol{L}}}$$

which we compare to a t-distribution with $(\sum_{i=1}^{s} n_i) - s - 1$ degrees of freedom (DF) for a model that includes a covariate and $(\sum_{i=1}^{s} n_i) - s$ DF for a model without a covariate. We note that in particular for small samples, any gain in efficiency due to including the covariate may be offset by the loss of a degree of freedom. For SAS users, we provide the OLS code below where ydiff_AB refers to $Y_{iAk} - Y_{iBk}$, seq refers to the parameters $\gamma_{iA} - \gamma_{iB}$, and LCB refers to an LCB covariate (centered at zero). Note that while the LCB is chosen to reflect an assumption about the covariance structure of the baseline and outcome measurements, the model fit and hypothesis test assumes an unstructured covariance structure that is identical for each individual.

```
PROC MIXED DATA=example_data;
CLASSES seq;
MODEL ydiff_AB = seq LCB;
ESTIMATE 'tau_A-tau_B' intercept 1 LCB 0 /CL;
RUN;
```

2.5. Application to Uniform Crossover Designs

2.5.1. Description of Designs

The specific designs of interest are:

- The 2×2 crossover design with sequences: AB, BA
- The 3×3 crossover design with sequences: ABC, BAC, CAB, CBA, ACB, BCA

- The 4×4 crossover design with sequences: ABCD, BDAC, CADB, DCBA

Note that each sequence, for any design, defines the ordering of the treatments. For example, sequence ABC, indicates that treatment A is administered in period 1, treatment B is administered in period 2, and treatment C is administered in period 3. For the 2×2 and 3×3 design, all possible treatment orderings occur. However, for a four treatment design, there are a maximum of 24 sequences. To simplify, we use the commonly used Williams Design with only 4 sequences. Notably, while these uniform designs have the same number of periods as the number of treatments, and hence are complete, our models still pertain to designs where $p \neq Z$. For example, our methods could be applied to the crossover design ABAB/BABA. However, for this design, it is unclear whether a period-ordered or treatment-ordered approach would be more efficient. Finally, we now apply our methods to each of these designs under several plausible covariance structures.

2.5.2. Plausible Covariance Structures

We consider four plausible covariance structures: Compound Symmetry (CS), Double Compound Symmetry (DCS), Equipredictability (EP), and Autoregressive(1) (AR(1)), as described in Table 2.1. CS, discussed by a variety of authors (Mehrotra, 2014; Metcalfe, 2010; Yan, 2012), is the most restrictive, assuming a single variance parameter for all measurements and a common correlation between all measurements. DCS was used in the work of Chen, Meng, and Zhang (Chen, Meng, and Zhang, 2012); it is similar to CS, but allows each baseline and outcome with the same treatment (or same period) to have a separate correlation. EP goes one step beyond DCS allowing each baseline and outcome with different treatments (different periods) to have a different correlation. EP, considered by (Mehrotra, 2014), is a simplified version of a six-parameter covariance structure described by Kenward and Roger (Kenward and Roger, 2010). In Table 2.1, CS, DCS, and EP are defined with respect to a treatment ordering, but these structures could be equivalently defined with respect to a period ordering. Indeed, for the CS, DCS, and EP covariance structures, the treatment-ordered and period-ordered approaches yield identical covariance structures (and identical optimal LCBs). This point is illustrated in Table 2.1. AR(1), used by Mehrotra and Metcalfe (Mehrotra, 2014; Metcalfe, 2010), is only used with temporally ordered baselines and outcomes ($\boldsymbol{W}_{ik}$). This assumes a common variance and a single correlation parameter. Note that in Table 2.1, $\boldsymbol{W}_{ik}[t]$ refers to the t^{th} element of the temporally ordered baselines and outcomes. Additionally, while our setup was for an unstructured (UN) covariance matrix, we do not consider optimal LCBs under

an UN covariance. This is because at small sample sizes, the variance components needed to estimate the optimal LCB are unstable. This resulting uncertainty will typically lead to type I error inflation.

Table 2.1: Covariance Structure Assumptions: CS, DCS, EP, and AR(1)

Structure	Assumption σ_d^2	$\rho^\star_{dd'}$	ρ^{XY}_{dd}	$\rho^{XY}_{dd'}$	Parameters	Comments
Σ_{CS}	σ^2	ρ	ρ	ρ	2	Common variance and common correlation for all pairwise measurements.
Σ_{DCS}	σ^2	ρ_2	ρ_1	ρ_2	3	Similar to CS, but allows baselines & outcomes within treatments (or periods) to have different correlations than between treatments (periods).
Σ_{EP}	σ^2	ρ_2	ρ_1	ρ_3	4	Similar to DCS, but allows baselines & outcomes between treatments (periods) to have different correlations.
Σ_{AR}	-	-	-	-	2	cov($\boldsymbol{W}_{ik}[t], \boldsymbol{W}_{ik}[t']) = \sigma^2\rho^{\lvert t-t'\rvert}$; Auto-regressive (1), two parameters.

Notes: $\rho^\star_{dd'}$ refers to either $\rho^{XX}_{dd'}$ $(corr(X_d, X_{d'}))$ or $\rho^{YY}_{dd'}$ $((corr(Y_d, Y_{d'})))$; $\rho^{XY}_{dd'} = corr(X_d, Y'_d)$. CS, DCS, and EP can be equivalently defined with respect to period ordering by substituing the d subscript (treatment-ordered covariance) with a j subscript (period-ordered covariance). $\boldsymbol{W}_{ik}[t]$ refers to the t^{th} element of the temporally ordered baselines/outcomes $(X_1, Y_1, ..., X_p, Y_p)$. CS = Compound Symmetry; DCS = Double Compound Symmetry; EP = Equipredictability; AR = Auto-regressive (1).

Next, Table 2.2 shows the optimal LCBs under these four considered structures. Derivation of the optimal LCBs can be found in Appendix B.1. Note that sequence and subject subscripts are dropped. Additionally, Table 2.2 shows the within-subject contrast that corresponds to the expected treatment effect, $\tau_A - \tau_B$. When the baselines and outcomes are ordered by treatment, this contrast is always $Y_A - Y_B$; when the baselines and outcomes are ordered by period, the contrast will differ by sequence.

Compound Symmetry (CS) Covariance Structure

As previously shown for the 2×2 design (Mehrotra, 2014), the variance conditional on the LCB (2.10) is

$$V(Y_{iAk} - Y_{iBk} | \boldsymbol{a}^T \boldsymbol{X}_{ik}, \Sigma_{CS}) = 2\sigma^2(1-\rho)$$

for each considered uniform design. This is simply the variance of $Y_{iAk} - Y_{iBk}$ and thus there is no LCB that improves the efficiency of the treatment estimate. If an LCB were included in the model, a hypothesis test or confidence interval would require an additional degree of freedom, thus reducing

the efficiency of the inferential procedure. This effect would be most pronounced in small samples, a frequent characteristic of crossover trials in practice. In general, when cov$(Y_{iAk} - Y_{iBk}, \boldsymbol{a}^T \boldsymbol{X}_{ik}) = 0$ for all $\boldsymbol{a}^T$, there is no LCB that improves the efficiency of the treatment estimate.

Double Compound Symmetry (DCS)/ Equipredictability (EP) Covariance Structure

For any of the considered designs, the conditional variances (2.10) under DCS and EP are:

$$V(Y_{iAk} - Y_{iBk} | \boldsymbol{a}^T \boldsymbol{X}_{ik}, \Sigma_{DCS}) = 2\sigma^2(1-\rho_2) - \frac{\sigma^4(a_1-a_2)^2(\rho_1-\rho_2)^2}{V(\boldsymbol{a}^T \boldsymbol{X}_{ik})}$$
$$V(Y_{iAk} - Y_{iBk} | \boldsymbol{a}^T \boldsymbol{X}_{ik}, \Sigma_{EP}) = 2\sigma^2(1-\rho_2) - \frac{\sigma^4(a_1-a_2)^2(\rho_1-\rho_3)^2}{V(\boldsymbol{a}^T \boldsymbol{X}_{ik})}$$

Notably, while $V(\boldsymbol{a}^T \boldsymbol{X}_{ik})$ depends on the specific design, $\boldsymbol{a}_\star^T = (1, -1, 0, ...0)$ minimizes the conditional variance in each case. Thus, $\boldsymbol{a}_\star^T \boldsymbol{X}_{ik} = X_{iAk} - X_{iBk}$ is the optimal LCB when the underlying covariance structure is DCS or EP. This result is also valid if we assume separate variances for the baselines and outcomes (i.e. $V(X_{idk}) = \sigma_X^2$, $V(Y_{idk}) = \sigma_Y^2$, for all d). Next, to obtain an unbiased estimate of $\tau_A - \tau_B$ (Section 2.4), we either need to condition at the sample mean of the LCB, $\frac{1}{n}\sum_i^s (X_{iAk} - X_{iBk})$, or shift $X_{iAk} - X_{iBk}$ by the sample mean and condition at zero. For simplicity, we refer to this LCB as $X_A - X_B$.

Autoregressive(1) [AR(1)] Covariance Structure

The AR(1) structure is only sensible for a period-ordered model. In this setting, we assume that the baselines/outcomes are ordered temporally and thus the within-subject contrast of interest differs by sequence. Table 2.2 shows the LCBs that minimize the conditional variance of the treatment effect for each sequence, under an AR(1) assumption. These optimal LCBs will simply be referred to as the AR(1) covariate. For the 2×2 and 3×3 designs, the optimal LCBs in pairs of sequences are just negatives of each other. The AR(1) covariate also depends on the AR(1) correlation, ρ. In practice, ρ will need to be estimated to use the AR(1) covariate. In Section 2.5.3, we provide an approach to estimate ρ. Finally, to obtain an unbiased estimate of the treatment effect, we either condition at the sample mean of the AR(1) covariate, or shift the AR(1) covariate by the sample mean and condition at zero.

Table 2.2: Uniform Design: Optimal LCB by Design and Covariance Structure

Crossover Design	Σ	Sequence	Contrast	Baseline Covariate ($\boldsymbol{a}_{i\star}^T\boldsymbol{X}$)
All	Σ_{CS}	All	$Y_A - Y_B$	No Baselines
All	Σ_{DCS}	All	$Y_A - Y_B$	$X_A - X_B$
All	Σ_{EP}	All	$Y_A - Y_B$	$X_A - X_B$
2x2	Σ_{AR}	AB	$Y_1 - Y_2$	$\rho^{-2}X_1 - X_2$
		BA	$Y_2 - Y_1$	$-(\rho^{-2}X_1 - X_2)$
3x3	Σ_{AR}	ABC	$Y_1 - Y_2$	$X_1 - X_3$
		BAC	$Y_2 - Y_1$	$-(X_1 - X_3)$
		CAB	$Y_2 - Y_3$	$\rho^{-2}X_2 - X_3$
		CBA	$Y_3 - Y_2$	$-(\rho^{-2}X_2 - X_3)$
		ACB	$Y_1 - Y_3$	$X_1 + X_2 - cX_3$
		BCA	$Y_3 - Y_1$	$-(X_1 + X_2 - cX_3)$
4x4	Σ_{AR}	ABCD	$Y_1 - Y_2$	$X_1 - X_3$
		BDAC	$Y_3 - Y_1$	$-(X_1 + X_2 - X_3 - X_4)$
		CADB	$Y_2 - Y_4$	$X_2 + X_3 - (1+\rho^2)X_4$
		DCBA	$Y_3 - Y_4$	$-(X_3 - \rho^{-2}X_4)$

Notes: $c = \frac{(1+\rho^2+\rho^4)+\sqrt{(1+\rho^2+\rho^4)^2+4\rho^2(1+\rho^2)^2}}{2(1+\rho^2)}$. CS = Compound Symmetry; DCS = Double Compound Symmetry; EP = Equipredictability; AR = Auto-regressive (1).

2.5.3. Adaptive Data-Based Approach

While the optimal LCB depends on the covariance structure, in practice the covariance structure is unknown. This motivates a data-based adaptive approach that chooses an analytic strategy based on the most likely covariance structure. Our adaptive approach, which is similar to Mehrotra's Method X (Mehrotra, 2014), is as follows:

1. For the given data set, fit four models of the treatment ordered baselines and measurements where $\Sigma = V(\boldsymbol{Z}_{ik})$ is assumed to be CS, EP, DCS, or UN (unstructured) and one model where the temporally ordered baselines/outcomes are assumed to follow an AR(1) covariance structure. All models use a saturated means model for the combined vector of baselines and outcomes

2. Obtain the corrected Akaike Information Criterion (AICC) for each model. AICC is a small sample correction of AIC (Hurvich and Tsai, 1989).

3. Use $X_A - X_B$ as the LCB except when (1) the AICC is smallest under CS; in this case use no baselines, or (2) the AICC is smallest under AR(1); in this case use the AR(1) covariate

(Table 2.2)

4. Fit the appropriate regression model based on the covariate choice from Step 3. If the AR(1) covariate is chosen, then estimate ρ based on the AR(1) model from Step 1.

We did not derive an LCB for UN, since this result rapidly becomes complex with increasing number of periods. However, given (2.9) and (2.13), reductions in the conditional variance can be obtained by including $X_A - X_B$ as a covariate. In Chapter 3 we derive the optimal LCB under an UN covariance for a 2-period design.

2.6. Application to Data Analysis

In this section, we apply the methods to two real data sets, one each from a 2×2 and 3×3 crossover design. For each data example, we consider the results obtained using each of the variance-minimizing LCBs described in Table 2.2, using no baselines, the adaptive method, and where change from baseline (CFB) scores are used as the outcome (and no baselines are included as covariates in the model). CFB is widely used in practice and is a natural benchmark for comparison of our proposed methods. As is typically done (Jones and Kenward, 2003), our CFB model includes random subject effects and fixed effects for period and treatment. Note that by assuming random subject effects, the underlying covariance structure of the CFB outcomes is assumed to be CS. Table 2.3 (2×2 design) and Table 2.4 (3×3 design) show the estimated unstructured covariance matrix of the treatment ordered baselines and outcomes, the adaptive AICC results, and the treatment effect estimates, standard errors, and p-values for all considered methods.

2.6.1. Real Data Example I: 2×2 *Crossover Design*

In this example (previously published as Example 2 from Mehrotra, 2014), a biomarker associated with renal function was measured for each of 20 subjects at baseline and after treatment in a 2×2 (AB/BA) crossover trial. The estimated treatment-ordered UN covariance matrix appears in Table 2.3; the upper left-hand corner corresponds to the estimates of Σ_{XX}, the lower right to Σ_{YY} and the upper right to Σ_{XY}. For this example, the AICC favored the AR(1) structure. Notably, the CS structure had the largest AICC relative to DCS, EP, and AR(1). This result suggests that using an LCB should improve the efficiency of the estimate, and indeed, the SEs and the p-values using either $X_A - X_B$ or the AR(1) covariate are reduced relative to no baselines. In this case, the

adaptive method uses the AR(1) covariate. However, while the adaptive method chooses the AR(1) covariate over $X_A - X_B$, the two LCB's yielded very similar results. Lastly, we note that the CFB approach, relative to the AR(1) covariate and $X_A - X_B$, yielded a larger estimated effect and SE, but a similar p-value.

Table 2.3: Uniform Design Baseline Models: 2×2 Real Data Example I (N=20)

$$\hat{\Sigma}_{UN} = \begin{array}{c} \begin{array}{cccc} X_A & X_B & Y_A & Y_B \end{array} \\ \left[\begin{array}{cc|cc} 0.31 & 0.12 & 0.19 & 0.15 \\ [0.62] & 0.13 & 0.12 & 0.14 \\ \hline [0.77] & [0.59] & 0.21 & 0.14 \\ [0.59] & [0.86] & [0.70] & 0.20 \end{array}\right] \end{array}$$

Σ	AICC
Σ_{CS}	73.5
Σ_{DCS}	69.9
Σ_{EP}	72.1
$\boldsymbol{\Sigma_{AR(1)}}$	**66.4**
Σ_{UN}	76.1

Method	Outcome	LCB	Estimate	SE	p-value
No Baselines	Y_{idk}	None	0.155	0.079	0.065
CFB	$Y_{ijk} - X_{ijk}$	None	0.235	0.092	0.0201
$X_A - X_B$	Y_{idk}	$X_{iAk} - X_{iBk}$	0.186	0.073	0.0212
AR(1)	Y_{ijk}	AR(1) Covariate	**0.190**	**0.075**	**0.0216**
Adaptive	Depends on AICC	Depends on AICC	**0.190**	**0.075**	**0.0216**

Notes: $\hat{\Sigma}_{UN}$ is the estimated unstructured covariance matrix of the treatment ordered responses. [] refer to correlation estimates. AICC values for the joint vector of responses under the various covariance structures are displayed. CFB=Change from Baseline. Method AR(1) refers to the covariates derived in Table 2.2 under AR(1). The adaptive method, based on the AICC values, chooses between methods No Baselines, $X_{iAk} - X_{iBk}$, and Method AR(1). CS = Compound Symmetry; DCS = Double Compound Symmetry; EP = Equipredictability; AR = Auto-regressive (1); UN=Unstructured.

2.6.2. Real Data Example II: 3×3 Crossover Design

The following example is available online at `http://www.stat.ufl.edu/CourseINFO/STA6167/crossoverSFLM.pdf`. This example contains data from a study which compared the effects on heart rate of three treatments: a test drug, a standard drug, and a placebo. These treatments were assigned to the six possible sequences, with four subjects in each sequence (N=24 total). For each of the three visits and for each subject, heart rate was measured one hour following administration of treatment. We illustrate the standard treatment compared to the placebo, but other pairwise comparisons show similar results.

The estimated covariance matrix under UN appears in Table 2.4. We note that for this example, correlations between different baselines and outcomes designated by the same treatment (or period) tend to have higher correlations than those from different treatments (or periods). In this example, the AICC favors the DCS structure and additionally, the CS structure had the largest AICC com-

pared to DCS, EP, and AR(1). This suggests that using an LCB should improve the efficiency of the estimate. Indeed, relative to no baselines, the SEs and p-values using either $X_A - X_B$ or the AR(1) covariate are reduced. Here the adaptive approach would choose $X_A - X_A$ as the LCB covariate. Lastly, $X_A - X_B$ has a noticeably lower SE and p-value than CFB, but has a slightly higher SE and p-value than the AR(1) covariate.

Table 2.4: Uniform Design Baseline Models: 3×3 Real Data Example (N=24)

$$\hat{\Sigma}_{UN} = \begin{array}{c} \begin{array}{cccccc} X_A & X_B & X_C & Y_A & Y_B & Y_C \end{array} \\ \left[\begin{array}{cccc|cc} 124.5 & 15.3 & 30.7 & 48.7 & 41.2 & 26.7 \\ [0.12] & 121.9 & 72.1 & 48.2 & 101.9 & 92.4 \\ [0.22] & [0.53] & 150.3 & 23.1 & 85.9 & 85.4 \\ \hline [0.42] & [0.42] & [0.18] & 106.1 & 60.5 & 56.3 \\ [0.34] & [0.85] & [0.65] & [0.54] & 117.9 & 83.8 \\ [0.19] & [0.67] & [0.56] & [0.44] & [0.62] & 156.7 \end{array}\right] \end{array}$$

Σ	AICC
Σ_{CS}	853.4
$\boldsymbol{\Sigma_{DCS}}$	**850.2**
Σ_{EP}	852.3
$\Sigma_{AR(1)}$	852.8
Σ_{UN}	872.7

Method	Outcome	LCB	Estimate	SE	p-value
No Baselines	Y_{idk}	None	5.17	2.07	0.023
CFB	$Y_{ijk} - X_{ijk}$	None	5.67	2.77	0.175
$\boldsymbol{X_A - X_B}$	Y_{idk}	$X_{iAk} - X_{iBk}$	**5.31**	**1.95**	**0.014**
AR(1)	Y_{ijk}	AR(1) covariate	5.70	1.75	0.005
Adaptive	Depends on AICC	Depends on AICC	**5.31**	**1.95**	**0.014**

Note: $\hat{\Sigma}_{UN}$ is the estimated unstructured covariance matrix of the treatment ordered responses. [] refer to correlation estimates. AICC values for the joint vector of responses under the various covariance structures are displayed. CFB=Change from Baseline. Method AR(1) refers to the covariates derived in Table 2.2 under AR(1). The adaptive method, based on the AICC values, chooses between methods No Baselines, $X_{iAk} - X_{iBk}$, and Method AR(1). CS = Compound Symmetry; DCS = Double Compound Symmetry; EP = Equipredictability; AR = Auto-regressive (1); UN=Unstructured.

2.7. Simulations

The simulation study was designed to answer the following questions: (1) Is the Type I error rate maintained, both for the two benchmarks (no baselines and CFB), for each of the LCB's under the correct covariance structure and when the covariance structure is misspecified, and for the adaptive method; (2) When the LCB is optimal for the underlying covariance structure, does including the optimal LCB offer an increase in power over the benchmarks; (3) Does the adaptive method capture any power gains seen by using the optimal LCB under the correct covariance structure; and (4) Is there any overall recommendation that can be made for practioners?

We simulated 20,000 trials for a variety of scenarios. The number of simulated data sets is rather large, but we wanted to know with high accuracy where our adaptive method suffers from type I error

inflation. The simulation scenarios were defined by the hypotheses (null, $\tau_A - \tau_B = 0$; alternative, $\tau_A - \tau_B \neq 0$), Σ (CS, DCS, EP, AR(1), UN), average pairwise correlation ($\bar{\rho}$), and sample sizes. Hypothesis tests were used with a nominal type I error rate of 0.05. Response vectors for each subject within each simulated trial were generated from a multivariate normal with a sequence-invariant Σ; either CS, DCS, EP, or AR(1) as described in Table 2.1, or UN under a treatment ordering (Equations 2.5,2.6). Throughout we assumed a common variance $\sigma^2 = 1$ and $\bar{\rho} \approx 0.60$. For all designs: under CS, $\rho = 0.6$; under EP, $\rho_2 = 0.60$, $\rho_1 = 0.70$, $\rho_3 = 0.50$. DCS, AR(1), and UN correlations vary by design (see Appendix B.2).

Under CS, DCS, EP, and UN, the response vector was generated under a treatment ordering, based on the mean models in (2.3, 2.4) with $\gamma_{id} = 0$ for all d. For 2×2, we set $\mu = 6.5$ and $E(X_{idk}) = \zeta_{id} = 0.5$ such that $E[(X_A, X_B, Y_A, Y_B)^T] = (0.5, 0.5, 6.5 + \tau_A, 6.5 + \tau_B)^T$; for 3×3, $\mu = 7$ and $E(X_{idk}) = 1$ such that $E[(X_A, X_B, X_C, Y_A, Y_B, Y_C)^T] = (1, 1, 1, 7 + \tau_A, 7 + \tau_B, 7 + \tau_C)^T$; for 4×4, $\mu = 7.38$ and $E(X_{idk}) = 1.375$ such that $E[(X_A, X_B, X_C, X_D, Y_A, Y_B, Y_C, Y_D)^T] = (1.375, 1.375, 1.375, 1.375, 7.38 + \tau_A, 7.38 + \tau_B, 7.38 + \tau_C, 7.38 + \tau_D)^T$. Under AR(1), the response vector is defined based on the period ordered mean models (2.12). See the Appendix for the specific parameter values. For 2×2, $E[(X_1, Y_1, X_2, Y_2)^T] = (0, 6 + \tau_{d[i,1]}, 1, 7 + \tau_{d[i,2]})^T$; for 3×3, $E[X_1, Y_1, X_2, Y_2, X_3, Y_3)^T] = (0, 6 + \tau_{d[i,1]}, 1, 7 + \tau_{d[i,2]}, 2, 8 + \tau_{d[i,3]})^T$; for the 4×4 design, $E[(X_1, Y_1, X_2, Y_2, X_3, Y_3, X_4, Y_4)^T] = (0, 6 + \tau_{d[i,1]}, 1, 7 + \tau_{d[i,2]}, 2, 8 + \tau_{d[i,3]}, 2.5, 8.5 + \tau_{d[i,4]})^T$. Under the null, $\tau_A = \tau_B = 0$ (and $\tau_C = \tau_D = 0$ as appropriate), while under the alternative, for each scenario, $\tau_A - \tau_B$ was fixed such that using the no baselines model yielded 80% power (see Appendix B.2). For the 3×3 and 4×4 design, $\tau_C - \tau_B$ and $\tau_D - \tau_B$ (4×4 only) were set equal to $\tau_A - \tau_B$. As expected, estimates of $\tau_A - \tau_B$ were approximately unbiased for all methods under all scenarios (results not shown). Power results are shown for the 2×2 (Table 2.5), 3×3 (Table 2.6), and 4×4 designs (Table 2.7), with indications of where type I error inflation occurs, while type I error results for all designs can be found in Appendix B.3. Lastly, while our CFB model used random subject effects (and thus assumed a CS structure), power results were comparable to when we used a CFB model with an unstructured covariance structure.

2.7.1. 2×2 *Crossover Design: Simulation Results*

The type I error was maintained across all simulations when baselines were excluded from the analysis. However, the CFB approach yielded type I error inflation for smaller sample sizes ($<$

28), particularly under AR(1). This is most likely due to how the CFB method assumes that the CFB outcomes follow a CS covariance structure. As we pointed out earlier, misspecifying the covariance structure in a mixed model can lead to type I error inflation, regardless of sample size (Gurka, Edwards, and Muller, 2011). Type I error inflation was rare when $X_A - X_B$ was included as the LCB, but including the optimal LCB under AR(1) yielded type I error inflation at smaller sample sizes. Lastly, the adaptive method exhibited type I error inflation at the lower sample sizes, especially under DCS, EP, and AR(1). As to why the adaptive method inflates the type I error, the adaptive method's success rate, or how often the AICC correctly picks the "right" covariance structure gives us some insight. For the 2×2 design, especially at small sample sizes, there is insufficient information to accurately choose the correct structure and the AICC criterion has difficulty picking the "correct" covariance structure (and thus the optimal LCB). Further, from our simulations, we found that when the incorrect covariance structure was chosen, the standard error of the treatment effect estimate was often deflated, leading to the inflated type I error.

Compared to the benchmarks (no baselines or CFB), including the optimal LCB increased power for both DCS and EP. Similarly, using no baselines under a CS structure yielded the most power and largely outperformed CFB. Under an AR(1) structure, the optimal LCB increased power at sample sizes greater than 20, but could not be evaluated at smaller sample sizes due to type I error inflation. In general, the adaptive method captured the power gains seen by using the optimal LCB, and it also matched the highest power observed using $X_A - X_B$ under an unstructured matrix. However, as mentioned above, at smaller sample sizes, the adaptive method often suffers from type I error inflation. Despite this, for N≥ 28, we see that the adaptive method does not suffer from any type I error inflation, suggesting that this method could be used for larger sample sizes. Overall, given that the adaptive method does not maintain the type I error in a variety of scenarios, our recommendation for a 2×2 design is to use $X_A - X_B$ as a covariate. This method consistently outperformed CFB and uniformly did well across all the covariance structures.

2.7.2. 3×3 Crossover Design: Simulation Results

With the exception of the adaptive method at N=18 under AR(1) and CFB under UN, all methods maintained the type I error. The lack of type I error inflation for the adaptive method can be attributed to the high success rate of the AICC criterion in the 3×3 design. For example, at N=18 under EP, the adaptive method correctly picks the optimal LCB ($X_A - X_B$) 83% of the time (out of 20,000

Table 2.5: 2×2 Simulations: Under the Alternative Hypothesis, Power

Truth	**Method**	N=12	N=16	N=20	N=24	N=28	N=32
Σ_{CS}	**No Baselines**	**79.4**	**80.2**	**79.6**	**80.1**	**80.4**	**79.6**
Σ_{CS}	CFB	[50.6]	50.6	50.8	51.2	51.1	51.1
Σ_{CS}	$X_A - X_B$	**74.4**	**76.5**	77.0	**78.5**	78.5	**78.4**
Σ_{CS}	AR(1)	74.0	76.2	76.9	78.0	78.4	78.0
Σ_{CS}	Adaptive	[[78.6]]	[79.6]	**79.4**	[80.1]	**80.4**	[79.7]
Σ_{DCS}	No Baselines	79.4	80.0	79.7	80.5	80.3	79.6
Σ_{DCS}	CFB	**86.9**	88.0	87.5	88.1	88.3	87.8
Σ_{DCS}	$\boldsymbol{X_A - X_B}$	**89.7**	**91.6**	**91.9**	**92.8**	**93.0**	**92.9**
Σ_{DCS}	AR(1)	84.9	87.3	**88.0**	88.9	89.1	88.9
Σ_{DCS}	Adaptive	[[89.6]]	**91.5**	[[91.8]]	**92.7**	**92.9**	**92.8**
Σ_{EP}	No Baselines	79.8	79.9	79.8	80.4	80.4	79.5
Σ_{EP}	CFB	79.7	79.9	[79.7]	80.2	80.1	79.9
Σ_{EP}	$\boldsymbol{X_{Ak} - X_B}$	**85.4**	**87.1**	**87.5**	**88.6**	**88.9**	**88.6**
Σ_{EP}	AR(1)	**80.4**	**82.3**	**83.0**	**84.2**	84.5	84.0
Σ_{EP}	Adaptive	[[85.6]]	[[86.8]]	[[86.9]]	[[88.1]]	**88.6**	**88.2**
Σ_{AR}	No Baselines	**79.6**	**79.7**	**79.6**	80.1	80.0	80.0
Σ_{AR}	CFB	[69.8]	[69.7]	[70.3]	70.2	70.0	69.8
Σ_{AR}	$X_A - X_B$	[79.8]	**81.8**	**82.7**	83.1	83.8	84.2
Σ_{AR}	AR(1)	[[81.7]]	[[83.5]]	[84.3]	**85.1**	**85.5**	**85.7**
Σ_{AR}	Adaptive	[[82.4]]	[[83.9]]	[[84.4]]	**85.1**	**85.5**	**85.7**
Σ_{UN}	No Baselines	**79.6**	80.2	79.8	79.6	80.0	79.9
Σ_{UN}	CFB	64.6	65.5	65.6	65.1	64.9	64.6
Σ_{UN}	$\boldsymbol{X_A - X_B}$	**80.1**	**82.1**	**82.6**	**82.9**	**83.6**	**83.6**
Σ_{UN}	AR(1)	78.4	80.8	81.4	81.7	82.5	82.6
Σ_{UN}	Adaptive	[80.4]	**82.3**	**82.6**	**82.9**	**83.6**	**83.6**

Notes: Values (Power %) are shown in bold if method yields the highest or second highest power in that sample size/covariance structure combination without type I error inflation. Entries are in brackets/double brackets if under the same scenario, but under the null hypothesis, the type I error is two/three SE's above 5% ($> 5.31\%$,$> 5.46\%$) based on 20,000 simulations. CFB=Change from Baseline. Method AR(1) refers to the covariates derived in Table 2.2 under AR(1). The adaptive method, based on AICC values, chooses between methods No Baselines, $X_A - X_B$, and Method AR(1). CS = Compound Symmetry; DCS = Double Compound Symmetry; EP = Equipredictability; AR = Auto-regressive (1); UN=Unstructured.

simulations).Next, compared to the benchmarks (no baselines and CFB), the optimal LCBs led to increased power under all scenarios. Furthermore, the adaptive method captured these power gains (superior to the benchmarks) by using the optimal LCBs and also matched the highest power observed using $X_A - X_B$ under UN. Coupled with the fact that the adaptive method did not suffer from type I error inflation, the adaptive method could be used for 3×3 crossover design.

2.7.3. 4×4 *Crossover Design: Simulations*

All methods maintained the type I error, except for CFB under UN. Like in the 2×2 design, the type I error inflation seen for CFB under UN is likely caused by the misspecification of the underlying

Table 2.6: 3×3 Simulations: Under the Alternative Hypothesis, Power

Truth	**Assume**	N=18	N=24	N=30	N=36	N=42	N=48
Σ_{CS}	**No Baselines**	**79.2**	**79.9**	**80.4**	**80.0**	**80.5**	**80.4**
Σ_{CS}	CFB	54.5	53.4	53.4	52.3	52.2	52.0
Σ_{CS}	$\boldsymbol{X_A - X_B}$	75.0	77.3	78.5	78.5	79.5	79.4
Σ_{CS}	AR(1)	74.7	77.1	78.5	78.3	79.1	79.4
Σ_{CS}	**Adaptive**	**78.8**	**79.6**	**80.2**	**79.9**	**80.3**	**80.2**
Σ_{DCS}	No Baselines	79.9	79.7	80.5	80.2	79.7	79.8
Σ_{DCS}	CFB	80.2	77.9	78.5	77.8	76.9	76.4
Σ_{DCS}	$\boldsymbol{X_A - X_B}$	**84.2**	**85.1**	**86.6**	**87.0**	**86.8**	**86.8**
Σ_{DCS}	AR(1)	79.6	81.0	82.6	83.0	82.4	82.9
Σ_{DCS}	**Adaptive**	**84.4**	**85.2**	**86.7**	**87.0**	**86.8**	**86.8**
Σ_{EP}	No Baselines	79.5	80.1	79.7	79.9	79.6	80.3
Σ_{EP}	CFB	83.6	82.5	81.5	81.0	80.7	81.2
Σ_{EP}	$\boldsymbol{X_A - X_B}$	**85.8**	**87.7**	**88.1**	**88.5**	**88.8**	**89.2**
Σ_{EP}	AR(1)	79.3	81.8	82.1	83.0	82.6	83.7
Σ_{EP}	**Adaptive**	**85.9**	**87.8**	**88.1**	**88.5**	**88.7**	**89.2**
Σ_{AR}	No Baselines	79.8	80.4	79.9	80.0	79.6	79.8
Σ_{AR}	CFB	84.6	83.6	81.9	81.4	80.8	81.5
Σ_{AR}	$\boldsymbol{X_A - X_B}$	**86.9**	88.4	88.5	88.8	88.9	89.2
Σ_{AR}	**AR(1)**	**91.2**	**92.8**	**92.8**	**93.3**	**93.3**	**93.6**
Σ_{AR}	Adaptive	[91.2]	**92.8**	**92.8**	**93.3**	**93.3**	**93.6**
Σ_{UN}	No Baselines	80.1	80.0	79.9	80.5	79.6	79.5
Σ_{UN}	CFB	67.3	65.6	64.8	65.0	63.5	63.2
Σ_{UN}	$\boldsymbol{X_A - X_B}$	**82.9**	**84.1**	**84.6**	**85.9**	**84.9**	**85.2**
Σ_{UN}	AR(1)	79.8	81.2	81.9	82.9	81.9	82.0
Σ_{UN}	**Adaptive**	**82.9**	**84.1**	**84.6**	**85.9**	**84.9**	**85.2**

Notes: Values (Power %) are shown in bold if method yields the highest or second highest power in that sample size/covariance structure combination without type I error inflation. Entries are in brackets/double brackets if under the same scenario, but under the null hypothesis, the type I error is two/three SE's above 5% ($> 5.31\%$,$> 5.46\%$) based on 20,000 simulations. CFB=Change from Baseline. Method AR(1) refers to the covariates derived in Table 2.2 under AR(1). The adaptive method, based on AICC values, chooses between methods No Baselines, $X_A - X_B$, and Method AR(1). CS = Compound Symmetry; DCS = Double Compound Symmetry; EP = Equipredictability; AR = Auto-regressive (1); UN=Unstructured.

covariance structure. Importantly, the type I error was maintained for the adaptive method in all scenarios. Like in the 3×3 design, the optimal LCBs outperformed both benchmarks (CFB and no baselines) under all scenarios. Additionally, the adaptive method did approximately as well as the optimal LCBs (and also beat out the benchmarks), while also matching the highest power observed using $X_A - X_B$ under UN. Given this, along with the fact that the adaptive method does not suffer from type I error inflation, the adaptive method could be used in the 4×4 crossover design, but with more potential for efficiency gain (relative to the 3×3 design).

Table 2.7: 4×4 Simulations: Under the Alternative Hypothesis, Power

Truth	**Method**	N=16	N=20	N=24	N=28	N=32	N=36
Σ_{CS}	**No Baselines**	**80.0**	**80.0**	**80.2**	**80.2**	**80.2**	**80.1**
Σ_{CS}	CFB	55.9	54.7	54.4	53.7	53.0	52.7
Σ_{CS}	$\boldsymbol{X_A - X_B}$	75.5	77.0	78.0	78.4	78.6	78.6
Σ_{CS}	AR(1)	75.7	77.0	78.0	78.3	78.5	78.8
Σ_{CS}	**Adaptive**	**79.4**	**79.7**	**80.0**	**80.0**	**79.9**	**79.9**
Σ_{DCS}	No Baselines	79.0	78.0	79.8	80.1	80.3	80.7
Σ_{DCS}	CFB	88.6	87.3	87.9	87.2	86.8	86.1
Σ_{DCS}	$\boldsymbol{X_A - X_B}$	**88.8**	**89.2**	**90.9**	**91.7**	**91.0**	**91.3**
Σ_{DCS}	AR(1)	82.5	81.7	84.9	85.2	85.3	85.2
Σ_{DCS}	**Adaptive**	**88.9**	**89.1**	**90.9**	**91.7**	**90.9**	**91.3**
Σ_{EP}	No Baselines	80.1	80.0	79.8	79.9	79.9	79.9
Σ_{EP}	CFB	85.5	84.2	83.3	82.8	82.3	81.8
Σ_{EP}	$\boldsymbol{X_A - X_B}$	**86.6**	**87.8**	**88.4**	**88.5**	**88.5**	**88.9**
Σ_{EP}	AR(1)	80.6	82.0	82.7	82.8	83.4	83.5
Σ_{EP}	**Adaptive**	**86.6**	**87.8**	**88.4**	**88.6**	**88.5**	**88.9**
Σ_{AR}	No Baselines	80.0	79.8	80.1	80.0	80.1	79.9
Σ_{AR}	CFB	89.8	88.5	87.9	87.6	87.2	86.7
Σ_{AR}	$\boldsymbol{X_A - X_B}$	89.8	90.6	91.1	91.6	91.6	91.6
Σ_{AR}	**AR(1)**	**95.6**	**96.2**	**96.6**	**96.7**	**96.8**	**96.8**
Σ_{AR}	**Adaptive**	**95.6**	**96.2**	**96.6**	**96.7**	**96.8**	**96.8**
Σ_{UN}	No Baselines	80.2	80.1	80.1	80.1	80.1	79.9
Σ_{UN}	CFB	[[74.9]]	[[73.9]]	[[73.1]]	[[72.1]]	[[71.7]]	[[71.2]]
Σ_{UN}	$\boldsymbol{X_A - X_B}$	**83.4**	**84.9**	**85.0**	**85.5**	**85.6**	**85.8**
Σ_{UN}	AR(1)	79.8	81.2	81.9	82.2	82.1	82.6
Σ_{UN}	**Adaptive**	**83.4**	**84.9**	**85.0**	**85.5**	**85.6**	**85.8**

Note: Values (Power %) are shown in bold if method yields the highest or second highest power in that sample size/covariance structure combination without type I error inflation. Entries are in brackets/double brackets if under the same scenario, but under the null hypothesis, the type I error is two/three SE's above 5% ($> 5.31\%$,$> 5.46\%$) based on 20,000 simulations. CFB=Change from Baseline. Method AR(1) refers to the covariates derived in Table 2.2 under AR(1). The adaptive method, based on AICC values, chooses between methods No Baselines, $X_A - X_B$, and Method AR(1). CS = Compound Symmetry; DCS = Double Compound Symmetry; EP = Equipredictability; AR = Auto-regressive (1); UN=Unstructured.

2.7.4. Simulation Results Summary

Overall, the simulations verify the theoretical arguments made with regards to baseline utilization. For example, if the underlying covariance structure is truly AR(1), then using the optimal LCB (AR(1) covariate (Table 2.2, Σ_{AR}) will best leverage the covariance structure, resulting in a large gain in efficiency relative to other methods. For the 2×2 crossover design, given the type I error inflation seen in the adaptive method, we recommend using $X_A - X_B$ as the analytical method, which is the same as previous recommendations (Mehrotra, 2014; Metcalfe, 2010). $X_A - X_B$ uniformly did well across all scenarios and also outperformed CFB in all cases.

For the 3×3 and 4×4 crossover designs, the adaptive method did as well as the best performing method in terms of power (optimal LCB), while controlling the nominal type I error rate. This is due to the AICC criteria better differentiating between covariance structures as the number of treatments (or periods) increases. The adaptive method also significantly outperformed the benchmark CFB. Given the simulation results for the 3×3 and 4×4 crossover design, we recommend using the adaptive method.

2.8. Discussion

By incorporating LCBs as covariates, we have developed an approach that uses the information in the underlying covariance structure of the baselines and outcomes to yield a more precise estimate of a desired treatment effect. Importantly, our proposed approach is less restrictive than most models (e.g., CFB with random subject effects), given that we make no assumptions about the covariance structure of the outcomes in the estimation of $\tau_A - \tau_B$. This approach is easily implemented, as one must simply add the LCB as a covariate in the regression model. Furthermore, while we derived optimal LCBs under CS, EP, DCS, and AR(1), this approach could be extended to any covariance structure one may deem plausible. The trick is simply deriving the appropriate conditional variance of a treatment effect, then determining the optimal LCB that minimizes this variance. In some cases, there may be an analytic solution, in other cases one may need to use an optimization algorithm. From our experience, using an optimization algorithm to determine the optimal LCB yielded similar if not identical results compared to using an analytically derived optimal LCB.

This work does confirm previous results showing limited gains in efficiency for the AB/BA design. Under a CS covariance structure, each pair of measurements has the same covariance, irrespective of the temporal separation between the measurements, or the treatment administered. Under a CS assumption, Yan showed that using baselines has a minimal effect on the variance of the treatment estimate for the AB/BA design (Yan, 2012). Mehrotra confirmed this finding, but also showed several data examples where CS appeared to be an oversimplification of the underlying structure (Mehrotra, 2014). Kenward and Roger similarly present a number of examples where a more flexible covariance structure appears better suited to the data (Kenward and Roger, 2010).

Importantly our proposed methods outperform commonly used models, specifically CFB or omitting

the baselines from the analysis. CFB is both intuitive and widespread, but is known to underperform, even compared to omitting the baselines altogether (Kenward and Roger, 2010; Mehrotra, 2014). Kenward and Roger recommend against the CFB approach and instead advocated conditioning on baselines as covariates. In fact, one example given showed an efficiency gain equivalent to a 40% decrease in sample size by using baselines as covariates (Kenward and Roger, 2010).

It may also be tempting to apply these methods to a joint model framework, for example using $((Y_{iAk} - Y_{iBk}, \boldsymbol{a}_\star^T \boldsymbol{X}_{ik})^T)$ as the outcome, as in Chen, Meng, and Zhang, 2012. However in a joint model, the test statistic is obtained using a linear mixed model, and the restricted maximum likelihood results are exact only asymptotically. One solution to this is to use the Kenward-Roger degrees of freedom (DDFM=KR in SAS PROC MIXED) which adjusts the standard error of the treatment effect estimate by accounting for the uncertainty in the estimate of the covariance structure. This has been shown to deliver a more accurate t-test in small samples (Kenward and Roger, 1997). However, even with this adjustment, Mehrotra illustrated through simulations that the type I error rate for testing a treatment effect could still be inflated (Mehrotra, 2014). Given all this, we recommend using an OLS based approach.

The adaptive method relies on the use of AICC criterion to determine what LCB to include in the analysis. From our simulations, the adaptive method's success rate (how often the AICC criterion chose the correct covariance structure) depended on the design. For 2×2 crossover trials, the adaptive method visibily exhibited type I error inflation. This was largely due to the AICC criterion having difficultly discerning between different covariance structures, especially at the smaller sample sizes. This was not an issue for the 3×3 and 4×4 crossover designs, where the adaptive method showed promising properties. In general, it is easier to differentiate between the covariance structures considered as the number of treatments/periods increases.

An attractive alternative to the adaptive method is a fully robust method of choosing the LCB. The idea here would be to minimize the conditional variance (2.11) while assuming an unstructured covariance matrix. This makes no assumptions about the covariance matrix of the baselines/outcomes which means there is no "choosing" between baseline utilization methods (which the adaptive method does). However, at low sample sizes, the variance component estimates necessary to determine the optimal LCB are unstable. Additionally, this approach requires some type of resampling method (i.e. bootstrap, permutation test) to maintain the nominal type I error rate.

This idea is pursued further in Chapter 3 and 4.

Overall, given the underlying covariance structure of the baselines and outcomes, we can construct LCBs that minimize the conditional variance of the within-subject contrasts corresponding to a desired treatment effect. Inclusion of these LCBs into a regression model can increase the efficiency of a treatment effect estimate. Compared to standard methods, such as CFB, using LCBs as covariates can yield large gains in power in 2×2, 3×3, and 4×4 crossover designs.

CHAPTER 3

INCOMPLETE BLOCK DESIGNS

3.1. Motivation and Literature Review

In an incomplete block design, the number of treatments exceeds the number of periods, meaning that this design is not uniform within sequence (UWS) and thus not uniform. As a consequence, estimation of a treatment effect is not necessarily obtained entirely through within-subject contrasts (Chi, 1991). Recall that in the Chapter 2 (uniform designs), all estimation was through within-subject contrasts. Given this, relative to a uniform design with the same number of treatments, an incomplete block design will yield a less efficient estimate of a treatment effect. Despite this, an incomplete block design can be appealing, given that each subject in the study receives fewer treatments, reducing the risk of incomplete data. (Jones and Kenward, 2003). For example, if the goal was to compare three treatments (A, B, or C), a uniform design could include all six possible sequences of the treatments (ABC, BAC, CAB, CBA, ACB, BCA), requiring each subject to receive three treatments administered across three periods. An analagous 3-treatment 2-period (3×2) incomplete block design would also have six sequences (AB, BA, BC, CB, AC, CA). But for this design, each subject would receive two out of the three treatments administered across two periods. As the number of treatments increases, a uniform design can become too arduous for a subject to complete. Thus, an incomplete block design may have the advantage of maintaining a high level of completion among subjects.

In Chapter 2, we showed that for a general uniform crossover design, an optimal linear combination of baselines (LCB) could be used as a covariate to efficiently incorporate all available baselines. The optimal LCB minimizes the conditional variance of a treatment effect contrast, leveraging the underlying joint covariance structure of the outcomes and baselines. Furthermore, by using a single metric to incorporate all the available baselines, only one degree of freedom is sacrificed in a hypothesis test. For small sample sizes, this can greatly increase the power of the hypothesis test. To our knowledge, there are no specific publications related to the use of baselines in an incomplete block crossover design. Notably, the work in Chapter 2 rested on the premise that all estimation was done through within-subject contrasts (Jemielita, Putt, and Mehrotra, 2016). However, for an

incomplete block design, estimation may also include between-subject contrasts.

For an incomplete block design, Chi illustrated that the overall estimate of a treatment effect is a weighted combination of within-subject and between-subject information (Chi, 1991). Moreover, the overall estimate of a treatment effect can be obtained through generalized least squares (GLS), efficiently combining the within-subject and the between-subject information (Chi, 1991). Kenward and Jones also pointed this out, noting that an approximate estimate of the overall treatment effect could be obtaining through inverse variance weighting of the between-subject estimate of the treatment effect and the within-subject estimate of the treatment effect (Jones and Kenward, 2003). To illustrate, consider the 3×2 crossover design, with sequences AB, BA, BC, CB, AC, CA. With the goal of estimating the treatment effect comparing A and B, note that the within-subject estimate is obtained through taking appropriate within-subject contrasts in sequences AB and BA. Effectively, these two sequences could be viewed together as a 2×2 or AB/BA design. The between-subject estimate of the treatment effect is obtained through a summation of period-specific outcomes in the remaining sequences. The overall estimate of the treatment effect is then calculated by using inverse variance weights on the within-subject and between-subject portions.

While this approximate estimate is conceptually insightful, in practice the overall estimate of a treatment effect is usually obtained by fitting a mixed effects model, where estimation is done through GLS and restricted maximum likelihood (REML). To account for the uncertainty of the estimated covariance structure, a Kenward-Roger degrees of freedom adjustment should be used. Especially at small sample sizes, this delivers more accurate inference (Kenward and Roger, 1997). This is particularly salient for crossover designs where sample sizes are small. However, even with this adjustment, problems may persist for small sample sizes. For a repeated measures design with small sample sizes and a complex covariance structure, Schaalje et al showed that hypothesis tests with a KR adjustment still resulted in inflated type I error (Schaalje, Mcbride, and Fellingham, 2002). Additionally, Mehrotra showed that REML based models for a 2×2 crossover design did not maintain the nominal type I error rate at small sample sizes (Mehrotra, 2014).

In this Chapter, potential LCB models are explored for incomplete block designs, with an emphasis on the 3×2 design. Importantly, we explore baseline utilization in the framework of a mixed model, where the overall treatment estimate is a weighted combination of the between-subject and within-subject information, and also in the framework of an OLS model, where the treatment effect is

estimated using only within-subject contrasts. Under either case, we explicitly incorporate LCBs into the analysis such that the precison of the treatment effect estimate is increased. We hypothesize that the OLS method might be particularly useful for small sample sizes.

Setup and notation for an incomplete block crossover design are defined in Section 3.2. In Section 3.3, we describe various baseline models for the incomplete block design. These models are developed under a mixed effects model framework and an OLS framework. Section covers estimation for the proposed methods. In Section 3.2, the baseline models are evaluated on a real 3×2 data set. In Section 3.6, all models are compared through simulations for the 3×2 design. Lastly, in section 3.7, the overall findings are summarized.

3.2. Setup and Notation

Consider an incomplete block crossover design. Briefly, it is more convenient for incomplete block designs to consider the baselines and outcomes ordered by period (as in Section 2.3.2). Accordingly, let:

$$\begin{aligned} \mathbf{X}_{ik} &= (X_{i1k}, ..., X_{ijk}, ..., X_{ipk})^T \\ \mathbf{Y}_{ik} &= (Y_{i1k}, ..., Y_{ijk}, ..., Y_{ipk})^T \end{aligned} \tag{3.1}$$

be the vectors of baseline and outcome measurements respectively, where $i = 1, ..., s$ indexes sequence, $j = 1, ..., p$ indexes period, and $k = 1, ..., n_i$ indexes subject k in sequence i where subjects are assumed to be independent of each other. We then assume that:

$$\begin{pmatrix} \mathbf{X}_{ik} \\ \mathbf{Y}_{ik} \end{pmatrix} \sim N\left(\begin{pmatrix} E(\mathbf{X}_{ik}) \\ E(\mathbf{Y}_{ik}) \end{pmatrix}, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{XY}^T & \Sigma_{YY} \end{pmatrix} \right) \tag{3.2}$$

Note that the covariance matrices are sequence-invariant. The expectations under a no carryover assumption are defined by their individual elements:

$$E(Y_{ijk}) = \mu + \pi_j + \tau_{d[i,j]} \tag{3.3}$$

$$E(X_{ijk}) = \zeta_j \tag{3.4}$$

where μ is the overall mean, π_j is the effect of period j with $\sum_j \pi_j = 0$, $\tau_{d[i,j]}$ is the effect of treatment $d = A, B, .., Z$ (defined by the period j and sequence i) with $\sum_d \tau_d = 0$, and ζ_j represents the mean of a baseline in period j. For an incomplete block design, the number of treatments exceeds the number of periods such that $Z > p$. Note that (3.4) is sequence-invariant, but allows for period-effects. In the absence of carryover, it may be reasonable to assume that $\mu + \pi_j = \zeta_j$ which implies that $E(Y_{ijk} - X_{ijk}) = \tau_{d[i,j]}$. However, this restriction is not necessary for our methods.

Our methods largely revolve around using the underlying joint covariance of the baselines and outcomes. While Equation (3.2) shows the joint covariance of the Xs and Ys in block matrices, it is more natural to think of the joint covariance structure in terms of the baselines and outcomes ordered temporally, or in the order in which they are measured. Let the baselines and outcomes ordered temporally be denoted as $\mathbf{W}_{ik} = (X_{i1k}, Y_{i1k}, ..., X_{ijk}, Y_{ijk}, ..., X_{ipk}, Y_{ipk})^T$. Then, from (3.2), it follows that:

$$\mathbf{W}_{ik} \sim N\Big(E(\mathbf{W}_{ik}), V(\mathbf{W}_{ik})\Big) \tag{3.5}$$

where $V(\mathbf{W}_{ik})$ is just a re-ordering of the elements in the sub-matrices defined by (3.2), while $E(\mathbf{W}_{ik})$ is defined by (3.3) and (3.4).

Next, while our baseline models are developed under an unstructured (UN) covariance assumption, for simulations and comparison of variance formulas, three plausible covariance structures are considered: Compound symmetry (CS), Equipredictability (EP), and Autoregressive (1) (AR(1)). These structures were discussed in detail in Section 2.5.2 for a general uniform crossover design, but can apply to any setting in which there are repeated measurements (Jemielita, Putt, and Mehrotra, 2016). Notably, while CS and EP were described under a treatment ordered setting in Section 2.5.2, these structures are easily described in the setting where baselines and outcomes are ordered temporally. Compound Symmetry assumes a common variance and common correlation for all measurements; $V(Y_j) = V(X_j) = \sigma$, corr$(Y_j, Y_{j'}) =$ corr$(X_j, X_{j'}) =$ corr$(Y_j, X_{j'}) = \rho$ for all j, j'. Equipredictability extends Compound Symmetry, allowing baselines and outcomes between and within periods to have a separate correlations; corr$(Y_j, X_j) = \rho_1$ for all $j = j$, corr$(Y_j, Y_{j'}) =$ corr$(X_j, X_{j'}) =$ corr$(Y_j, X_{j'}) = \rho_2$ for all j, j', and corr$(Y_j, X_{j'}) = \rho_3$ for all $j \neq j'$. Autoregressive(1) is a two-parameter covariance structure, in which measurements that more

distant temporally are less correlated. For the 3×2 design under these covariance structures, $V(\mathbf{W}_{ik}) = V((X_{i1k}, Y_{i1k}, X_{i2k}, Y_{i2k})^T)$ is written as:

$$\Sigma_{CS} = \sigma \begin{bmatrix} 1 & \rho & \rho & \rho \\ & 1 & \rho & \rho \\ & & 1 & \rho \\ & & & 1 \end{bmatrix} \Sigma_{AR(1)} = \sigma \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ & 1 & \rho & \rho^2 \\ & & 1 & \rho \\ & & & 1 \end{bmatrix} \tag{3.6}$$

$$\Sigma_{EP} = \sigma \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ & 1 & \rho_3 & \rho_2 \\ & & 1 & \rho_1 \\ & & & 1 \end{bmatrix} \Sigma_{UN} = \begin{bmatrix} \sigma_1 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ & \sigma_2 & \sigma_{23} & \sigma_{24} \\ & & \sigma_3 & \sigma_{34} \\ & & & \sigma_4 \end{bmatrix}$$

We examine the performance of our methods under these covariance structures through 3×2 crossover design simulations in Section 6.

Lastly, let $\mathbf{a}^T = (a_1, ..., a_p)$ be a p-length vector of constants. Then $\mathbf{a}^T\mathbf{X}_{ik} = \sum_j^p a_j X_{ijk}$ is a linear combination of baselines (LCB). Furthermore, an LCB could also be sequence-specific ($\mathbf{a}_i^T$), discussed in Section 2.3.2, or period-specific ($\mathbf{a}_j^T$). Period-specific LCBs are discussed in Section 3.3.1. Regardless, given (3.2), the following conditional distribution is of interest:

$$\mathbf{Y}_{ik}|\mathbf{a}^T\mathbf{X}_{ik} \sim N\Big(E(\mathbf{Y}_{ik}) - \Sigma_{XY}^T\mathbf{a}(\mathbf{a}^T\Sigma_{XX}\mathbf{a})^{-1}(\mathbf{a}^T E(\mathbf{X}_{ik}) - \mathbf{a}^T\mathbf{X}_{ik}), \\ \Sigma_{YY} - \Sigma_{XY}^T\mathbf{a}(\mathbf{a}^T\Sigma_{XX}\mathbf{a})^{-1}\mathbf{a}^T\Sigma_{XY}\Big) \tag{3.7}$$

In general, LCBs are chosen based on a minimization criteria that is related to the conditional variance of a desired treatment effect.

3.3. Baseline Models

As discussed in the introduction, two modeling frameworks for baseline utilization are explored. The first modeling option is a mixed effects model. Under this option, the overall estimate of a treatment effect is a weighted combination of the within-subject information and the between-subject information (Chi, 1991). GLS, with REML to estimate the necessary variance components, is used to efficiently combine the within-subject and between-subject information. Further, a Kenward-Roger

degrees of freedom adjustment should be used to obtain valid inference. The second modeling option for estimation only includes within-subject contrasts and thus OLS can be used. Under each modeling framework, various baseline strategies are developed. The primary goal is to identify a baseline utilization method that consistently outperforms common analysis methods, specifically the change from baseline (CFB) model (3.9).

3.3.1. Mixed Models: Baseline Utilization

1. **Mixed Model: No Baselines, $\mathbf{Y}$**

$$Y_{ijk} = \mu + \tau_{d[i,j]} + \pi_j + \epsilon_{ijk} \tag{3.8}$$

   In this model, no baseline covariate is included and the outcome is the dependent variable. Notably, the overall estimate of some treatment effect obtained from a mixed effects model (and thus GLS) can be viewed as a weighted summation of period-specific outcomes. This is the case for any of the proposed mixed models and is illustrated in Appendix C.1 for the 3×2 design.

2. **Mixed Model: Change from Baseline (CFB), $\mathbf{Y} - \mathbf{X}$**

$$(Y_{ijk} - X_{ijk}) = \mu + \tau_{d[i,j]} + \pi_j + \epsilon_{ijk} \tag{3.9}$$

   The CFB model is frequently used in crossover designs. While it remains a popular and intuitive model, there is no guarantee that this model will outperform the simpler model without baselines (Kenward and Roger, 2010; Mehrotra, 2014).

3. **Mixed Model: $\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$**

$$Y_{ijk} = \mu + \tau_{d[i,j]} + \pi_j + \beta_j X_{ijk} + \gamma(\sum_{j}^{p} X_{ijk}) + \epsilon_{ijk} \tag{3.10}$$

   In this model, period-specific baselines are regresssed against the corresponding period-specific outcome. The baseline regression parameters (β_j,γ) depend on the joint covariance of the baselines and outcomes. See Appendix for details and the exact formulas for β_j, γ. The term $(\sum_j^p X_{ijk})$ is added to remove any bias in the estimate of a pairwise treatment effect.

Kenward and Roger showed that under designs where the overall estimate of a treatment effect is a combination of between-subject and within-subject information (i.e. incomplete block designs), fitting a mixed model with period-specific covariates can add bias to the treatment effect estimates (Kenward and Roger, 2010). For example, consider the 3×2 design and say we fit a model using only within-subject information (sequences AB,BA) and a model using only between-subject information (sequences BC,CB,AC,CA). In this case, there is no guarantee that the two model's baseline regression coefficients are equal and fitting (3.10) without the summation term implicitly assumes that the between-subject and within-subject regression coefficients are equivalent. Specifically, fitting (3.10) without the summation term yields unbiased estimates only when the between-subject and within-subject regression coefficients of X_{ijk} are equivalent. By adding the summation term, the between-subject and within-subject baseline regression coefficients are allowed to differ, thus removing the potential bias.

4. **Mixed Model:** $\mathbf{Y}|\mathbf{X}\mathbf{a}_\star, \overline{\mathbf{X}\mathbf{a}}_\star$

$$Y_{ijk} = \mu + \tau_{d[i,j]} + \pi_j + \beta_j \mathbf{a}_{j\star}^T \mathbf{X}_{ik} + \gamma(\sum_j^p \mathbf{a}_{j\star}^T \mathbf{X}_{ik}) + \epsilon_{ijk} \tag{3.11}$$

where β_j and γ depend on the covariance of the outcomes and baselines (Appendix) and:

$$\mathbf{a}_{j\star}^T = (a_{j1}^\star, ..., a_{jp}^\star)^T = \underset{\boldsymbol{a}^T}{\text{Argmin}} V(Y_{ijk}|\mathbf{a}^T\mathbf{X}_{ik}) \tag{3.12}$$

Given (3.7), it follows that:

$$V(Y_{ijk}|\mathbf{a}^T\mathbf{X}_{ik}) = V(Y_{ijk}) - \frac{\text{cov}(Y_{ijk}, \mathbf{a}^T\mathbf{X}_{ik})^2}{V(\mathbf{a}^T\mathbf{X}_{ik})} \tag{3.13}$$

In general, (3.12) can be solved using an optimization algorithm such a Newton-Raphson (proc NLP in SAS or optim in R). For the 3×2 design, the solutions to (3.12) for $j = 1, 2$, have analytical solutions under an unstructured covariance structure (3.6):

$$\mathbf{a}_{1\star}^T = \left\{ a_{11}^\star = \sigma_3\sigma_{12} - \sigma_{23}\sigma_{13}; a_{12}^\star = \sigma_1\sigma_{23} - \sigma_{13}\sigma_{12} \right\} \tag{3.14}$$

$$\mathbf{a}_{2\star}^T = \left\{ a_{21}^\star = \sigma_3\sigma_{14} - \sigma_{13}\sigma_{34}; a_{22}^\star = \sigma_1\sigma_{34} - \sigma_{13}\sigma_{14} \right\} \tag{3.15}$$

Importantly, regardless of the design, certain covariance parameters must be estimated to estimate the period-specific LCBs. The intuition behind this model comes from the observation that the overall estimate of any pairwise treatment difference is a weighted combination of period-specific outcomes. Thus, the idea here is to reduce the overall variance of some treatment effect estimate by reducing the variance of each individual period-specific conditional outcome. As in (3.10), the term $(\sum_j^p \mathbf{a}_{j\star}^T \mathbf{X}_{ik})$ is added to remove any bias in an estimate of a pairwise treatment effect.

3.3.2. Within-Subject Contrast Models: Baseline Utilization

The previous subsection dealt with baseline utilization in mixed models. Under a mixed model framework, the estimate of a treatment effect can be viewed as a weighted summation of period-specific outcomes. This motivated using period-specific LCBs as covariates. On the other hand, the within-subject (WS) contrast model approach turns the longitudinal model into a cross-sectional model where each subject effectively has a single derived outcome. Under this setting, incorporating baselines in an efficient way is similar to the approach taken in Chapter 2.

For a general incomplete block design, the first step is to find a set of p-length vectors $\mathbf{b} = \{\mathbf{b}_1, ..., \mathbf{b}_i, ..., \mathbf{b}_s\}$ such that for some constant K:

$$\frac{1}{K} E\left(\sum_i^s \frac{1}{n_i} \sum_k^{n_i} \mathbf{b}_i \mathbf{Y}_{ik}\right) = \tau_A - \tau_B \tag{3.16}$$

For example, if $p = 2$ and $\mathbf{b}_i = (1, -1)$, then $\mathbf{b}_i \mathbf{Y}_{ik} = Y_{i1k} - Y_{i2k}$. Notably, based on the expectation of the outcomes (4.2), the contrasts $\mathbf{b}_i \mathbf{Y}_{ik}$ are chosen to eliminate period-effect parameters (π_j). This is straightforward for an incomplete block design that is uniform within period, like the 3×2 design. For an incomplete block design that is not uniform within period, the WS contrast model may not be appropriate. For this work, these designs are not considered. Regardless, to incorporate baselines, consider an approach which incorporates optimal LCBs ($\mathbf{a}_{i\star}^T \mathbf{X}_{ik}$) for each sequence. Thus:

$$\mathbf{a}_{i\star}^T = (a_{i1}^\star, ..., a_{ip}^\star)^T = \underset{\boldsymbol{a}^T}{\text{Argmin}} V(\mathbf{b}_i \mathbf{Y}_{ik} | \mathbf{a}^T \mathbf{X}_{ik}) \tag{3.17}$$

In general, this can be solved through numerical optimization. For the 3×2 design, there are exact

analytical solutions. Further, the optimal LCB is the same for sequences with the same within-subject contrasts. Next, while (3.16) and (3.17) show the fundamental strategy of a WS model, we illustrate how to implement this approach for the 3×2 design.

- 3×2 **Design, WS Model**: Given the expectations from (4.2), an estimate of $\tau_A - \tau_B$ can be wholly obtained through within-subject contrasts. Note that:

$$\begin{aligned}
E(Y_{AB,1} - Y_{AB,2}) &= (\pi_1 - \pi_2) + (\tau_A - \tau_B) \\
E(Y_{BA,1} - Y_{BA,2}) &= (\pi_1 - \pi_2) - (\tau_A - \tau_B) \\
E(Y_{BC,1} - Y_{BC,2}) &= (\pi_1 - \pi_2) + (\tau_B - \tau_C) \\
E(Y_{CB,1} - Y_{CB,2}) &= (\pi_1 - \pi_2) - (\tau_B - \tau_C) \\
E(Y_{AC,1} - Y_{AC,2}) &= (\pi_1 - \pi_2) + (\tau_A - \tau_C) \\
E(Y_{CA,1} - Y_{CA,2}) &= (\pi_1 - \pi_2) - (\tau_A - \tau_C)
\end{aligned}$$

Relating this back to our general WS model formulation (3.16), it follows that:

$$\begin{aligned}
\frac{1}{4}E\left(\sum_i^s \frac{1}{n_i}\sum_k^{n_i} \mathbf{b}_i \mathbf{Y}_{ik}\right) &= \frac{1}{4}E\left(\sum_i^{AB,CB,AC} \frac{1}{n_i}\sum_k^{n_i}(1,-1)\mathbf{Y}_{ik} - \sum_i^{BA,BC,CA} \frac{1}{n_i}\sum_k^{n_i}(1,-1)\mathbf{Y}_{ik}\right) \\
&= \frac{1}{4}\left(3(\pi_1 - \pi_2) + 2(\tau_A - \tau_B) + 2(\tau_C - \tau_B) + 2(\tau_A - \tau_C) + 3(\pi_2 - \pi_1)\right) = \tau_A - \tau_B
\end{aligned}$$

Note that only $Y_{i1k} - Y_{i2k}$ contrasts are used, meaning there is a single optimal LCB for all sequences. Given this, consider the following WS regression model:

$$Y_{i1k} - Y_{i2k} = (\pi_1 - \pi_2) + \tau_{AB} * S_{1,i} + \tau_{BC} * S_{2,i} + \tau_{AC} * S_{3,i} + \beta \mathbf{a}^T \mathbf{X}_{ik} + \epsilon_{ik} \qquad (3.18)$$

where $\tau_{AB} = \tau_A - \tau_B$, $S_{1,i} = 1, -1$ for $i = AB, BA$ and zero otherwise, $\tau_{BC} = \tau_B - \tau_C$, $S_{2,i} = 1, -1$ for $i = BC, CB$ and zero otherwise, $\tau_{AC} = \tau_A - \tau_C$, and $S_{3,i} = 1, -1$ for $i = AC, CA$ and zero otherwise. To incorporate the full data to compare treatment A to treatment B, it follows that: $\tau_A - \tau_B = \frac{1}{2}(\tau_{AB} + \tau_{AC} - \tau_{BC})$. Next, this model can be further improved by noting that $\tau_{BC} = \tau_{AC} - \tau_{AB}$. Given this constraint, our model can be re-parameterized as:

$$\begin{aligned}
Y_{i1k} - Y_{i2k} &= (\pi_1 - \pi_2) + \tau_{AB} * S_{1,i} + (\tau_{AC} - \tau_{AB}) * S_{2,i} + \tau_{AC} * S_{3,i} + \beta \mathbf{a}^T \mathbf{X}_{ik} + \epsilon_{ik} \\
&= (\pi_1 - \pi_2) + \tau_{AB} * (S_{1,i} - S_{2,i}) + \tau_{AC} * (S_{2,i} + S_{3,i}) + \beta \mathbf{a}^T \mathbf{X}_{ik} + \epsilon_{ik} \qquad (3.19)
\end{aligned}$$

This modification reduces the number of regression parameters and degrees of freedom by one, thus increasing the overall efficiency of a hypothesis test. SAS code for this model is given at the end of the Section 3.4. An optimal LCB ($\mathbf{a}_\star^T\mathbf{X}_{ik}$) can be found by solving:

$$\mathbf{a}_\star^T = (a_1^\star, a_2^\star)^T = \underset{\boldsymbol{a}^T}{\text{Argmin}} V(Y_{i1k} - Y_{i2k}|\mathbf{a}^T\mathbf{X}_{ik}) \tag{3.20}$$

$$V(Y_{i1k} - Y_{i2k}|\mathbf{a}^T\mathbf{X}_{ik}) = V(Y_{i1k} - Y_{i2k}) - \frac{\text{cov}(Y_{i1k} - Y_{i2k}, \mathbf{a}^T\mathbf{X}_{ik})^2}{V(\mathbf{a}^T\mathbf{X}_{ik})} \tag{3.21}$$

. Equation (3.21) follows directly from (3.7). The analytical solution to (3.20) under an unstructured covariance (3.6) is:

$$\mathbf{a}_\star^T = \left\{a_1^\star = \sigma_3(\sigma_{12} - \sigma_{14}) - \sigma_{13}(\sigma_{23} - \sigma_{34}); a_2^\star = \sigma_1(\sigma_{23} - \sigma_{34}) - \sigma_{13}(\sigma_{12} - \sigma_{14})\right\}$$

- **WS Baseline Models**

The WS regression model estimates a pairwise treatment effect using only within-subject contrasts. Here, each subject has a single outcome and OLS can be used for estimation. Thus, efficient baseline utilization is identical to the uniform design (Jemielita, Putt, and Mehrotra, 2016). Recall that $\mathbf{b}_i\mathbf{Y}_{ik}$ refers to the within-subject contrast used in sequence i. Consider the following baseline models:

1. **WS Model: No Baselines**; This model does not include a baseline covariate ($\mathbf{a}_\star^T\mathbf{X}_{ik} = 0$). Under a WS framework, this model is optimal under CS (Jemielita, Putt, and Mehrotra, 2016).

2. **WS Model**: $\mathbf{b}_i\mathbf{X}_{ik}$; Let $\mathbf{a}_\star^T\mathbf{X}_{ik} = \mathbf{b}_i\mathbf{X}_{ik}$ such that the LCB used corresponds to the within-subject contrast. For the 3×2 design, since the outcome contrasts are between period 1 and period 2, $\mathbf{a}_\star^T\mathbf{X}_{ik} = X_{i1k} - X_{i2k}$. This follows the recommendation of previous research to use the difference between the baselines at period 1 and period 2 as a covariate for a 2×2 crossover design (Jemielita, Putt, and Mehrotra, 2016; Mehrotra, 2014; Metcalfe, 2010).

3. **WS Model: LCB**; Estimate sequence specific optimal LCBs ($\mathbf{a}_{i\star}^T\mathbf{X}_{ik}$) where $\mathbf{a}_{i\star}^T$ (3.17) is solved under an unstructured covariance structure. For the 3×2 design, there is

a single optimal LCB ($\mathbf{a}_\star^T\mathbf{X}_{ik}$). We previously found little improvement in efficiency in 2-period designs by estimating the optimal LCB (Jemielita, Putt, and Mehrotra, 2016). Our simulations confirmed this observation for the 3×2 design and this model is not considered further.

- **WS Model: Adaptive**

  Under the WS model framework, there are certain conditions where it is not efficient to include baselines (or LCBs) as covariates. In general, it is inefficient to include LCBs if cov($\mathbf{b}_i\mathbf{Y}_{ik}, \mathbf{a}_i^T\mathbf{X}_{ik}) = 0$ for all i and $\mathbf{a}_i^T$. For the 3×2 design, if cov$(Y_{i1k} - Y_{i2k}, \mathbf{a}^T\mathbf{X}_{ik}) = 0$ for all $\mathbf{a}^T$, then including an LCB will not decrease the variance of a treatment effect estimate (3.21). As an example, this occurs when the underlying covariance structure is CS. This was discussed by Mehrotra and Yan for the 2×2 design (Mehrotra, 2014; Yan, 2012) and in Section 2.5.2 for the general uniform design. Indeed, when the WS model includes an LCB covariate and the true covariance is CS, the variance is not reduced and the hypothesis test for a treatment effect uses up an extra unnecessary degree of freedom. For small sample sizes, compared to a model without a baseline covariate, this results in reduced power. Subsequently, as in Section 2.5.3, information criteria are used to guide an adaptive procedure. Our procedure is as follows:

  1. Fit four joint models of the temporally ordered baselines and outcomes $(X_{i1k}, Y_{i1k}$,..., $X_{ipk}, Y_{ipk})^T$, where the assumed underlying covariance structure is Σ_{CS}, Σ_{EP}, Σ_{AR}, or Σ_{UN} (3.6). All models use a saturated means model for the combined vector of temporally ordered measurements.

  2. For each model, estimate the AICC, a small-sample corrected version of the AIC (Hurvich and Tsai, 1989).

  3. If the AICC favors Σ_{CS}, do not include a baseline covariate in the WS Model ($\mathbf{a}^T\mathbf{X}_{ik} = 0$). Else, use **WS Model: $\mathbf{b}_i\mathbf{X}_{ik}$** ($X_1 - X_2$ in the 3×2 design).

All considered methods are summarized below in Table 3.1. Since our real data example and simulations are for the 3×2 design, the methods summary table is specifically for the 3×2 design. Additional insights on the estimation are also included, which are described in Section 3.4.

Importantly, for the mixed model with period-specific LCBs ($\mathbf{Y}|\mathbf{X}\mathbf{a}_\star, \overline{\mathbf{X}\mathbf{a}}_\star$), unstructured covariance estimates obtained from Step (1) in the adaptive procedure are used to construct our period-specific LCBs.

Table 3.1: Baseline Utilization Methods: Incomplete Block Design

Framework	Method	Covariate(s)	Estimation	Covariance
Mixed	No Baselines	None	GLS, REML, KR DF	Σ_{UN}
Mixed	CFB	None	GLS, REML, KR DF	Σ_{UN}
Mixed	$\mathbf{Y}\|\mathbf{X}, \overline{\mathbf{X}}$	Period-specific: X_{ijk}, $\sum_j X_{ijk}$	GLS, REML, KR DF	Σ_{UN}
Mixed	$\mathbf{Y}\|\mathbf{X}\mathbf{a}_\star, \overline{\mathbf{X}\mathbf{a}}_\star$	Period-Specific LCBs: $\mathbf{a}_{j\star}^T\mathbf{X}_{ik}$, $\sum_j \mathbf{a}_{j\star}^T\mathbf{X}_{ik}$	GLS, REML, KR DF	Σ_{UN}
WS	No Baselines	None	OLS, exact test	Common σ^2
WS	$X_1 - X_2$	Single LCB: $X_{i1k} - X_{i2k}$	OLS, exact test	Common σ^2
WS	Adaptive	IF AICC favors CS, No Baselines. Else $X_1 - X_2$.	OLS, exact test	Common σ^2

Notes: WS: Within-Subject. See Section 3.3 for detailed explanation on models. Estimation refers to estimation methods used (see Section 3.4). KR DF=Kenward Roger Degrees of freedom; Covariance refers to what is the assumed covariance structure during estimation.

3.3.3. 3×2 *Design: Comparison of Models*

Briefly, variance formulas for the treatment effect estimate are compared for the proposed models (Table 3.1). Derivations of these variance formulas can be found in the Appendix C. For a balanced design ($n = n_i$ for all i), the variance of the estimated treatment effect for any of the mixed models can be placed in the following form:

$$V(\widehat{\tau_A - \tau_B})_M = \frac{1}{n}\frac{V(Y_{i1k}^\star)V(Y_{i2k}^\star) - \text{cov}(Y_{i1k}^\star, Y_{i2k}^\star)^2}{V(Y_{i1k}^\star) + V(Y_{i2k}^\star) + \text{cov}(Y_{i1k}^\star, Y_{i2k}^\star)} \tag{3.22}$$

where $Y_{ijk}^\star$ depends on the chosen mixed model. For the No Baselines mixed model, $Y_{ijk}^\star = Y_{ijk}$; for CFB, $Y_{ijk}^\star = Y_{ijk} - X_{ijk}$; for the period-specific LCB model, $Y_{ijk}^\star = Y_{ijk} - \beta_j(\mathbf{a}_{j\star}\mathbf{X}_{ik}) - \gamma(\mathbf{a}_{1\star}\mathbf{X}_{ik} + \mathbf{a}_{2\star}\mathbf{X}_{ik})$. Note that $\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$ is corresponds to a period-specific LCB model with $\mathbf{a}_{1\star} = (1, 0)$ and $\mathbf{a}_{2\star} = (0, 1)$. For any of the WS models, the variance of interest is:

$$V(\widehat{\tau_A - \tau_B})_{WS} = \frac{1}{3n}\left(V(Y_{i1k} - Y_{i2k}) - \frac{\text{cov}(Y_{i1k} - Y_{i2k}, \mathbf{a}^T\mathbf{X}_{ik})^2}{V(\mathbf{a}^T\mathbf{X}_{ik})}\right) \tag{3.23}$$

For the WS No Baselines Model, $\mathbf{a}^T\mathbf{X}_{ik} = 0$ and the variance reduces to $\frac{1}{3n}V(Y_{i1k} - V_{i2k})$. For the WS $X_1 - X_2$ model, substitute $\mathbf{a}^T\mathbf{X}_{ik} = X_{i1k} - X_{i2k}$. Note that the variance of WS $X_1 - X_2$ is never lower than the variance of WS No Baselines.

Under the plausible covariance structures CS, EP, and AR(1), these variance formulas can more

readily be compared. See the Appendix for the exact formulas and details. Under CS, the No Baselines mixed model always has a smaller variance than CFB or any of the WS models. Further, if the CS correlation (ρ) is greater than zero, which is a reasonable assumption, $\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$ has a smaller variance than the No Baselines Mixed model. Under AR(1), the models can be directly compared by calculating ratios of variances for a range of AR(1) correlation values (ρ). Again, assume that $\rho > 0$. Here, the No Baselines Mixed model only has smaller variance than CFB for roughly $\rho \leq 0.73$ and a smaller variance than the WS $X_1 - X_2$ model for roughly $\rho \leq 0.67$. Further, $\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$ consistently has a smaller variance than the No Baselines/CFB mixed models and WS $X_1 - X_2$. Under EP, directly comparing all models is more difficult given that there are three different correlation parameters. However, a simple grid search across all positive values (0.01 to 0.99 by 0.01) of the EP correlation parameters (ρ_1, ρ_2, ρ_3) gives us some insight. Across the range of parameters, $\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$ had a smaller variance than CFB about 90% of the time and a smaller variance than the No Baselines mixed model or the WS $X_1 - X_2$ model about 83% of the time.

Lastly, the two mixed models with baseline covariates are compared. In practice, the period-specific LCB coefficients ($\mathbf{a}_{j\star}$) must be estimated using certain covariance parameters (3.14). Under CS, $\mathbf{a}_{1\star} = \{\sigma^2(\rho - \rho^2), \sigma^2(\rho - \rho^2)\} = \{\sigma^2(\rho - \rho^2), \sigma^2(\rho - \rho^2)\} = \mathbf{a}_{2\star}$, meaning all optimal LCB coefficients are equal. A period-specific optimal LCB is then $X_{i1k} + X_{i2k}$. This actually suggests using $X_{i1k} + X_{i2k}$ as an overall covariate (Equation (3.10) without $\beta_j X_{ijk}$). For a CS structure, since $\beta_j = 0$ for $\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$ (See Appendix C), these models are equivalent. Under EP, $\mathbf{a}_{1\star} = \{\sigma^2(\rho_2 - \rho_1\rho_3), \sigma^2(\rho_3 - \rho_1\rho_2)\}$ and $\mathbf{a}_{2\star} = \{\sigma^2(\rho_3 - \rho_1\rho_2), \sigma^2(\rho_2 - \rho_1\rho_3)\}$. However, based on a grid search across all EP correlation parameters, $\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$ performs similarly to $\mathbf{Y}|\mathbf{X}\mathbf{a}_\star, \overline{\mathbf{X}\mathbf{a}_\star}$. Under AR(1), $\mathbf{a}_{1\star} = \{\sigma^2(\rho - \rho^3), \sigma^2(\rho - \rho^3)\}$ and $\mathbf{a}_{2\star} = \{0, \sigma^2(\rho - \rho^5)\}$. In this case, $\mathbf{Y}|\mathbf{X}\mathbf{a}_\star, \overline{\mathbf{X}\mathbf{a}}_\star$ offers additional gains in efficiency relative to $\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$. This is similar to Chapter 2 (uniform designs), which suggested that LCBs were especially effective in AR(1) type structures. Simulation results further back up these overall comments on model efficiency.

3.4. Estimation

The hypotheses of interest are:

$$\begin{aligned} H_0 &: \tau_A - \tau_B = 0 \\ H_A &: \tau_A - \tau_B \neq 0 \end{aligned}$$

For both the mixed models and WS contrast models, assume that the LCBs are centered at zero. In practice, this is done by subtracting the sample mean out of the LCB(s).

3.4.1. Mixed Models

For any of the mixed models, estimation is done through generalized least squares (GLS) and restricted maximum likelihood (REML). For estimation, we allow the outcomes or conditional outcomes to have an unstructured covariance structure. This is a robust choice given that misspecification of a covariance structure can lead to type I error inflation, regardless of the sample size (Gurka, Edwards, and Muller, 2011). Next, denote the point estimates for τ_A and τ_B as $\hat{\eta} = (\hat{\tau}_A, \hat{\tau}_B)$ with an estimated covariance matrix $V(\hat{\eta})$. Next, let $\mathbf{L}_1^T = (1, -1)$ such that $E(\mathbf{L}_1^T \hat{\eta}) = \tau_A - \tau_B$. Our test statistic is:

$$t_1 = \frac{\mathbf{L}_1^T \hat{\eta}}{\sqrt{\mathbf{L}_1^T V(\hat{\eta}) \mathbf{L}_1}}$$

Notably, by using GLS where the variance components are estimated by REML, this test statistic only approximately follows a t-distribution (Mehrotra, 2014). Thus, the corresponding t-test is only exact asymptotically and the degrees of freedom for a hypothesis test must be approximated. The usual t-test degrees of freedom ($N - p - 1$) is not recommended. For small sample sizes, it is better to use the Kenward-Roger (KR) degrees of freedom adjustment (DDFM=KR in SAS PROC MIXED). This approach adjusts the standard error of the treatment effect estimate by capturing the uncertainty in the estimated covariance structure. Kenward and Roger showed that this delivers a more accurate t-test in small samples (Kenward and Roger, 1997). Finally, we illustrate how to model (3.11), or the period-specific LCB model. Note that LCBprd refers to the period-specific LCBs, while LCBsum refers to the sum of the period-specific LCBs. Data is assumed to be in long format with a row for each period-specific outcome.

```
PROC MIXED DATA=long;
CLASSES seq prd trt subjid;
MODEL y = trt prd LCBprd  LCBprd*prd  LCBsum / ddfm=KR;
REPEATED prd/SUBJECT=subjid(seq) TYPE=UN;
ESTIMATE 'Trt A vs Trt B' trt 1 -1 0 LCBprd 0 LCBsum 0; RUN;
```

3.4.2. WS Models

For the WS baseline models, estimation is done through OLS. For ease of notation, consider the 3×2 design. Using the re-parameterized WS model (3.19), denote the estimated treatment effect as $\widehat{\tau_A - \tau_B}$ with a corresponding estimated variance of $\hat{V}(\widehat{\tau_A - \tau_B})$. Then, our test statistic is:

$$t_2 = \frac{\widehat{\tau_A - \tau_B}}{\sqrt{\hat{V}(\widehat{\tau_A - \tau_B})}}$$

This test statistic is compared to a t-distribution with $(\sum_{i=1}^{s} n_i) - 4$ DF for a model that includes a baseline covariate and $(\sum_{i=1}^{s} n_i) - 3$ DF for a model without a baseline covariate. This test is exact under the assumption of normality. This approach is valid for the general incomplete block design, although the number of nuisance parameters and re-parameterized treatment parameters will vary by design. Lastly, we show how to fit a WS Baseline Model (3.19) for the 3×2 design, where LCB refers to the baseline covariate. Data is assumed to be in wide format with one outcome $(Y_{i1k} - Y_{i2k})$ per subject.

```
/*** Generate Variables based on Re-Parameterized WS Baseline Model: Data in Wide Format***/
data wide; set wide;
if seq IN ('ab','cb') then G1mG2=1; else if seq IN ('ba','bc') then G1mG2=-1; else G1mG2=0;
if seq IN ('bc','ac') then G2pG3=0; else if seq IN ('cb','ca') then G2pG3=-1; else G2pG3=0;
run;
proc mixed data=wide;
class subjid;
MODEL ydiff12 = G1mG2 G2pG3 LCB;
ESTIMATE 'A vs B' G1mG2 1 / CL; run;
```

3.5. Application to a Clinical Trial

In this section, we apply our methods to a real 3×2 data set. The baselines and outcomes are measures of a cardiac safety biomarker. Three treatments were compared: Drug 1 (A), Drug 2 (B), and a placebo (C). These treatments were assigned to the six possible sequences (AB, BA, CA, CB, AC, BC) with five subjects per sequence (N=30 total). The main comparison of interest was A v C. A between-treatment mean difference of 4 units or higher is deemed clinically interesting for this biomarker.

Table 3.2 below shows the treatment effect estimates, standard errors, and p-values for all considered methods (Table 3.1). For this data set, the AICC was lowest under the DCS covariance structure. Thus, the WS adaptive method uses $X_1 - X_2$ as a covariate. Notably, the mixed models with baseline covariates ($\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$; $\mathbf{Y}|\mathbf{X}\mathbf{a}_\star, \overline{\mathbf{X}\mathbf{a}}_\star$) performed similarly to WS $X_1 - X_2$. Further, these baseline models all noticeably outperformed CFB. Overall, the mixed models with baseline covariates and WS $X_1 - X_2$ yielded the smallest standard errors and p-values.

Table 3.2: Real Data Example: 3×2 Crossover Design

Framework	Method	Estimate	SE	p-value
Mixed	No Baselines	4.51	2.14	0.045
Mixed	CFB	4.85	2.88	0.100
Mixed	$\mathbf{Y}\|\mathbf{X}, \overline{\mathbf{X}}$	4.67	1.92	0.021
Mixed	$\mathbf{Y}\|\mathbf{X}\mathbf{a}_\star, \overline{\mathbf{X}\mathbf{a}}_\star$	4.69	1.92	0.021
WS	No Baselines	4.53	2.13	0.043
WS	$X_1 - X_2$	4.80	1.91	0.019
WS	Adaptive	4.80	1.91	0.019

Notes: The AICC did not favor Compound Symmetry and WS Adaptive uses WS $X_1 - X_2$. All methods are described in detail in Section 3.3 and in Table 3.1.

3.6. Simulations

A 3×2 design simulation study was implemented to compare the proposed baseline methods. For this study, we are particularly interested in: (1) Is the Type I error rate maintained at the nominal level for the proposed methods?; (2) How do the baseline mixed models perform with and without a Kenward-Roger degrees of freedom adjustment? (3) Do the WS models maintain the type I error at small sample sizes? (4) How well do the proposed methods perform against the standard CFB model (Table 3.1)?; (5) Is there an overall best method?

We simulated 20,000 trials for a variety of scenarios. The scenarios were defined by the hypotheses (null, $\tau_A - \tau_B = 0$; alternative $\tau_A - \tau_B \neq 0$), the underlying covariance structure Σ, and sample sizes. Sample sizes are balanced across the sequences. For the hypothesis tests, the nominal type I error rate was set at 5%. The temporally ordered baseline and outcome response vector for each subject within each simulated crossover trial were generated from a multivariate normal with a sequence invariant covariance. Covariances CS, EP, and AR(1) (3.6) were explored, where a common variance of $\sigma = 1$ was assumed for all structures. Under CS, $\rho_1 = 0.6$; under EP, $\rho_1 = 0.6$, $\rho_2 = 0.7$, $\rho_3 = 0.5$; under AR(1), $\rho = 0.73$. These are the same covariance settings as the 2×2 design simulations (Section 2.7).

The response vector (X_1, Y_1, X_2, Y_2) was generated with mean $(0, 6, 1, 7)$ under the null. Under the alternative, the response vector was generated with mean $(0, 6 + \tau_{d[i,1]}, 1, 7 + \tau_{d[i,2]})$. For example, for sequence AC, the mean vector was $(0, 6 + \tau_A, 1, 7 + \tau_C)$. Furthermore, for each scenario, $\tau_A - \tau_B = \tau_C - \tau_B$ was chosen such that the mixed model with no baselines approach yielded 80% power (see Appendix B.2). For all methods explored, the estimates of $\tau_A - \tau_B$ were approximately unbiased (results not shown). Of particular interest was the power and type I error rate for the proposed baseline methods. Type I error results and Powerresults are shown in Tables 3.3 and 3.4 respectively. Lastly, all mixed models use a KR DF adjustment unless otherwise stated.

3.6.1. Simulation Results

A number of methods exhibit type I error inflation (Table 3.3). In particular, for all sample sizes and covariance structures, hypothesis tests that did not use a KR DF adjustment showed type I error inflation. In addition, at N=18, the smallest sample size considered, even with the KR adjustment the mixed models all had inflated type I error. The mixed model with period-specific LCBs ($\mathbf{Y}|\mathbf{X}\mathbf{a}_\star, \overline{\mathbf{X}\mathbf{a}}_\star$, Table 1) also exhibited type I error inflation at larger sample sizes. As expected, the WS No Baselines and $X_1 - X_2$ models maintained the nominal type I error rate. The WS Adaptive method, which chooses between including $X_1 - X_2$ as a covariate or not adjusting for any baseline covariate, showed signs of type I error inflation at N=18 under CS and at N=18,24 under EP.

Table 3.4 evaluates the power of each considered method. Compared to the benchmark CFB mixed model, the mixed models with baseline covariates ($\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$; $\mathbf{Y}|\mathbf{X}\mathbf{a}_\star, \overline{\mathbf{X}\mathbf{a}}_\star$) and the WS baseline models were consistently more powerful. Overall, purely in terms of power, the best performers

were the mixed models with baselines ($\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$; $\mathbf{Y}|\mathbf{X}\mathbf{a}_\star, \overline{\mathbf{X}\mathbf{a}}_\star$). While these two methods yielded similar power under CS and EP, $\mathbf{Y}|\mathbf{X}\mathbf{a}_\star, \overline{\mathbf{X}\mathbf{a}}_\star$ does slightly better under AR(1), validating theoretical findings in Section 3.3.3. The WS $X_1 - X_2$ model uniformly exhibited better power than CFB and WS No Baselines across all scenarios and did especially well under the EP covariance structure. Under CS, the WS No Baselines model offers additional gains relative to WS $X_1 - X_2$. WS Adaptive captures this added efficiency under CS while performing approximately as well as WS $X_1 - X_2$ under the other covariance structures.

3.6.2. Simulation Results Summary

Overall, the uniformly best method was the simple $\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$ mixed model. This method was consistently the most powerful while maintaining the type I error rate for $N > 18$. The more complex period-specific LCB mixed model ($\mathbf{Y}|\mathbf{X}\mathbf{a}_\star, \overline{\mathbf{X}\mathbf{a}}_\star$) suffered from type I error inflation for most of the considered sample sizes. However, at larger sample sizes, the period-specific LCB mixed model maintained the nominal type I error rate and was approximately as powerful as the simpler $\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$ mixed model. Under AR(1), the period-specific LCB mixed model was slightly more powerful than $\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$. For larger sample size incomplete block designs, the period-specific LCB model is an attractive data-driven option.

These simulations clearly showed the importance of a KR DF adjustment for mixed models. In all considered scenarios, mixed models without a KR DF adjustment demonstrated signs of type I error inflation. This was true even at N=60 (10 subjects per sequence). Further, all mixed models with a KR adjustment still had type I error inflation at N=18. On the other hand, the WS models had valid type I error rates. This is because estimation is through OLS and there is no need to estimate a covariance matrix. In particular, the WS $X_1 - X_2$ model never had type I error inflation and consistently was more powerful than the benchmark CFB. In a small sample size setting, the WS $X_1 - X_2$ model appears to be more appropriate than the baseline mixed models. Further, if a KR DF adjustment is not available due to software limitations, a mixed model approach should be avoided. In this case, the WS $X_1 - X_2$ model is preferred.

Table 3.3: 3×2 Simulations: Under the Null Hypothesis, Type I Error

Truth	**Estimation**	**Method**	N=18	N=24	N=30	N=36	N=60
Σ_{CS}	Mixed	No Baselines	[5.38]	5.23	4.90	4.91	4.67
Σ_{CS}	Mixed, No KR	No Baselines	[[7.14]]	[[6.51]]	[[5.98]]	[[5.79]]	5.04
Σ_{CS}	Mixed	CFB	[[5.77]]	5.28	5.26	5.17	5.08
Σ_{CS}	Mixed, No KR	CFB	[[7.85]]	[[6.57]]	[[6.21]]	[[6.03]]	[5.46]
Σ_{CS}	Mixed	$\mathbf{Y}\|\mathbf{X}, \overline{\mathbf{X}}$	[[5.80]]	5.27	5.12	5.11	4.96
Σ_{CS}	Mixed, No KR	$\mathbf{Y}\|\mathbf{X}, \overline{\mathbf{X}}$	[[8.32]]	[[6.85]]	[[6.27]]	[[6.06]]	[[5.50]]
Σ_{CS}	Mixed	$\mathbf{Y}\|\mathbf{X}\mathbf{a}_\star, \overline{\mathbf{X}\mathbf{a}_\star}$	[[6.32]]	[5.43]	[5.36]	5.24	5.08
Σ_{CS}	Mixed, No KR	$\mathbf{Y}\|\mathbf{X}\mathbf{a}_\star, \overline{\mathbf{X}\mathbf{a}}_\star$	[[8.61]]	[[6.97]]	[[6.54]]	[[6.22]]	[[5.51]]
Σ_{CS}	WS	No Baselines	5.05	4.80	4.73	4.88	4.80
Σ_{CS}	WS	$X_1 - X_2$	5.09	4.88	4.76	4.84	4.79
Σ_{CS}	WS	Adaptive	[5.39]	5.09	4.99	5.00	4.88
Σ_{EP}	Mixed	No Baselines	[5.46]	5.26	5.13	5.07	5.08
Σ_{EP}	Mixed, No KR	No Baselines	[[7.26]]	[[6.42]]	[[6.16]]	[[5.88]]	[[5.49]]
Σ_{EP}	Mixed	CFB	[[5.58]]	5.11	5.03	5.14	5.01
Σ_{EP}	Mixed, No KR	CFB	[[7.37]]	[[6.50]]	[[6.09]]	[[5.97]]	[[5.51]]
Σ_{EP}	Mixed	$\mathbf{Y}\|\mathbf{X}, \overline{\mathbf{X}}$	[[5.88]]	5.30	5.11	5.21	5.00
Σ_{EP}	Mixed, No KR	$\mathbf{Y}\|\mathbf{X}, \overline{\mathbf{X}}$	[[8.07]]	[[6.62]]	[[6.21]]	[[6.09]]	[5.41]
Σ_{EP}	Mixed	$\mathbf{Y}\|\mathbf{X}\mathbf{a}_\star, \overline{\mathbf{X}\mathbf{a}_\star}$	[[6.28]]	[5.41]	5.27	[5.33]	5.01
Σ_{EP}	Mixed, No KR	$\mathbf{Y}\|\mathbf{X}\mathbf{a}_\star, \overline{\mathbf{X}\mathbf{a}}_\star$	[[8.47]]	[[6.83]]	[[6.36]]	[[6.18]]	[[5.46]]
Σ_{EP}	WS	No Baselines	4.86	4.85	4.87	5.03	5.07
Σ_{EP}	WS	$X_1 - X_2$	4.99	4.93	4.77	4.83	4.80
Σ_{EP}	WS	Adaptive	[[5.71]]	[5.36]	5.05	5.04	4.83
Σ_{AR}	Mixed	No Baselines	[[5.61]]	5.22	5.25	5.07	5.20
Σ_{AR}	Mixed, No KR	No Baselines	[[7.34]]	[[6.37]]	[[6.11]]	[[5.75]]	[[5.62]]
Σ_{AR}	Mixed	CFB	[[5.88]]	5.13	5.21	5.22	5.01
Σ_{AR}	Mixed, No KR	CFB	[[7.88]]	[[6.58]]	[[6.32]]	[[5.97]]	[[5.58]]
Σ_{AR}	Mixed	$\mathbf{Y}\|\mathbf{X}, \overline{\mathbf{X}}$	[[5.82]]	5.23	5.04	4.96	4.91
Σ_{AR}	Mixed, No KR	$\mathbf{Y}\|\mathbf{X}, \overline{\mathbf{X}}$	[[8.66]]	[[7.19]]	[[6.52]]	[[6.08]]	[[5.58]]
Σ_{AR}	Mixed	$\mathbf{Y}\|\mathbf{X}\mathbf{a}_\star, \overline{\mathbf{X}\mathbf{a}_\star}$	[[6.66]]	[[5.74]]	[5.38]	5.25	5.23
Σ_{AR}	Mixed, No KR	$\mathbf{Y}\|\mathbf{X}\mathbf{a}_\star, \overline{\mathbf{X}\mathbf{a}}_\star$	[[9.00]]	[[7.21]]	[[6.83]]	[[6.08]]	[[5.68]]
Σ_{AR}	WS	No Baselines	4.83	4.91	4.76	5.01	5.16
Σ_{AR}	WS	$X_1 - X_2$	4.95	4.90	4.73	5.09	4.98
Σ_{AR}	WS	Adaptive	5.11	5.00	4.79	5.12	4.97

Notes: Values (type I error %) are shown. Entries are in brackets/double brackets if the type I error is two/three SE's above 5% ($> 5.31\%$,$> 5.46\%$) based on 20,000 simulations. No Baselines (Mixed or WS) uses no baseline covariate; CFB uses change scores with no baseline covariate; $\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$ and $\mathbf{Y}|\overline{\mathbf{X}\mathbf{a}}_\star$ respectively use period-specific baselines and period-specific LCBs as covariates; WS $X_1 - X_2$ includes $X_1 - X_2$ as a covariate in a mixed model. WS Adaptive chooses between WS No Baselines and WS $X_1 - X_2$ based on AICC. All methods are summarized in Table 3.1. CS = Compound Symmetry; EP = Equipredictability; AR = Auto-regressive(1). N refers to total sample size.

3.7. Discussion

For the incomplete block crossover design, we explored a number of approaches for using outcome and baseline measurements in hypothesis testing of the main treatment effect. Baseline utilization

Table 3.4: 3×2 Simulations: Under the Alternative Hypothesis, Power

Truth	Estimation	Method	N=18	N=24	N=30	N=36	N=60
Σ_{CS}	Mixed	No Baselines	[80.1]	**80.4**	**80.2**	79.9	79.9
Σ_{CS}	Mixed	CFB	[[59.9]]	59.7	59.6	58.7	58.9
Σ_{CS}	Mixed	$\mathbf{Y}\|\mathbf{X}, \overline{\mathbf{X}}$	[[78.2]]	**80.5**	**81.5**	**81.4**	**82.3**
Σ_{CS}	Mixed	$\mathbf{Y}\|\mathbf{X}\mathbf{a}_\star, \overline{\mathbf{X}\mathbf{a}}_\star$	[[78.3]]	[80.5]	[81.4]	**81.3**	**82.3**
Σ_{CS}	WS	No Baselines	**79.3**	78.8	78.5	77.6	77.5
Σ_{CS}	WS	$X_1 - X_2$	**76**	76.6	76.9	76.3	76.7
Σ_{CS}	WS	Adaptive	[78.8]	78.4	78.3	77.5	77.4
Σ_{EP}	Mixed	No Baselines	[80.1]	80.3	80.4	79.8	80
Σ_{EP}	Mixed	CFB	[[83.1]]	83.8	83.5	82.6	82.9
Σ_{EP}	Mixed	$\mathbf{Y}\|\mathbf{X}, \overline{\mathbf{X}}$	[[86.9]]	**88.6**	**89.7**	**89.5**	**90.4**
Σ_{EP}	Mixed	$\mathbf{Y}\|\mathbf{X}\mathbf{a}_\star, \overline{\mathbf{X}\mathbf{a}}_\star$	[[86.9]]	[88.7]	**89.6**	[89.6]	**90.5**
Σ_{EP}	WS	No Baselines	**79.5**	79	78.9	77.8	77.6
Σ_{EP}	WS	$X_1 - X_2$	**86.9**	**87.4**	87.8	**87.2**	87.5
Σ_{EP}	WS	Adaptive	[[86.3]]	[86.8]	87.4	86.8	87.5
Σ_{AR}	Mixed	No Baselines	[[80.3]]	80.1	80.3	80.1	79.7
Σ_{AR}	Mixed	CFB	[[80.5]]	80.3	80.8	80.4	80
Σ_{AR}	Mixed	$\mathbf{Y}\|\mathbf{X}, \overline{\mathbf{X}}$	[[86.8]]	**88.5**	**89.9**	**90**	**90.4**
Σ_{AR}	Mixed	$\mathbf{Y}\|\mathbf{X}\mathbf{a}_\star, \overline{\mathbf{X}\mathbf{a}}_\star$	[[89.8]]	[[91.3]]	[92]	**92.1**	**92.3**
Σ_{AR}	WS	No Baselines	78.9	77.9	77.9	77.4	76.4
Σ_{AR}	WS	$X_1 - X_2$	**81.2**	**81.4**	81.7	81.7	81.5
Σ_{AR}	WS	Adaptive	**81.4**	**81.4**	**81.8**	81.7	81.5

Note: Values (Power %) are shown in bold if method yields the highest or second highest power in that sample size/covariance structure combination without type I error inflation. Entries are in brackets/double brackets if under the same scenario, but under the null hypothesis, the type I error is two/three SE's above 5% ($> 5.31\%$,$> 5.46\%$) based on 20,000 simulations. No Baselines (Mixed or WS) uses no baseline covariate; CFB uses change scores with no baseline covariate; $\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$ and $\mathbf{Y}|\overline{\mathbf{X}\mathbf{a}}_\star$ respectively use period-specific baselines and period-specific LCBs as covariates; WS $X_1 - X_2$ includes $X_1 - X_2$ as a covariate in a mixed model. WS Adaptive chooses between WS No Baselines and WS $X_1 - X_2$ based on AICC. All methods are summarized in Table 3.1. CS = Compound Symmetry; EP = Equipredictability; AR = Auto-regressive(1). N refers to total sample size.

was examined for mixed effects models and for within-subject (WS) contrast only fixed effects models. In a mixed model, the overall estimate of a treatment effect is obtained from a weighted sum of within-subject and between-subject information. In the within-subject contrast models, the mixed model is re-parameterized and only within-subject information is used. Consequently, the form of a treatment effect estimate differs between the two modeling frameworks. Thus, our baseline utilization methods also differed depending on the underlying modeling approach.

For both frameworks, data-driven models were developed that took advantage of the joint covariance structure of the baselines and outcomes. For the mixed model framework, we developed a period-specific LCB model. The overall estimate of a treatment effect in a mixed model is a summation of period-specific outcomes. Consequently, the period-specific LCBs minimized the conditional

variance of the period-specific outcomes. Based on simulation results, this period-specific LCB model yielded similar power to the simpler $\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$ mixed model but with increased type I error inflation. However, the period-specific LCB model did perform as well or better than the $\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$ mixed model at large sample sizes. In particular, the period-specific LCB model did better under an AR(1) structure, reflecting the fact that estimated LCBs are especially useful for temporally related covariance structures (Jemielita, Putt, and Mehrotra, 2016).

For the WS models, two data-driven methods were discussed. Although not shown in the data example or simulations, the WS LCB model finds the LCB that minimizes the conditional variance corresponding to the WS model treatment effect estimate. While theoretically appealing, this approach offers little for a 2-period design and is not recommended. On the other hand, the WS Adaptive method used information criteria to guide the choice of baseline covariates. While this method did as well or better than the WS $X_1 - X_2$ model, the nominal type I error rate was not always maintained.

The mixed model with estimated period-specific LCBs failed to consistently maintain the nominal type I error rate across a number of sample sizes and covariance structures. This is primarily due to the fact that our LCB is estimated and then included in a regression model as if it is a fixed covariate. Essentially, the estimation of the LCB is not accounted for during inference. This typically resulted in a deflated standard error. One way to obtain valid inference is to use permutation or bootstrap resampling. However, at least for the 3×2 design, this isn't worth the effort as simpler methods yielded similar power gains but without the inflated type I error.

A large focus of this paper was comparing the mixed model and the within-subject (WS) model. While the mixed models with baseline covariates were more efficient than the WS baseline models, there are some theoretical advantages to using the WS model. First, while this point is minor, an OLS model yields exact estimates. The mixed model, fit with GLS and REML, yields approximate estimates since numerical optimization is required. Second, under a mixed model, hypothesis tests are only exact asymptotically, even if the data was truly normal. In contrast, the hypothesis tests for the WS models are exact under the assumption of normality. Moreover, for randomized clinical trials, Judkins and Porter showed that under the null hypothesis and a wide range of non-normal distributions, OLS models provide valid inference and control the type I error (Judkins and Porter, 2015). Lastly, given that the hypothesis tests are only exact asymptotically for the mixed

models, a Kenward-Roger degree of freedom adjustment is required (Kenward and Roger, 1997). This inflates the standard error of a treatment effect estimate to account for the uncertainty in the estimated covariance structure. This in itself reduces the efficiency of the treatment effect estimate. For the WS models, no such adjustment is needed.

The WS baseline models can be practical alternatives to the mixed models. First, WS baseline models may be preferable for small sample sizes. In our simulations, the mixed models with a KR DF adjustment all suffered type I error inflation at the smallest sample size (N=18, 3 per sequence). Notably, this result matches up to previous research. Specifically, Chen and Wei showed that for the 3×3 and $4\times$ design, a no baselines mixed model with a KR DF adjustment exhibited type I error for roughly $N < 24$ (Chen and Wei, 2003). In contrast, the WS $X_1 - X_2$ model consistently maintained the nominal type I error rate. Second, WS baseline models should be used if a KR DF adjustment is not available. Without a KR DF adjustment, the mixed models exhibited type I error inflation across all simulation scenarios. Overall, the WS $X_1 - X_2$ model should be used in place of the baseline mixed models if sample size is a concern or if a KR DF adjustment cannot be implemented.

Overall, we have examined and proposed a number of baseline utilization methods for an inomplete block crossover design. While various data driven approaches were explored for the 3×2 design, the mixed model where period-specific outcomes are regressed against their respective period-specific baselines ($\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$) proved to be the uniformly best method. This method largely outperformed the commonly used change from baseline mixed models. At larger sample sizes, the more complex period-specific LCB mixed model could instead be used. For smaller sample sizes, a mixed model approach may not be appropriate. Even with a Kenward-Roger degrees of freedom adjustment, a mixed model approach will likely inflate the type I error. Additionally, if a Kenward-Roger degrees of freedom adjustment is not available, a WS model should also be used. In either case, the WS $X_1 - X_2$ model is an attractive and easy to implement alternative.

CHAPTER 4

EFFICIENT BASELINE UTILIZATION IN CROSSOVER DESIGNS USING BOTH PARAMETRIC AND NON-PARAMETRIC REGRESSIONS

4.1. Motivation and Literature Review

For normally distributed data, including baseline measurements in an OLS model can substantially improve the efficiency of the estimated treatment effect. However, departures from normality can affect both the validity and the efficiency of OLS (Hettmansperger and Mckean, 2010; Huber, 1973). Since OLS models minimize a quadratic function, OLS estimates can be overly sensitive to extreme values. In this setting, OLS may not be the most efficient modeling approach and a robust or nonparametric alternative may be preferred. Thus, a primary aim of this paper is to develop the framework for efficient baseline utilization under a robust or nonparametric setting.

For crossover designs and particularly the AB/BA design, a variety of rank-based nonparametric models without baseline adjustment have been discussed in the literature. For a two-treatment design, Koch developed the framework for tests of treatment effects, period effects, and carryover effects based on the Wilcoxon rank sum test (Koch, 1972). Under the null hypothesis, the distribution of the Wilcoxon rank sum test is distribution-free or nonparametric (Hettmansperger and Mckean, 2010). For designs with more than two treatments, such as the 3×3 design, Ohrvik proposed a rank-based statistic based on aligned outcomes (Ohrvik, 1998). Aligned outcomes are obtained by substracting period effect estimates from each period-specific outcome. This approach is equivalent to the Wilcoxon rank sum test for the 2×2 design and is powerful across a range of distributions (Ohrvik, 1998; Putt and Chinchilli, 2004). For the 3×3 design, Bellavance and Tardif proposed a rank-based statistic based on transforming the analysis of the 3×3 design into one of a randomized block design (Bellavance and Tardif, 1995). This was also shown to be efficient across a range of distributions. Relative to the Ohvrik's method, this approach was shown to be especially powerful in situations where carryover was present, although less powerful for small carryover effects or no carryover (Correa and Bellavance, 2001). Relative to OLS based estimation, asymptotic theory indicates that rank-based tests show slight losses in efficiency under a normal distribution but are more efficient for non-normal distributions (Bellavance and Tardif, 1995; Hettmansperger

and Mckean, 2010; Ohrvik, 1998; Putt and Chinchilli, 2004).

There has been limited work on the use of baseline measurements with rank-based approaches. Baseline utilization through rank-based statistics has also been discussed for the 2×2 design. In their overview of nonparametric estimation for crossover designs, Tudor and Koch mention several approaches for baseline adjustment (Tudor and Koch, 1994). For this work, the focus was on baseline methods for hypothesis tests related to carryover effects. With the goal of increasing efficiency, Tudor and Koch suggested either using change scores (outcome-baseline differences) in a Wilcoxon rank sum test or using nonparametric covariance adjustment. Nonparametric covariance adjustment first regresses the outcomes against the baselines while ignoring sequence assignment. The estimated residuals are then used in a rank-based statistic, such as the Wilcoxon rank sum test. This approach depends both on the initial regression model as well as the chosen baseline covariate(s). While Tudor and Koch did not explicitly recommend a specific baseline covariate, they suggested obtaining the residuals by regressing the ranks of outcomes against the ranks of the baseline covariate (Tudor and Koch, 1994). This is a robust option and is called Rank analysis of covariance (ANCOVA) (Lavange and Koch, 2006; Quade, 1967). Similarly, Tsai and Patel proposed estimating the residuals by regressing each outcome against their respective period-specific baseline in a robust regression model (Tsai and Patel, 1996). In general, previous work focused on the 2×2 design and generally involves a two-step approach where the variability of the outcomes is reduced, either by regression or substracting the baseline, and then the adjusted outcomes are used in a Wilcoxon rank sum test.

Our earlier work suggests these methods might be improved in two ways. First, using mixed effects models or OLS, we observed that using linear combinations of baselines (LCBs) can be a powerful strategy in both uniform and incomplete block designs. We propose using R-estimation to directly estimate treatment effects while simultaneously adjusting for baselines. R-estimation is a robust nonparametric regression approach in which regression parameters are estimated by minimizing the Jaeckel dispersion function (Hettmansperger and Mckean, 2010). This function is based on ranks and some standardized score function (Jaeckel, 1972; Jureckova, 1971). Importantly, the score function can be adjusted to better fit certain distributions of data. Relative to OLS estimation, choosing the Wilcoxon score function yields minor losses in efficiency under normality but increased efficiency for other distributions (Hettmansperger and Mckean, 2010). Wilcoxon rank

sum tests, which are utilized in the two-step nonparametric baseline adjustment approaches (Tsai and Patel, 1996; Tudor and Koch, 1994), have similar properties. Efficient baseline utilization is straightforward in R-estimation. As we will show, treatment effects and the optimal LCB can be estimated simultaneously within the context of a regression model or loss function. For the two-step approaches, it is less clear as how to efficiently incorporate LCBs. Finally, while previous nonparametric baseline research focused on the AB/BA design, R-estimation models offer a unified nonparametric baseline framework for a general crossover designs.

The primary aim of this chapter is to efficiently incorporate linear combinations of baselines (LCBs) into the analysis of crossover designs under a general regression model or loss function. While this is framed for a general loss function, only OLS and R-estimation are considered. An adaptive model selection based procedure is also proposed. In general, among a set of OLS and R-estimation baseline models, the adaptive procedure selects the best fitting model. This is an extension of previous research which selected among a set of OLS baseline models (Jemielita, Putt, and Mehrotra, 2016). Overall, this work will show the advantage of R-estimation baseline models relative to previously recommended nonparametric baseline adjusted and unadjusted models. Further, we will demonstrate that a data-driven procedure which selects among a set of baseline models fit under OLS and R-estimation yields sizable gains in efficiency under a variety of covariance structures and distributions.

The model and notation for a general crossover design is defined in Section 4.2. Efficient baseline utilization under a general loss function is then discussed. Section 4.3 covers estimation under OLS and R-Estimation. Section 4.4 discusses optimal utilization of certain LCB covariate in a general setting. Section 4.5 discusses the adaptive model selection procedure and how to obtain valid inference through nonparametric bootstrap resampling. Section 4.6 covers the application of the proposed methods for a general crossover design, with emphasis on the 2×2 and 3×3 designs. In Section 4.7, the proposed methods are evaluated through simulations for the 2×2 and 3×3 design. For the 2×2 design, we additionally examine the accuracy of optimal LCB estimates based on either OLS or R-estimation. In section 4.8, the analysis of real data sets for the 2×2 and 3×3 designs are examined. Lastly, section 4.9 summarizes the overall findings.

4.2. Models and Notation

Consider a general crossover design. Let:

$$\begin{aligned}\mathbf{X}_{ik} &= (X_{i1k}, ..., X_{ijk}, ..., X_{ipk})^T \\ \mathbf{Y}_{ik} &= (Y_{i1k}, ..., Y_{ijk}, ..., Y_{ipk})^T \end{aligned} \tag{4.1}$$

be the vectors of baseline and outcome measurements respectively, where $i = 1, ..., s$ indexes sequence, $j = 1, ..., p$ indexes period, $d = A, B, ..., Z$ denotes treatment, and $k = 1, ..., n_i$ indexes subject k in sequence i, where subjects are assumed to be independent of each other. Next, assume that $(\mathbf{X}_{ik}, \mathbf{Y}_{ik})^T$ follows some unknown distribution with some sequence invariant covariance matrix Σ. The expectations are defined by their individual elements:

$$E(Y_{ijk}) = \mu + \pi_j + \tau_{d[i,j]} \tag{4.2}$$

$$E(X_{ijk}) = \zeta_j \tag{4.3}$$

where μ is the overall mean, π_j is the effect of period j with $\sum_j \pi_j = 0$, $\tau_{d[i,j]}$ is the effect of treatment d (defined by the period j and sequence i) with $\sum_d \tau_d = 0$, and ζ_j represents the mean of a baseline in period j. Note that a null carryover is assumed for both the baselines and outcomes.

Briefly, as in Chapters 2-3, consider the Compound Symmetry (CS), Equipredictability (EP), and AR(1) covariance structures. CS assumes a common variance and correlation. EP is a four parameter structure that assumes a common variance with $\rho_1 = \text{corr}(Y_j, X_j)$ for $j = j$, $\rho_2 = \text{corr}(X_j, X_{j'}) = \text{corr}(Y_j, Y_{j'})$ for $j \neq j'$, and $\rho_3 = \text{corr}(Y_j, X_{j'})$ for $j \neq j'$. Essentially, the correlation between baselines and outcomes is allowed to vary depending on whether pairs of measurements are within or between periods. AR(1) is the familiar two-parameter covariance structure where the correlation between pairs of measurements decreases over time. For the 2×2 design, $V((X_1, Y_1, X_2, Y_2)^T)$ under these covariance structures can be written as:

$$\Sigma_{CS} = \sigma \begin{array}{c} \begin{matrix} X_1 & Y_1 & X_2 & Y_2 \end{matrix} \\ \begin{bmatrix} 1 & \rho & \rho & \rho \\ & 1 & \rho & \rho \\ & & 1 & \rho \\ & & & 1 \end{bmatrix} \end{array} \quad \Sigma_{EP} = \sigma \begin{array}{c} \begin{matrix} X_1 & Y_1 & X_2 & Y_2 \end{matrix} \\ \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ & 1 & \rho_3 & \rho_2 \\ & & 1 & \rho_1 \\ & & & 1 \end{bmatrix} \end{array} \quad \Sigma_{AR(1)} = \sigma \begin{array}{c} \begin{matrix} X_1 & Y_1 & X_2 & Y_2 \end{matrix} \\ \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ & 1 & \rho & \rho^2 \\ & & 1 & \rho \\ & & & 1 \end{bmatrix} \end{array} \tag{4.4}$$

The goal is to efficiently incorporate baselines in the estimation of a treatment effect under a general loss function. Let $\boldsymbol{b} = [\mathbf{b}_1, ..., \mathbf{b}_s]$ be a set of p-length vectors of constants such that $Y^{\star}_{ik} = \mathbf{b}_i \mathbf{Y}_{ik}$ is a within-subject contrast. Within-subject contrasts are chosen to target a specific pairwise treatment effect. For example, in the AB/BA design, $\mathbf{b}_i = [1, -1]$ for $i = 1, 2$ such that $\frac{1}{2}E[\mathbf{b}_1\mathbf{Y}_1 - \mathbf{b}_2\mathbf{Y}_2] = \frac{1}{2}E[(Y_{11} - Y_{12}) - (Y_{21} - Y_{22})] = \tau_A - \tau_B$. For any crossover design, a linear combination of baselines (LCB) model can then be written as:

$$Y^{\star}_{ik} = \mathbf{W}_{ik}\boldsymbol{\gamma} + \beta \mathbf{a}^T \mathbf{X}_{ik} + \epsilon_{ik} \tag{4.5}$$

where $\mathbf{W}_{ik}$ is a vector of relevant contrasts related to nuisance, intercept, and treatment parameters ($\boldsymbol{\gamma}$), $\mathbf{a} = (a_1, ..., a_p)^T$ is a p-vector of constants such that our LCB is $\mathbf{a}^T\mathbf{X}_{ik} = \sum_{j=1}^{p} a_j \mathbf{X}_{ik}$, β is the regression coefficient corresponding to the LCB, and ϵ_{ik} is the error. Next, for some loss function L, the parameters and the optimal LCB $\mathbf{a}^T_{\star}\boldsymbol{X}_{ik}$ are estimated by solving:

$$(\hat{\boldsymbol{\gamma}}, \hat{\beta}, \boldsymbol{a}_{\star}) = \underset{\boldsymbol{\gamma}, \beta, \boldsymbol{a}}{\text{Argmin}} \sum_{i=1}^{s} \sum_{k=1}^{n_i} L(Y^{\star}_{ik} - \mathbf{W}_{ik}\boldsymbol{\gamma} - \beta \mathbf{a}^T \mathbf{X}_{ik}) \tag{4.6}$$

Generally, (4.6) can be solved numerically. This is accomplished for any loss function by finding the partial derivatives for the various parameters and using Newton optimization.

Under ordinary least squares (OLS), the estimated model parameters minimize the sum of squared residuals:

$$(\hat{\boldsymbol{\gamma}}, \hat{\beta}, \boldsymbol{a}_{\star}) = \text{Argmin} \sum_{i=1}^{s} \sum_{k=1}^{n_i} \epsilon^2_{ik} = \underset{\boldsymbol{\gamma}, \beta, \boldsymbol{a}}{\text{Argmin}} \sum_{i=1}^{s} \sum_{k=1}^{n_i} (Y^{\star}_{ik} - \mathbf{W}_{ik}\boldsymbol{\gamma} - \beta \mathbf{a}^T \mathbf{X}_{ik})^2 \tag{4.7}$$

OLS is equivalent to the maximum likelihood estimator under the assumption of normality. Further, the optimal LCB ($\mathbf{a}^T_{\star}\boldsymbol{X}_{ik}$) under a squared-loss function is equivalent to finding the optimal LCB under a normality assumption (See Appendix D.1).

R-estimation extends rank-based methods to a regression setting where estimated model parameters minimize the Jaeckel dispersion function (Jaeckel, 1972; Jureckova, 1971):

$$(\hat{\boldsymbol{\gamma}}, \hat{\beta}, \boldsymbol{a}_{\star}) = \text{Argmin} D_{\varphi}(\boldsymbol{\gamma}, \beta, \boldsymbol{a}) = \underset{\boldsymbol{\gamma}, \beta, \boldsymbol{a}}{\text{Argmin}} \sum_{i=1}^{s} \sum_{k=1}^{n_i} a(R(Y^{\star}_{ik} - \mathbf{W}_{ik}\boldsymbol{\gamma} - \beta \mathbf{a}^T \mathbf{X}_{ik}))(Y^{\star}_{ik} - \mathbf{W}_{ik}\boldsymbol{\gamma} - \beta \mathbf{a}^T \mathbf{X}_{ik}) \tag{4.8}$$

where $R(.)$ denotes the rank, and $a(i) = \varphi(i/(n+1))$ for some non-decreasing standardized score function $\varphi(u)$. Model fit and estimation efficiency can be enhanced by choosing a score function that best fits the shape of the observed data (Hettmansperger and Mckean, 2010). For example, Kloke discusses specific score functions that are geared towards increased efficiency for skewed data (Kloke and McKean, 2012). Our work focuses on the Wilcoxon score function. This function is efficient across a range of distributions (Hettmansperger and Mckean, 2010; Kloke and McKean, 2012), a property that is especially true for symmetric distributions. This score yields:

$$a(i) = \frac{\sqrt{12}}{n+1}(R(Y_{ik}^{\star} - \mathbf{W}_{ik}\boldsymbol{\gamma} - \beta\mathbf{a}^T\mathbf{X}_{ik}) - 0.5)$$

Relative to OLS, the Wilcoxon score function is 95% efficient under normality (Hettmansperger and Mckean, 2010). As in OLS, R-estimation yields point estimates, standard errors, p-values, and model fit statistics. The primary difference is that R-estimation uses a rank-based metric for estimation.

Lastly, for convex loss functions, including squared-loss (OLS) and the Jaeckel dispersion function (R-Estimation), a unique minimum exists for Equation (4.6). Of note, if the LCB or baseline covariate is fixed or pre-specified, Equation (4.6) is now solved with respect to γ and β. Section 4.6 illustrates this general LCB model for the 2×2 and 3×3 design.

4.3. Estimation and Hypothesis Testing

The two loss functions considered for our general LCB model (Equation 4.6) are squared-loss (OLS) and the Jaeckel dispersion function (R-Estimation). OLS is optimal if the joint distribution of the measurements is normal. R-Estimation is a robust nonparametric alternative that uses a rank-based loss function to estimate the desired regression parameters. For a heavy-tailed or mixture distributions, R-estimation is more efficient than OLS (Hettmansperger and Mckean, 2010).

Given our general LCB regression model (Equation 4.5) with a fixed LCB, for example $\mathbf{a}^T\mathbf{X}_{ik} = X_{i1k} - X_{i2k}$, assume that $\boldsymbol{\gamma} = (\mu, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ is a t-length vector where μ is the intercept and $\boldsymbol{\gamma}_1$ is q-length vector with $t \geq q$. Let $\mathbf{Z} = [\mathbf{1}, \mathbf{W}_1, \mathbf{W}_2, \mathbf{Xa}]$ be the overall design matrix with corresponding regression parameters $(\mu, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \beta)$ and let $\mathbf{Y}^{\star}$ be the stacked vector of within-subject contrasts

($Y_{ik}^{\star}$). Hypothesis tests are of estimated parameters:

$$H_0 : \boldsymbol{\gamma}_1 = \mathbf{0}$$
$$H_A : \boldsymbol{\gamma}_1 \neq \mathbf{0}$$

This formulation is useful for comparing nested models and forms the basis of a specific R-estimation test statistic. Hypothesis tests based on linear combinations of regression parameters can also be constructed for OLS and R-estimation through Wald type tests.

4.3.1. Ordinary Least Squares

Using OLS, parameter estimates and the covariance of parameter estimates have closed form solutions:

$$(\hat{\mu}, \hat{\boldsymbol{\gamma}}_1, \hat{\boldsymbol{\gamma}}_2, \hat{\beta})^T = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}^{\star}$$
$$V((\hat{\mu}, \hat{\boldsymbol{\gamma}}_1, \hat{\boldsymbol{\gamma}}_2, \hat{\beta})^T) = \sigma^2(\mathbf{Z}^T\mathbf{Z})^{-1}$$

where σ^2 is estimated with $s^2 = \hat{\boldsymbol{\epsilon}}^T\hat{\boldsymbol{\epsilon}}/((\sum_i^s n_i) - (t-1))$ and $\hat{\boldsymbol{\epsilon}}$ is the stacked vector of estimated residuals based on the OLS parameter estimates. Hypothesis testing typically proceeds through a Wald test, although a likelihood ratio test is another approach. Under normality, the tests are algebraically equivalent (Hettsmansperger and Mckean, 1983). Using the Wald test framework, our test statistic is:

$$F_{OLS} = \frac{\hat{\boldsymbol{\gamma}}_1^T(\hat{V}(\hat{\boldsymbol{\gamma}}_1))^{-1}\hat{\boldsymbol{\gamma}}_1}{q}$$

where $\hat{V}(\hat{\boldsymbol{\gamma}}_1)$ corresponds to the [2:(q+1), 2:(q+1)] elements of $s^2(\mathbf{Z}^T\mathbf{Z})^{-1}$. Irrespective of the distribution, asymptotically this statistic has a F distribution with q and $n-t-1$ degrees of freedom under the null hypothesis. This test is exact if the true distribution of the conditional outcomes ($Y_{ik}^{\star}|\mathbf{a}^T\mathbf{X}_{ik}$) is normal.

4.3.2. R-Estimation

Using R-estimation, parameter estimates are determined numerically through Newton optimization. Let $\mathbf{Z}_{\star} = [\mathbf{W}_1, \mathbf{W}_2, \mathbf{Xa}]$ be the design matrix corresponding to the non-intercept parameters,

$(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \beta)$. Under certain regularity conditions, the non-intercept and intercept estimates are independent (Hettmansperger and Mckean, 2010 pg 166). A key condition, called the Huber condition, is that $(\sum_i^s n_i)^{-1}\mathbf{Z}_\star^T\mathbf{Z}_\star$ converges to a positive definite matrix as $\sum_i^s n$ goes to infinity. Asymptotically:

$$(\hat{\boldsymbol{\gamma}}_1, \hat{\boldsymbol{\gamma}}_2, \hat{\beta})^T \sim N\left((\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \beta)^T, \tau_\varphi^2(\mathbf{Z}_\star'\mathbf{Z}_\star)^{-1}\right) \tag{4.9}$$

where τ_φ is the R-estimation scale parameter (Hettmansperger and Mckean, 2010 pg 147). The scale parameter is estimated through a density-type estimator based on the residuals (Koul, Sievers, and Mckean, 1987). As in OLS, Wald type tests can be utilized for inference. Alternatively, the dispersion test, the likelihood ratio test analagoue for R-estimation, can be used. The dispersion test is as follows:

1. Fit the full model and calculate the dispersion $D(\hat{\boldsymbol{\gamma}}, \hat{\beta}, \boldsymbol{a}_\star)_\varphi$. Estimate the scale parameter τ_φ.

2. Fit the reduced model where we do not include the variables corresponding to $\boldsymbol{\gamma}_1$. Use the reduced model parameters and calculate the reduced dispersion $D((\mathbf{0}, \hat{\boldsymbol{\gamma}}_2, \hat{\beta}, \boldsymbol{a}_\star)_R)_\varphi$.

3. Evaluate:

$$F = \frac{D((\mathbf{0}, \hat{\boldsymbol{\gamma}}_2, \hat{\beta}, \boldsymbol{a}_\star)_R)_\varphi - D(\hat{\boldsymbol{\gamma}}, \hat{\beta}, \boldsymbol{a}_\star)_\varphi}{q\hat{\tau}_\varphi/2}$$

   Under the null hypothesis, this statistic has an approximate χ^2 distribution with q degrees of freedom. Mckean and Sheather showed in small-sample studies that it is best to compare the test statistic to an F distribution with q and $n - t - 1$ degrees of freedom Mckean and Sheather, 1991. This modification better maintained the nominal type I error rate over a range of designs, sample sizes, and distributions (Hettmansperger and Mckean, 2010; Mckean and Sheather, 1991).

At small sample sizes, Mckean and Sheather showed that Wald tests are less efficient than the dispersion test (Mckean and Sheather, 1991). While this was shown for a parallel group trial, our own empirical simulations verify this observation for crossover designs. For R-estimation, Wald tests and dispersion tests are only asymptotically equivalent (Hettsmansperger and Mckean, 1983). While dispersion tests are preferred for hypothesis testing, this procedure does not readily yield

confidence intervals. Based on the asymptotic distribution of the 'R'-estimates (Equation 4.9), Wald type confidence intervals can instead be constructed. R-estimation models can be fit through the "Rfit" package in R (Kloke and McKean, 2012). Code examples are given in Section 4.6.

Lastly, if the LCB is estimated directly within Equation 4.6, the described OLS and R-estimation inferential procedures will have inflated type I error at small sample sizes. This is because our model includes an estimated covariate but the inferential procedure assumes that all covariates are fixed. We address this using nonparametric bootstrap resampling (Section 4.5).

4.4. Selecting the LCB

For normally distributed data, we previously showed that the optimal LCB ($\mathbf{a}_\star^T\mathbf{X}_{ik}$) can be found by minimizing the conditional variance of the desired treatment effect estimate (Jemielita, Putt, and Mehrotra, 2016). In a more general setting for some known distribution, the optimal LCB ($\mathbf{a}_\star^T\mathbf{X}_{ik}$) minimizes the total conditional variance of the treatment effect estimate:

$$\mathbf{a}_\star^T = \underset{\boldsymbol{a}^T}{\operatorname{Argmin}} \left(\sum_i^s V(Y_{ik}^\star | \mathbf{a}^T\mathbf{X}_{ik}) \right) \tag{4.10}$$

For normally distributed data:

$$V(Y_{ik}^\star|\mathbf{a}^T\mathbf{X}_{ik})_N = V(Y_{ik}^\star) - \frac{\text{cov}(Y_{ik}^\star, \mathbf{a}^T\mathbf{X}_{ik})^2}{V(\mathbf{a}^T\mathbf{X}_{ik})}$$

While we could use results for normally distributed data, we explored analytically how these results might generalize to the T-distribution, a heavy tailed symmetric distribution often used to explore the properties of robust and nonparametric estimators. Asymptotically, for a multivariate normal distribution and a multivariate T-distribution with the same covariance matrix Σ, the optimal LCB is identical. Assume that the baselines and outcomes follow a multivariate T-distribution with v degrees of freedom. Then, the distribution of the outcomes conditional on the baselines also follows a T-distribution (Ding, 2016). In this case:

$$V(Y_{ik}^\star|\mathbf{a}^T\mathbf{X}_{ik})_T = \frac{v + (\mathbf{a}\mathbf{X}_{ik} - E(\mathbf{a}^T\mathbf{X}_{ik}))^2 V(\mathbf{a}^T\mathbf{X}_{ik})^{-1}}{v+1} \left(V(Y_{ik}^\star) - \frac{\text{cov}(Y_{ik}^\star, \mathbf{a}^T\mathbf{X}_{ik})^2}{V(\mathbf{a}^T\mathbf{X}_{ik})} \right)$$

In general, as long as $(\mathbf{a}\mathbf{X}_{ik} - E(\mathbf{a}^T\mathbf{X}_{ik}))^2 \xrightarrow{p} V(\mathbf{a}^T\mathbf{X}_{ik})$, $V(Y_{ik}^\star|\mathbf{a}^T\mathbf{X}_{ik})_T \xrightarrow{p} V(Y_{ik}^\star|\mathbf{a}^T\mathbf{X}_{ik})_N$ and

the optimal LCB under both distributions is the same. Notably, this result requires that $v > 2$ else variances are undefined for the multivariate T-distribution. See the Appendix D.2 for details.

In practice, the distribution of the data is unknown and the optimal LCB can be estimated within the context of a loss function or regression model. However, there are certain plausible covariance scenarios in which certain LCB models are optimal. Consider the following covariance expression for a variation of the LCB baseline model where we allow the LCB regression parameter (β) to vary by sequence:

$$\text{cov}(Y_{ik}^{\star}, \mathbf{a}^T\mathbf{X}_{ik}) = \text{cov}(\mathbf{W}_{ik}\boldsymbol{\gamma}, \mathbf{a}^T\mathbf{X}_{ik}) + \beta_i \text{cov}(\mathbf{a}^T\mathbf{X}_{ik}, \mathbf{a}^T\mathbf{X}_{ik}) + \text{cov}(\epsilon_{ik}, \mathbf{a}^T\mathbf{X}_{ik}) \tag{4.11}$$

Given that subjects are randomized, the crossover design parameters ($\boldsymbol{\gamma}$) and the baseline measurements are independent, thus $\text{cov}(\mathbf{W}_{ik}\boldsymbol{\gamma}, \mathbf{a}^T\mathbf{X}_{ik}) = 0$. Under standard linear model assumptions, we also assume that $\text{cov}(\epsilon_{ik}, \mathbf{a}^T\mathbf{X}_{ik}) = 0$. It then follows that:

$$\beta_i = \frac{\text{cov}(Y_{ik}^{\star}, \mathbf{a}^T\mathbf{X}_{ik})}{V(\mathbf{a}^T\mathbf{X}_{ik})} \tag{4.12}$$

Let us consider these results for three general covariance scenarios which correspond to previously discussed covariance structures: Compound Symmetry (CS), Equipredictability (EP), and AR(1). Recall that $Y_{ik}^{\star} = \mathbf{b}_i\mathbf{Y}_{ik}$. First, if $\text{cov}(Y_{ik}^{\star}, \mathbf{a}^T\mathbf{X}_{ik}) = 0$ for all $\mathbf{a}^T$, then $\beta_i = 0$ and the inclusion of any LCB will not add any information but will add an extra degree of freedom to a hypothesis test. Thus, the optimal LCB uses no baseline covariate. This is the case when the underlying covariance of the baselines and outcomes is Compound Symmetry (CS). This result was previously shown under a normality assumption (Jemielita, Putt, and Mehrotra, 2016). Second, assume that $\text{cov}(X_{ijk}, Y_{ijk}) = \sigma_1^2\rho_1^{\star}$ for $j = j$ and $\text{cov}(X_{ijk}, Y_{ij'k}) = \sigma_2^2\rho_2^{\star}$ for $j \neq j'$. This plausible covariance assumption corresponds to the Equipredictability (EP) covariance, as well as the six-parameter covariance structure discussed by Kenward and Roger (Kenward and Roger, 2010). In this case, $\beta = \beta_i$ for all i and the optimal LCB is $\mathbf{b}_i\mathbf{X}_{ik}$ or XDIFF. A proof of this appears in Appendix D.3. For the 2×2 design, XDIFF = $X_{i1k} - X_{i2k}$. This result was previously shown under a normality assumption (Jemielita, Putt, and Mehrotra, 2016). Lastly, for other covariances, such as Unstructured or AR(1), there may be potential efficiency gains by estimating the optimal LCB directly from Equation 4.6.

Overall, these three scenarios cover a broad range of potential covariance structures. Further, this

illustrates the importance of the joint covariance of the baselines and outcomes in determining the optimal LCB for a linear model. Notably, in this section the LCB regression parameter was allowed to vary by sequence, thus accounting for the fact that cov$(Y_{ik}^{\star}, \mathbf{a}^T\mathbf{X}_{ik})$ may vary by sequence. In practice, a common regression parameter is used to reduce the number of model parameters and increase efficiency. For 2-period designs, since $Y_{i1k} - Y_{i2k}$ is the only contrast used for estimation, the LCB regression parameter is constant across all sequences. These LCB models are summarized in Table 4.1 with the additional point that OLS should be used for normal data and R-estimation should be used for other symmetric distributions.

Table 4.1: Crossover Design Optimal Modeling Strategies

		Baseline Covariate		
Distribution	**Estimation**	cov$(Y_{ik}^{\star}, \mathbf{a}^T\mathbf{X}_{ik}) = 0$	UN/AR(1)	Otherwise
Normal	OLS	No Baselines	$\mathbf{a}_{\star}^T\boldsymbol{X}_{ik}$	XDIFF ($\mathbf{a}^T\mathbf{X}_{ik} = \mathbf{b}_i\mathbf{X}_{ik}$)
Non-Normal	R-Estimation	No Baselines	$\mathbf{a}_{\star}^T\boldsymbol{X}_{ik}$	XDIFF ($\mathbf{a}^T\mathbf{X}_{ik} = \mathbf{b}_i\mathbf{X}_{ik}$)

Notes: $Y_{ik}^{\star} = \mathbf{b}_i\mathbf{Y}_{ik}$, or the chosen within-subject contrast. XDIFF uses the same contrast corresponding to the outcomes. XDIFF is the optimal LCB under an Equipredictability (EP) covariance. Ex: $Y_{i1k} - Y_{i3k}$ uses $X_{i1k} - X_{i3k}$. $\mathbf{a}_{\star}^T\boldsymbol{X}_{ik}$ is estimated directly through Equation 4.6.

4.5. Model Selection and Inference: A Bootstrap Approach

For the general LCB model (Equations 4.5, 4.6), the optimal LCB largely depends on the joint covariance of the baselines and outcomes. Further, certain loss functions should be used for different distributions of data. For example, if the joint distribution of the measurements is normal with a CS covariance structure, OLS should be used without including any baseline covariate (Jemielita, Putt, and Mehrotra, 2016; Mehrotra, 2014). However, the joint distribution is unknown and we can never know the true best approach. We search for an approach to adaptively choose the best fitting model among a set of models which are efficient under a range of conditions. With the aim of increasing the efficiency of a treatment effect estimate, our approach picks the model which yields the lowest p-value for the desired treatment effect estimate. A naive approach would be then to use the original SE and p-value from the chosen model for inference. However, this does not account for model selection and inference through the original p-value or confidence interval (CI) is not valid (Efron, 2014; Hurvich and Tsai, 1990). Here we use bootstrapping to account for the model selection and and obtain valid SEs, CIs, and p-values (Efron, 1979, 1987, 2014; Hesterberg, 2015).

Our 'Min-P' approach first fits a number of different baseline models, for example XDIFF (OLS) and XDIFF (R-est) from Table 4.1. Under the null hypothesis of no treatment effect, the model with the smallest p-value and the associated treatment effect estimate is chosen. The data are then resampled through nonparametric bootstrapping and within each resampled data set, the Min-P procedure is repeated. The bootstrap Min-P estimates are then used to obtain bootstrap SEs, CIs, and p-values. Bootstrap resamples are generated by resampling subjects with replacement within each sequence group. The study design determines the sequence groups (Section 4.6).

For the original sample, define the Min-P estimate and p-value as $\hat{\theta}$ and p_0. For each bootstrap resample $b = 1, ..., B$, calculate the Min-P estimate $\hat{\theta}_b$. The bootstrap SE and bootstrap CI are constructed based on the empirical distribution of the bootstrap Min-P estimates. The smoothed or bagged bootstrap estimate and bootstrap SE are defined respectively as:

$$\hat{\theta}_S = \frac{1}{B}\sum_{1}^{B}\hat{\theta}_b \tag{4.13}$$

$$SE(\hat{\theta})_B = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}_b - \hat{\theta}_S)^2} \tag{4.14}$$

Define the nonparametric cumulative distribution function as $\hat{G}(s) = \#\{\hat{\theta}_b < s\}/B$. Then the bootstrap percentile CI is defined as:

$$CI_{pct} = [\hat{G}^{-1}(\alpha/2), \hat{G}^{-1}(1-\alpha/2)] \tag{4.15}$$

Note that for the bootstrap percentile CI, $\hat{G}^{-1}(\alpha/2)$ is simply the $(\alpha/2)$-quantile, for example 2.5%, of the empirical bootstrap distribution of $\hat{\theta}_b$. To obtain a bootstrap p-value, resampling must be done under the null hypothesis. For each bootstrap resample, the null hypothesis is generated by randomly shuffling the treatment assignment. The treatment assignment varies by design (Section 4.6). In each null simulated bootstrap resample, calculate the Min-P p-value $p_b^\star$. The bootstrap p-value is then calculated as:

$$p^\star = \frac{1 + (\#p_b^\star \leq p_0)}{1 + B} \tag{4.16}$$

This formula reflects the fact that the original data set is a possible bootstrap resample and the

original p-value satisfies $p_0 \leq p_0$. Further, the addition of one on the numerator and denominator prevents reporting a p-value of zero.

Improvements can be made to this standard bootstrap approach. First, bootstrap standard errors can be improved by using the standard error of the smoothed bootstrap estimator, $SE(\hat{\theta}_S)$. Efron showed analytically that SE($\hat{\theta}_S$) is more efficient than SE($\hat{\theta}$)$_B$ (Efron, 2014). Generally, using a smoothed or bagged estimate reduces the variability (Friedman and Hall, 2007). Second, Bias-Corrected and Accelerated (BCa) bootstrap CIs are a better option than the simple percentile CI. The BCa improves on the percentile approach by correcting for potential parameter bias and acceleration (Diciccio and Tibshirani, 1987; Efron, 1987). The acceleration describes how much the standard error of a parameter estimate changes with respect to changes in the parameter estimate. The BCa CI is second order accurate, meaning that coverage probabilities differ from the nominal values by $O(n^{-1})$ (Diciccio and Tibshirani, 1987). The percentile method is first order accurate, $O(n^{-1/2})$. Thus, the BCa CI will converge to the nominal coverage faster than the percentile method. Third, for small sample sizes, bootstrap CIs tend to be too narrow (Hesterberg, 2015). To correct this, Hesterberg proposed the Expanded Interval, which adjusts the critical value α. Further details on these improvements can be found in the Appendix D.4.

Overall, bootstrapping allows us to obtain valid inference under a model selection procedure. One minor point is that inference obtained through CIs and bootstrap p-values will not have an exact 1-1 correspondence. This is because the bootstrap p-values require resampling under the null hypothesis while bootstrap CIs do not. However, our empirical simulation results showed an approximate 1-1 correspondence between inference through p-values versus CIs.

4.6. Application of Methods

Table 4.1 illustrates that there are six general scenarios with different preferred analytic approaches. In practice, we are ignorant of the distribution as well as the covariance structure and thus cannot know the best analytic approach. The Min-P method offers a solution. Based on Table 4.1, three versions of the Min-P method are considered:

1. **Min-P1:** Use XDIFF as the LCB and choose between OLS and R-estimation; 2 models
2. **Min-P2:** Choose between No Baselines and XDIFF (R-Est, OLS); 4 models

3. **Min-P3:** Choose between No Baselines, XDIFF, and $\mathbf{a}_\star^T \boldsymbol{X}_{ik}$ (R-Est, OLS); 6 models

Min-P1, Min-P2, and Min-P3 are designed to be efficient under both normal and non-normally distributed data with bootstrap resampling to obtain valid inference.

We benchmark our methods to previously proposed nonparametric methods, both with and without baseline adjustment. For unadjusted nonparametric methods, both the Wilcoxon rank sum test and Ohrvik's aligned rank test are considered. The Wilcoxon rank sum test is used for the 2×2 design and involves ranking the within-subject contrasts between period 1 and period 2 ($R(Y_{i1k} - Y_{i2k})$) (Koch, 1972; Putt and Chinchilli, 2004). Ohrvik's aligned rank test involves first removing period effects from the outcomes and then ranking the within-subject contrasts corresponding to a desired treatment effect (Ohrvik, 1998; Putt and Chinchilli, 2004). While a few nonparametric baseline models have been discussed in the literature, we illustrate Rank ANCOVA (Lavange and Koch, 2006; Quade, 1967) with XDIFF as a covariate. This approach regresses the ranks of relevant within-subject contrasts ($R(\mathbf{b}_i \mathbf{Y}_{ik})$) against the ranks of XDIFF ($R(\mathbf{b}_i \mathbf{X}_{ik})$). The residuals are then used in a Wilcoxon rank sum test. Of note, Rank ANCOVA does not produce treatment estimates or standard errors and is purely used for hypothesis testing. While Tsai and Patel suggested using robust regression models, for example M-estimation (Huber, 1964), to estimate the residuals, this approach performed similarly to Rank ANCOVA based on empirical simulations. Further, compared to simply using the unadjusted outcomes, using change scores in a Wilcoxon rank sum test yielded similar or worse efficiency. This approach is not presented in the data examples or simulations.

The methods considered here are described in Table 4.2. Note that the LCB model, where the baseline covariate ($\mathbf{a}_\star^T \mathbf{X}_{ik}$) is estimated, requires bootstrap resampling for valid inference. Another benchmark, Change from Baseline (CFB), is also described. CFB models the change scores $Y_{ijk} - X_{ijk}$ in a mixed model framework. Generalized least squares (GLS) and restricted maximum likelihood (REML) are used to estimate treatment parameters. A Kenward Roger degree of freedom adjustment is used to properly account for the estimation of the covariance (Kenward and Roger, 1997). This is a commonly used model for crossover designs. Finally, our methods are detailed for the 2×2 and 3×3 designs. The hypotheses of interest are:

$$\begin{aligned} H_0 &: \tau_A - \tau_B = 0 \\ H_A &: \tau_A - \tau_B \neq 0 \end{aligned}$$

Table 4.2: Parametric and Nonparametric Crossover Models

Method	Outcomes	Covariate(s)	Estimation
Wilcoxon Test	$R(Y_{i1k} - Y_{i2k})$	None	Rank-Based: 2×2 design
Rank ANCOVA	$R(\mathbf{b}_i\mathbf{Y}_{ik})$	$R(\mathbf{b}_i\mathbf{X}_{ik})$	Wilcoxon test on residuals
Ohrvik	Aligned Outcomes	None	Rank-Based: Higher order designs
CFB	$Y_{ijk} - X_{ijk}$	None	GLS, REML, KR DF
No Baselines	$\mathbf{b}_i\mathbf{Y}_{ik}$	None	OLS or R-Est
XDIFF	$\mathbf{b}_i\mathbf{Y}_{ik}$	$\mathbf{a}^T\mathbf{X}_{ik} = \mathbf{b}_i\mathbf{X}_{ik}$	OLS or R-Est
LCB	$\mathbf{b}_i\mathbf{Y}_{ik}$	Estimate $\mathbf{a}_\star^T\mathbf{X}_{ik}$	OLS or R-Est; Bootstrap
Min-P1	$\mathbf{b}_i\mathbf{Y}_{ik}$	$\mathbf{a}^T\mathbf{X}_{ik} = \mathbf{b}_i\mathbf{X}_{ik}$	Data-Driven: OLS vs R-Est; Bootstrap
Min-P2	$\mathbf{b}_i\mathbf{Y}_{ik}$	(1) $\mathbf{a}^T\mathbf{X}_{ik} = \mathbf{b}_i\mathbf{X}_{ik}$ or (2) None	Data-Driven: OLS vs R-Est; Bootstrap
Min-P3	$\mathbf{b}_i\mathbf{Y}_{ik}$	(1) $\mathbf{a}^T\mathbf{X}_{ik} = \mathbf{b}_i\mathbf{X}_{ik}$ or (2) None or (3) $\mathbf{a}_\star^T\mathbf{X}_{ik}$	Data-Driven: OLS vs R-Est; Bootstrap

Notes: $R(u)$=ranks. $\mathbf{b}_i\mathbf{Y}_{ik}$ are the chosen within-subject contrasts. Estimation refers to estimation methods used (See Section 4.3). Section 4.4 discusses when certain LCB models are optimal. Section 4.5 discusses the Min-P methods. Section 4.6 details these crossover models for the 2×2 and 3×3 designs. KR DF=Kenward Roger Degrees of freedom. Bootstrap indicates that resampling is needed for valid inference.

4.6.1. 2×2 *Design*

The 2×2 design is a two-period two-treatment design with two sequences such that $j = 1, 2$, $d = A, B$, and $i = AB, BA$. The LCB regression model from (4.5) can be written as:

$$Y_{i1k} - Y_{i2k} = \mu + \delta W_i + \beta \boldsymbol{a}^T \boldsymbol{X}_{ik} + \epsilon_{ik} \tag{4.17}$$

where $W_i = 1$ for sequence AB and 0 for sequence BA. Then $\frac{1}{2}\delta = \tau_A - \tau_B$, our desired treatment effect estimate. Since all outcomes are $Y_{i1k} - Y_{i2k}$, XDIFF = $X_{i1k} - X_{i2k}$. For this 2-period design, there little to be gained by estimating $\mathbf{a}_\star^T\mathbf{X}_{ik}$ (Jemielita, Putt, and Mehrotra, 2016). Given this, the LCB models and Min-P3 are not considered for the 2×2 design. For Min-P1 and Min-P2, bootstrap resampling is done separately within sequences AB and BA. For bootstrap p-values, the null hypothesis is simulated by randomly shuffling the sequence group assignment (W_i).

4.6.2. 3×3 *Design*

The 3×3 design is a three-period three-treatment design with six sequences such that $j = 1, 2, 3$, $d = A, B, C$ and $i = ABC, BAC, CAB, CBA, ACB, BCA$. For this design, it is advantageous to first "align" the within-subject contrasts such that period effects are eliminated. This was motivated

by the Orhvik's aligned rank test, which was shown to be quite powerful under the 3×3 design (Ohrvik, 1998; Putt and Chinchilli, 2004). Accordingly, with the goal of estimating $\tau_A - \tau_B$, the within-subject contrasts and "aligned" within-subject contrasts, along with XDIFF, are as follows:

Sequence	WS Contrast	Aligned WS Contrast	XDIFF
ABC	$E(Y_1 - Y_2) = \tau_A - \tau_B + (\pi_1 - \pi_2)$	$Y_1 - Y_2 - \widehat{\pi_1 - \pi_2}$	$X_1 - X_2$
BAC	$E(Y_1 - Y_2) = \tau_B - \tau_A + (\pi_1 - \pi_2)$	$Y_1 - Y_2 - \widehat{\pi_1 - \pi_2}$	$X_1 - X_2$
CAB	$E(Y_2 - Y_3) = \tau_A - \tau_B + (\pi_2 - \pi_3)$	$Y_2 - Y_3 - \widehat{\pi_2 - \pi_3}$	$X_2 - X_3$
CBA	$E(Y_2 - Y_3) = \tau_B - \tau_A + (\pi_2 - \pi_3)$	$Y_2 - Y_3 - \widehat{\pi_2 - \pi_3}$	$X_2 - X_3$
ACB	$E(Y_1 - Y_3) = \tau_A - \tau_B + (\pi_1 - \pi_3)$	$Y_1 - Y_3 - \widehat{\pi_1 - \pi_3}$	$X_1 - X_3$
BCA	$E(Y_1 - Y_3) = \tau_B - \tau_A + (\pi_1 - \pi_3)$	$Y_1 - Y_3 - \widehat{\pi_1 - \pi_3}$	$X_1 - X_3$

where $\frac{1}{6}\sum_i^s \frac{1}{n_i}\sum_k^{n_i}(Y_{ijk} - Y_{ij'k}) = \widehat{\pi_j - \pi_{j'}}$. Using these aligned contrasts ($Z^\star_{ik}$), our regression model becomes:

$$Z^\star_{ik} = \mu + \delta W_i + \beta \boldsymbol{a}^T \boldsymbol{X}_{ik} + \epsilon_{ik} \tag{4.18}$$

where $W_i = 1$ for sequences ABC/CAB/ACB (A is before B) and 0 otherwise (B is before A). Again, $\frac{1}{2}\delta = \tau_A - \tau_B$. For this higher order design, estimating $\mathbf{a}^T_\star \mathbf{X}_{ik}$ can lead to power gains and the LCB models (OLS, R-est) along with Min-P1, Min-P2, and Min-P3 are all considered. Bootstrap resampling is done separately within sequences ABC/CAB/ACB and sequences BAC/CBA/BCA after the contrasts have been aligned. This group resampling avoids resampling within each sequence, which could be unfeasible for low sample sizes. Further, this simplification is only possible with the aligned model since this alignment guarantees equal mean models within each group (ABC/CAB/ACB vs BAC/CBA/BCA). For bootstrap p-values, the null hypothesis is generated by shuffling the sequence group assignment (W_i). While this has not been proposed in the literature, we use Rank ANCOVA with aligned within-subject contrasts and XDIFF as the covariate. Finally, we illustrate R-estimation code applicable for the 2×2 and 3×3 design. Note that YDIFF refers to the within-subject contrasts, $Y_{i1k} - Y_{i2k}$ for the 2×2 or the aligned contrasts for the 3×3, delta corresponds to δ in (4.17) or (4.18), and LCB refers to $\mathbf{aX}_{ik}$. The data is assumed to be in wide format with one row per subject. For a dispersion test, it is recommended to use the fitted data from the reduced model as an initial fit for the full model.

```
#### R-estimation Linear Model: Treatment A vs B #####
reduced = rfit(YDIFF~delta+LCB, data=wide) ## Fit reduced model (for dispersion test)
full = rfit(YDIFF~delta+LCB, data=wide, yhat0=reduced$fitted) ##Fit full model
trtAB_est = 0.5*full$coefficients[2] ## Treatment estimate
trtAB_SE = sqrt(0.5^2 * vcov(full)[2,2] ) ## Asymptotic Standard Error
trtAB_pval = drop.test(full,red)$p.value ## Dispersion test p-value
```

4.7. Simulations

Simulation studies for the 2×2 and 3×3 were designed to answer the following questions: (1) Do the optimal LCBs estimated under OLS and R-estimation behave as expected? (2) How do the R-estimation baseline models compare to previously recommended baseline adjusted or unadjusted nonparametric methods? (3) How do the various baseline models (Table 4.1), fit by either OLS or R-estimation, perform for different distributions? (4) How efficient are the model selection based Min-P methods?

For both the 2×2 and 3×3 designs, we simulated 5,000 trials for a variety of scenarios. The simulation scenarios were defined by the hypotheses (null, $\tau_A - \tau_B = 0$; alternative, $\tau_A - \tau_B \neq 0$), the distribution (Normal or T), Σ (CS, EP, AR(1)), and sample sizes. Hypothesis tests were used with a nominal type I error rate of 0.05. Response vectors for each subject within each simulated trial were generated from either a multivariate normal or a multivariate T-distribution with 3 degrees of freedom, which simulates a heavy tailed distribution. In either case, covariance structures CS, EP, and AR(1) were considered. Throughout, we assumed a common variance $\sigma = 1$. For all designs and distributions: under CS, $\rho = 0.6$; under EP, $\rho_1 = 0.70$, $\rho_2 = 0.60$, $\rho_3 = 0.50$. Under AR(1), $\rho = 0.739$ for the 2×2 design while $\rho = 0.789$ for the 3×3 design. These are the same covariance settings as before (Section 2.7).

For the 2×2 and 3×3 design, response vectors were generated with respect to the previously defined mean models (4.2,4.3). For the 2×2 design, setting $\zeta_1 = \pi_1 = 0$, $\zeta_2 = \pi_2 = 1$, and $\mu = 6$, the response vector was generated such that $E[(X_1, Y_1, X_2, Y_2)^T] = (0, 6 + \tau_{d[i,1]}, 1, 7 + \tau_{d[i,2]})^T$. For 3×3, setting $\zeta_1 = \pi_1 = 0$, $\zeta_2 = \pi_2 = 1$, $\zeta_3 = \pi_3 = 2$ and $\mu = 6$, the response vector was generated such that $E[(X_1, Y_1, X_2, Y_2, X_3, Y_3)^T] = (0, 6 + \tau_{d[i,1]}, 1, 7 + \tau_{d[i,2]}, 2, 8 + \tau_{d[i,3]})^T$. Under the null, $\tau_A = \tau_B = 0$ (and $\tau_C = \tau_A = 0$ for the 3×3 design), while under the alternative, for

each scenario, $\tau_A - \tau_B$ was fixed such that using the No Baselines OLS model yielded 80% power under the normal distribution (see Appendix B.2). For the 3×3, $\tau_C - \tau_B$ was set equal to $\tau_A - \tau_B$. Estimates of $\tau_A - \tau_B$ were approximately unbiased for all methods under all scenarios (results not shown).

To evaluate the theoretical arguments from Section 4.4, the optimal LCB was estimated under both OLS and R-estimation under EP and AR(1) for both the normal distribution and T-distribution. For the 2×2 design, the absolute value of the median and interquartile range estimates for the ratio of the estimated LCB coefficients ($a_1^\star/a_2^\star$) can found in Table 4.3. Of particular interest was the power and type I error of each method. Power and type I error results for the Min-P methods, which use 1000 bootstrap resamples, are through bootstrap p-values. Inference through bootstrap CIs (percentile or BCa) yielded similar results. For the 2×2 design, Figure 4.1 compares XDIFF (OLS, R-est), the Wilcoxon Rank Sum Test, and Rank ANCOVA (XDIFF as a covariate) while Figure 4.2 compares XDIFF (OLS,R-est) to Min-P1 and Min-P2. For the 3×3 design, Figure 4.3 compares XDIFF (OLS, R-est), Ohrvik's Aligned Rank Test, and Rank ANCOVA (XDIFF as a covariate) while Figure 4.4 compares XDIFF (OLS,R-est) to Min-P1, Min-P2, and Min-P3. With the exception of the No Baseline models, power and type I error results for all considered models (Table 4.2) and designs can be found in Appendix D.5. While optimal under CS, the No Baseline models are only included through Min-P2 and Min-P3. The LCB models and Min-P3 are not considered for the 2×2 design.

4.7.1. 2×2 *Crossover Design: Estimated Optimal LCBs*

Table 4.3 illustrates the absolute value of the estimated median of the ratio of estimated LCB coefficients, $a_1^\star/a_2^\star$. The median was used since ratio estimates in the simulations can have very extreme values. For EP, the optimal LCB is $X_{i1k} - X_{i2k}$ and thus $a_1^\star/a_2^\star = -1$. For AR(1), the optimal LCB under normality is $\rho^{-2}X_{i1k} - X_{i2k}$ and thus $a_1^\star/a_2^\star = -\rho^{-2} = -1.86$; under the T-distribution, the optimal LCB converges to the optimal LCB under normality. In general, OLS and R-estimation both better estimated the optimal LCB as the sample size increased. Under normality, results between OLS and R-estimation were similar. Under the T-distribution, the R-estimation LCB estimates were slightly more precise. Further, we can see that as the sample size increases, the estimated optimal LCB converges to the optimal LCB under a normality assumption for an AR(1) covariance.

Table 4.3: 2×2 Simulations: Estimated LCBs

		OLS		R-estimation	
Distribution	**Sample Size**	EP	AR(1)	EP	AR(1)
Normal	32	0.99	1.62	0.99	1.60
		[0.80, 1.26]	[1.08, 2.54]	[0.79, 1.26]	[1.05, 2.52]
Normal	100	1.00	1.85	1.00	1.85
		[0.89, 1.13]	[1.49, 2.44]	[0.89, 1.13]	[1.48, 2.46]
Normal	200	1.00	1.87	1.00	1.87
		[0.92, 1.09]	[1.61, 2.26]	[0.92, 1.09]	[1.60, 2.27]
T-Dist	32	0.96	1.28	0.98	1.36
		[0.66, 1.34]	[0.56, 2.30]	[0.70, 1.33]	[0.69, 2.38]
T-Dist	100	1.00	1.61	1.00	1.76
		[0.78, 1.25]	[1.05, 2.50]	[0.83, 1.19]	[1.27, 2.55]
T-Dist	200	1.00	1.73	1.00	1.86
		[0.84, 1.19]	[1.27, 2.53]	[0.88, 1.14]	[1.47, 2.46]

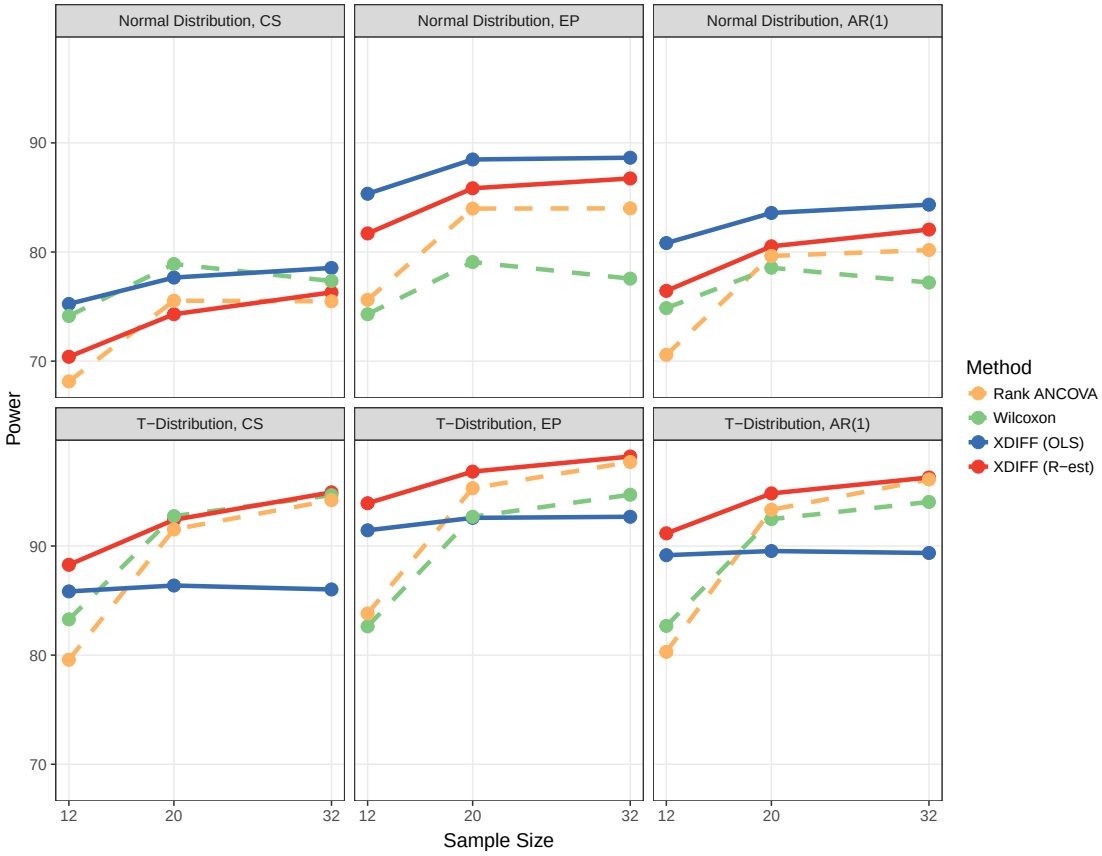
Notes: For each scenario, Entries show the absolute value of the median [IQR] estimated ratio of $a_1^\star/a_2^\star$ based on 5000 simulated data sets. The optimal LCB under EP is $X_1 - X_2$ such $a_1^\star/a_2^\star = -1$. The optimal LCB under AR(1) is $\rho^{-2}X_1 - X_2$ such for $\rho = 0.73$, $a_1^\star/a_2^\star = -\rho^{-2} = -1.86$. OLS/R-estimation columns indicate how the coefficients of optimal LCB were estimated.

4.7.2. 2×2 Crossover Design: Simulation Results

Power and type I error rates for all considered methods can be found in Tables D.1, D.2 in the Appendix while the primary comparisons of interest can be found in Figures 4.1, 4.2. Type I error rates were well controlled under all scenarios. Next, ours and work by others have previously found that change from baseline (CFB) can be highly inefficient (Jemielita, Putt, and Mehrotra, 2016; Kenward and Roger, 2010; Mehrotra, 2014). The simulation results for both distributions confirm this result. Under normality, previous results with OLS suggest (1) ignoring baseline information under CS, (2) that XDIFF is the optimal LCB under EP, and that (3) XDIFF performs reasonably well under AR(1). Indeed, the XDIFF OLS model was our previous recommendation and is our benchmark for the 2×2 design. As expected, we observed a reduction in power when using XDIFF with R estimation versus OLS for the normal, and an increase in power when using XDIFF with R estimation versus OLS for the T-distribution. For the T-distribution, we found that including a baseline covariate in the rank-based method (R-est with XDIFF) improved the power of the hypothesis test for EP and AR(1) compared to a rank-based approach without a covariate, here the Wilcoxon rank sum. For the T-distribution under CS, where we might expect a decrease in efficiency when using a base-

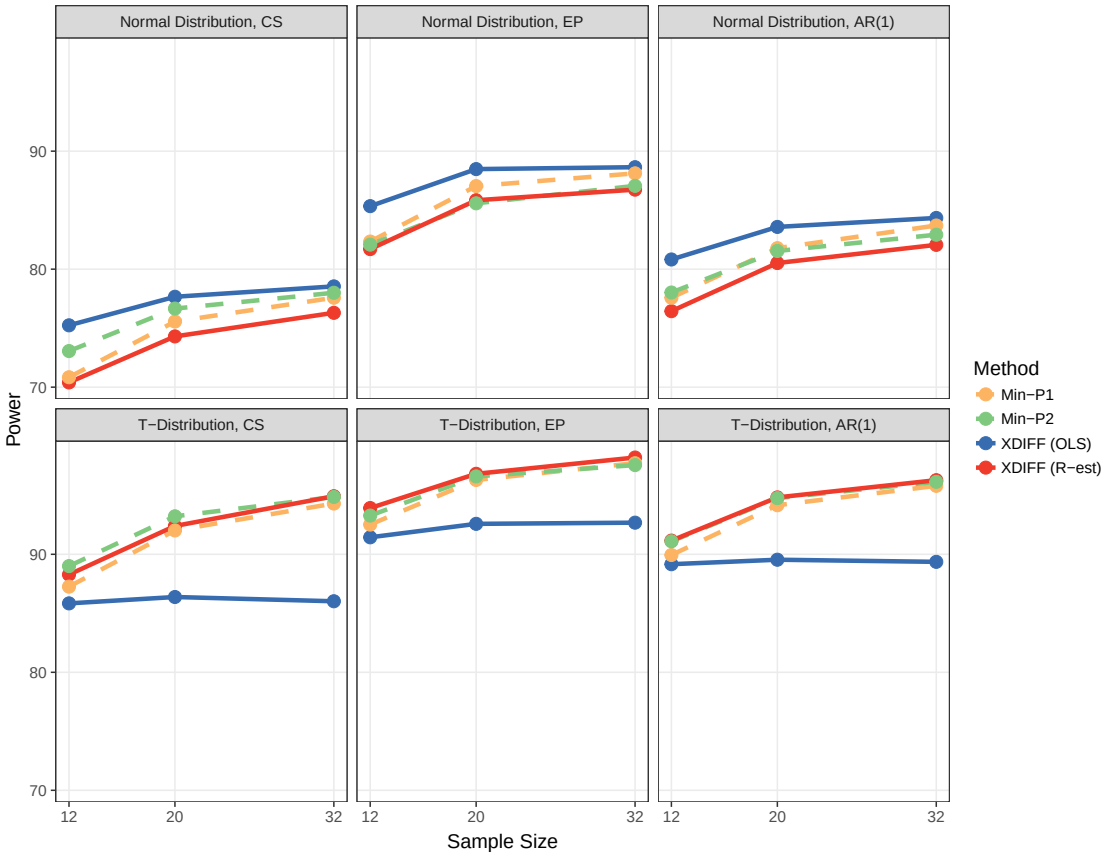
line covariate, we found little difference in the efficiency of rank-based tests with (R-est, XDIFF) and without (Wilcoxon Rank Sum) the covariate for N=20,32. Interestingly, at N=12, XDIFF (R-est) performed better than the Wilcoxon rank sum test. However, this is likely due to the fact that the Wilcoxon rank sum used an exact test. For the normal distribution under EP and AR(1), XDIFF (R-est) consistently yielded better performance than the Wilcoxon Rank sum. Further, we found that Rank ANCOVA with XDIFF as a covariate consistently yielded lower power than XDIFF (R-est). Lastly, as hoped, the Min-P approaches yielded no type I error inflation and offered intermediate power to the optimal approach for each scenario.

Figure 4.1: 2x2 Simulations: Benchmark Comparisons



***Notes**:* XDIFF (OLS) is the most powerful under normality while XDIFF (R-est) is the most powerful under the T-distribution. CS=Compound Symmetry, EP=Equipredictability, AR(1)=Autoregressive(1). Wilcoxon refers to the Wilcoxon rank sum test with no baseline adjustment. Rank ANCOVA regresses the ranks of the outcomes against the ranks of XDIFF and then uses the residuals in a Wilcoxon rank sum test. XDIFF (OLS) uses XDIFF as a covariate in an OLS model. XDIFF (R-est) uses XDIFF as a covariate in a R-estimation model. See Section 4.6 and Table 4.2 for details.

Figure 4.2: 2x2 Simulations: Min-P Comparisons



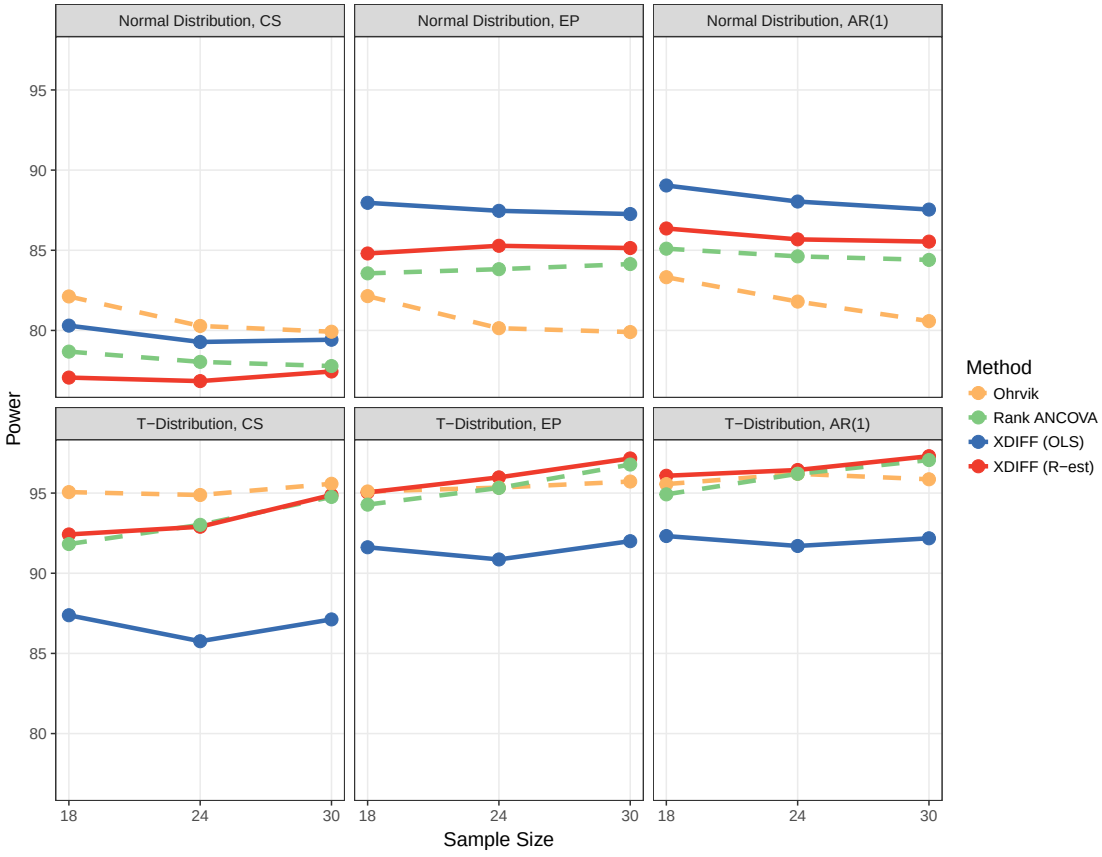
***Notes**:* Under normality, Min-P1 and Min-P2 capture additional power relative to XDIFF (R-est). Under the T-distribution, Min-P1 and Min-P2 are approximately as powerful as XDIFF (R-est). CS=Compound Symmetry, EP=Equipredictability, AR(1)=Autoregressive(1). XDIFF (OLS) uses XDIFF as a covariate in an OLS model. XDIFF (R-est) uses XDIFF as a covariate in a R-estimation model. Min-P1 chooses between XDIFF (OLS, R-est) models. Min-P2 chooses between No Baseline and XDIFF (OLS, R-est) models. See Section 4.6 and Table 4.2 for details.

4.7.3. 3×3 Crossover Design: Simulation Results

Power and type I error rates for all considered methods can be found in Tables D.3, D.4 in the Appendix while the primary comparisons of interest can be found in Figures 4.3, 4.4. For the normal distribution under CS, we observed slight type I error inflation for the XDIFF covariate (both OLS and R-estimation at N=18) and for R-estimation at N=30. The LCB under OLS also yielded some Type I error inflation at N=18,30. Next, the simulation results for both the normal and the T-distribution indicate generally poor performance for the CFB. Consistent with the 2×2 design, we observed a reduction in power when using XDIFF with R estimation versus OLS for the normal, and an increase in power when using XDIFF with R estimation versus OLS for the T-distribution.

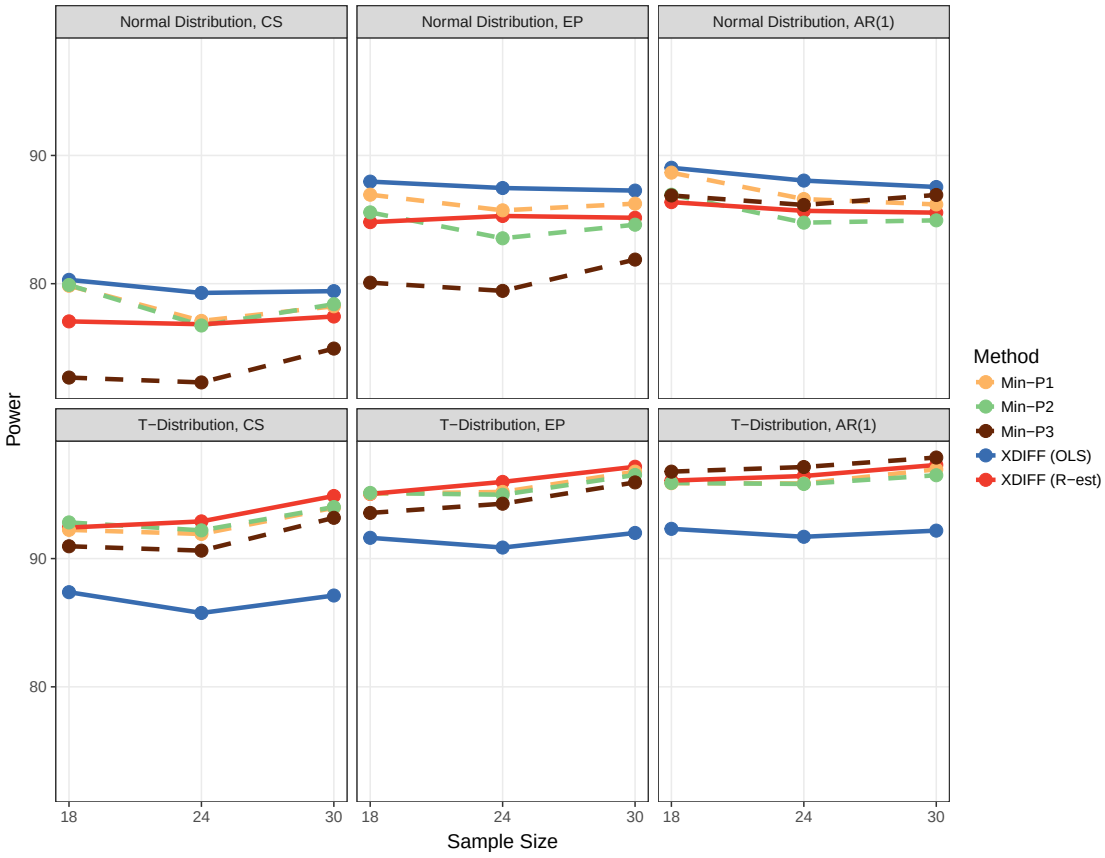
For the T-distribution, we found that including a baseline covariate in the rank-based method (R-est with XDiff) increased the power of the hypothesis test for EP and AR(1) compared to a rank-based approach without a covariate, here the aligned rank test described by Ohrvik (Ohrvik, 1998). For the T-distribution under CS, Ohrvik's aligned rank test showed an increase in efficiency related to XDIFF (R-est). For the normal distribution, XDIFF (R-est) again yielded better performance than Ohrvik's aligned rank test and yielded only slightly lower power to XDIFF (OLS). For the 3x3 design, we found that Rank-based ANCOVA yielded similar or slightly less power than XDIFF (R-est) for the T-distribution. For the normal distribution, Rank-based ANCOVA consistently yielded lower power than XDIFF (R-est). As in the 2×2 design, the Min-P approaches yielded no type I error inflation and offered intermediate power to the optimal approach for each scenario.

Figure 4.3: 3x3 Simulations: Benchmark Comparisons



***Notes**:* XDIFF (OLS) is uniformly the most powerful under normality while XDIFF (R-est) is uniformly the most powerful under the T-distribution.CS=Compound Symmetry, EP=Equipredictability, AR(1)=Autoregressive(1). Ohrvik refers to Ohrvik's aligned rank test with no baseline adjustment. Rank ANCOVA regresses the ranks of the aligned outcomes against the ranks of XDIFF and then uses the residuals in a Wilcoxon rank sum test. XDIFF (OLS) uses XDIFF as a covariate in an OLS model. XDIFF (R-est) uses XDIFF as a covariate in a R-estimation model. See Section 4.6 and Table 4.2 for details.

Figure 4.4: 3x3 Simulations: Min-P Comparisons



***Notes**: Under normality, Min-P1 and Min-P2 capture additional power relative to XDIFF (R-est). Under the T-distribution, Min-P1 and Min-P2 are approximately as powerful as XDIFF (R-est). CS=Compound Symmetry, EP=Equipredictability, AR(1)=Autoregressive(1). XDIFF (OLS) uses XDIFF as a covariate in an OLS model. XDIFF (R-est) uses XDIFF as a covariate in a R-estimation model. Min-P1 chooses between XDIFF (OLS, R-est) models. Min-P2 chooses between No Baselines and XDIFF (OLS, R-est) models. Min-P3 chooses between No Baselines, XDIFF, and LCB (OLS, R-est) models. See Section 4.6 and Table 4.2 for details.*

4.7.4. Simulation Results Summary

Overall, these simulations give credence to previous theoretical findings. Using either OLS or R-estimation, the optimal LCB is well represented by the estimated LCB. Under a normal or T-distribution with identical covariance matrices, the optimal LCB is the same in large samples. For normal data, OLS and R-estimation estimated the optimal LCB with similar accuracy. For T-distributed data, R-estimation estimated the optimal LCB with slightly more precision.

Previously recommended nonparametric methods for the 2×2 and 3×3 were compared to R-estimation baseline models. For the 2×2 design, XDIFF (R-est) consistently outperformed the

Wilcoxon rank-sum test and Rank ANCOVA. For the 3×3 design, XDIFF (R-est) was generally more efficient than Ohrvik's aligned rank test and Rank ANCOVA. Ohrvik's aligned rank test was especially powerful under a CS covariance assumption.

Min-P methods were also considered for both designs. For the 2×2 design, Min-P1 and Min-P2 did similarly, although Min-P2 demonstrated increased efficiency gains under CS. For the 3×3 design, Min-P1 seemed to be the best adaptive method. In contrast to the 2×2 design, Min-P2, which incorporates the no baseline models, showed little efficiency gains under CS. Min-P3, which incorporates six different baseline models, proved to inefficient under CS and EP but slightly more efficient under AR(1). Overall, the proposed data-driven Min-P methods were efficient for a variety of distributions and covariance structures. In both designs, Min-P1 and Min-P2 performed similarly to XDIFF (OLS) under normality and similarly to XDIFF (R-est) under a T-distribution. These data-driven methods also outperformed CFB and previously recommended nonparametric methods.

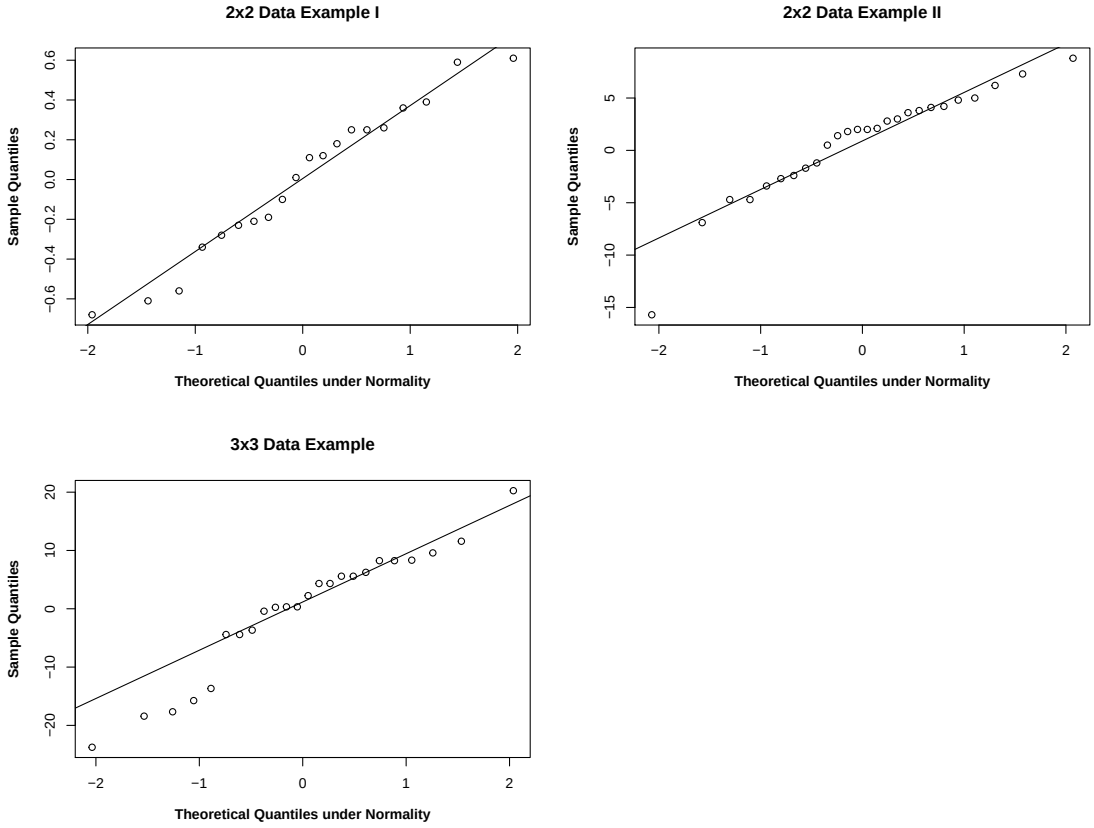
4.8. Real Data Analysis

For each data set, estimates, standard errors, and p-values are provided for the methods described in Table 4.2. Normal Q-Q plots of the relevant within-subject contrasts, $Y_{i1k} - Y_{i2k}$ for the 2×2 and the aligned contrasts for the 3×3, are shown in Figure 4.5. The LCB models and Min-P methods, based on 5000 bootstrap resamples, show smoothed bootstrap standard errors along with bootstrap p-values. Additionally, previous research recommended the XDIFF OLS model for the 2×2 design (Jemielita, Putt, and Mehrotra, 2016; Mehrotra, 2014; Metcalfe, 2010) and an information criteria based adaptive approach for the 3×3 design (Jemielita, Putt, and Mehrotra, 2016). This information criteria approach, which selects an LCB to include as a covariate in an OLS model based on AICC, a small sample correction of AIC (Hurvich and Tsai, 1989), is provided for the 3×3 data set for comparison.

4.8.1. 2×2 Real Data Example I

This is the same example presented in Section 2.6.1. In this example, a biomarker associated with renal function was determined for each of 20 subjects at baseline and after treatment. AICC results indicated that the AR(1) covariance structure was the most likely covariance. This suggests that XDIFF should be used as the baseline covariate. Estimate results are shown in Table 4.4. A Normal

Figure 4.5: Normal Q-Q Plots: Crossover Design Real Data Examples



***Notes**:* Normal Q-Q Plots compare observed sample quantiles to theoretical quantiles under the assumption of normality; departures from the straight line indicate departures from normality. For the 2×2 examples, sample quantiles are of outcomes $Y_{i1k} - Y_{i2k}$. For the 3×3 example, sample quantiles are of aligned within-subject contrasts.

Q-Q plot of the outcomes ($Y_{i1k} - Y_{i2k}$) is shown in the top left corner of Figure 4.5.

Based on Figure 4.5, there does appear to be deviations from normality and nonparametric or robust models may be preferable. This observation holds true, as the R-estimation models do quite well relative to their respective OLS models. Both R-estimation models yield similar results to the Wilcoxon rank sum test and Rank ANCOVA. Min-P1 and Min-P2 both pick the XDIFF (R-est) model and have the smallest standard errors among all considered models. Lastly, Min-P1 and Min-P2 greatly outperform CFB and are competitive with the previously recommended XDIFF OLS model.

Table 4.4: Parametric and Nonparametric Comparisons: 2×2 Real Data Example I (N=20)

Method	LCB	Estimate	SE	p-value
CFB	None	0.235	0.092	0.0201
Wilcoxon Rank Sum	None	0.193	-	0.0376
Rank ANCOVA	$R(X_{i1k} - X_{i2k})$	-	-	0.0126
No Baselines (OLS)	None	0.155	0.079	0.0650
XDIFF (OLS)	$X_{i1k} - X_{i2k}$	0.186	0.073	0.0212
No Baselines (R-est)	None	0.195	0.072	0.0171
XDIFF (R-est)	$X_{i1k} - X_{i2k}$	0.202	0.073	0.0160
Min-P1	$X_{i1k} - X_{i2k}$	0.202	0.064^a	0.0248^b
Min-P2	Data Driven	0.202	0.064^a	0.0338^b

Notes: P-values are based on Wald tests for OLS, dispersion tests for R-est, and Wilcoxon rank sum exact tests for Rank ANCOVA. R-est standard errors are based on Equation 4.9. ^{a}: Bootstrap Smoothed SE. ^{b}: Bootstrap p-values. Rank ANCOVA is used for hypothesis testing.

4.8.2. 2×2 Real Data Example II

In this second 2×2 example, 24-hour mean arterial blood pressure, assessed via ambulatory blood pressure monitoring, was obtained for each of 26 subjects at baseline and after a fixed duration of receiving the assigned treatment for the given period (Example 1, Mehrotra 2014) (Mehrotra, 2014). AICC results indicated that the EP covariance was the most likely covariance. This suggests that XDIFF should be used as the baseline covariate. Results are shown in Table 4.5. A Normal Q-Q plot of the outcomes ($Y_{i1k} - Y_{i2k}$) is shown in the top right corner of Figure 4.5.

In this example, while there appears to be a single extreme value, the outcomes do not seem to deviate strongly from normality. This indicates that while OLS may be reasonable, the extreme value is likely to have an adverse impact on efficiency. In term of standard errors, the best model is No Baselines (R-est). Inference based on the Wiloxon Rank Sum and Rank ANCOVA are similar to the OLS models. Min-P1 picks XDIFF (OLS) while Min-P2 picks No Baselines (R-est). In this case, Min-P2 has a smaller SE than Min-P1, illustrating the fact that a no baseline model is sometimes desirable. Again, in terms of standard errors, Min-P1 and Min-P2 outperform the benchmark CFB and the previously recommended XDIFF (OLS) model.

4.8.3. 3×3 Real Data Example

The 3×3 example is the same as in Section 2.6.2. This study compared effects on heart rate of three treatments; a test drug, a standard drug, and a placebo. Treatments were assigned in the six

Table 4.5: Parametric and Nonparametric Comparisons: 2×2 Real Data Example II (N=24)

Method	LCB	Estimate	SE	p-value
CFB	None	-2.04	0.96	0.045
Wilcoxon Rank Sum	None	-2.00	-	0.045
Rank ANCOVA	$R(X_{i1k} - X_{i2k})$	-	-	0.038
No Baselines (OLS)	None	-1.94	0.95	0.053
XDIFF (OLS)	$X_{i1k} - X_{i2k}$	-1.86	0.87	0.043
No Baselines (R-Est)	None	-2.00	0.81	0.029
XDIFF (R-est)	$X_{i1k} - X_{i2k}$	-1.57	0.82	0.080
Min-P1	$X_{i1k} - X_{i2k}$	-1.86	0.82^a	0.051^b
Min-P2	Data Driven	-2.00	0.75^a	0.048^b

Notes: P-values are based on Wald tests for OLS, dispersion tests for R-est, and Wilcoxon rank sum exact tests for Rank ANCOVA. R-est standard errors are based on Equation (4.9). ^{a}: Bootstrap Smoothed SE. ^{b}: Bootstrap p-values. Rank ANCOVA is used for hypothesis testing.

possible sequences to four patients each. We compare standard to placebo. AICC results indicate that the most likely covariance structure is EP. This suggests that XDIFF is a good covariate choice. Model results are given in Table 4.6. A Normal Q-Q plot of the aligned outcomes is shown in the bottom left corner of Figure 4.5.

Simply looking at the Normal Q-Q plot of the aligned outcomes (Figure 4.5), there appears to be considerable deviation from normality. In agreement with the AICC results, the XDIFF models yield smaller standard errors than the No Baseline models. Relative to the other non-adaptive methods, the LCB models yielded smaller standard errors. Min-P1 and Min-P2 both pick the XDIFF (OLS) model while Min-P3 picks the LCB (OLS) model. Overall, with the exception of the LCB (OLS) model, the Min-P methods yielded the smallest standard errors.

Table 4.6: Parametric and Nonparametric Comparisons: 3×3 Real Data Example (N=24)

Method	LCB	Estimate	SE	p-value
CFB	None	5.67	2.77	0.047
Ohrvik Q	None	4.68	-	0.0347
RANK ANCOVA	$R(\mathbf{b}_i\mathbf{X}_{ik})$	-	-	0.068
No Baselines (OLS)	None	5.12	2.07	0.023
XDIFF (OLS)	XDIFF	5.31	1.95	0.014
LCB (OLS)	$\mathbf{a}_\star^T\mathbf{X}_{ik}$	6.64	1.71^a	0.002^b
No Baselines (R-Est)	None	4.73	2.71^a	0.086
Xdiff (R-est)	XDIFF	4.47	1.89^a	0.024
LCB (R-est)	$\mathbf{a}_\star^T\mathbf{X}_{ik}$	6.77	2.23^a	0.014^b
Min-P1	XDIFF	5.31	1.85^a	0.010^b
Min-P2	Data Driven	5.31	1.86^a	0.015^b
Min-P3	Data Driven	6.64	1.73^a	0.003^b

Notes: AICC favors EP covariance and the information based adaptive method uses OLS with the XDIFF covariate. P-values are based on Wald tests for OLS, dispersion tests for R-est, and Wilcoxon rank sum exact tests for Rank ANCOVA. R-est standard errors are based on Equation (4.9). ^{a}: Bootstrap Smoothed SE. ^{b}: Bootstrap p-values. Rank ANCOVA is used for hypothesis testing.

4.9. Discussion

By incorporating linear combinations of baselines (LCB) in a general loss function, efficient baseline models can be constructed for varying regression models. This allows the use of nonparametric or robust regression models, which may be preferable depending on the underlying distribution of the measurements. In particular, we proposed various baseline models for both OLS and R-estimation regression models. R-estimation is a rank-based regression approach which is robust across a range of distributions. Further, as far as we are aware, this study is the first to directly show that including baseline measurements in a robust analysis of a crossover trial offers improved efficiency over not including baselines. Based on a range of simulation studies and real data examples, R-estimation baseline models were more efficient than previously recommended nonparametric baseline adjusted or unadjusted models.

For baseline utilization in crossover designs, R-estimation models have a number of advantages. First, relative to OLS, R-estimation models yields 95% efficiency under normality (Hettmansperger and Mckean, 2010). Notably, this property is shared among other nonparametric alternatives, such as the Wilcoxon rank sum. Second, R-estimation has good small sample properties (Mckean and Sheather, 1991). As a frame of reference, M-estimation is another popular robust regression model

with similar advantages as R-estimation (Huber, 1964). Despite this, for small sample sizes, variance estimates tend to be unreliable (Fox and Weisberg, 2011). This can lead to inflated type I error. Our own empirical studies in crossover designs confirm this observation (results not shown). Third, R-estimation allows both estimation and hypothesis testing. In contrast, most nonparametric alternatives do not easily give SEs, although CIs can be constructed, for example by inverting the hypothesis test. For Rank ANCOVA, estimation procedures are not readily available. Lastly, the R-estimation score function could be chosen to best fit the underlying data. For symmetric data, the Wilcoxon rank sum is recommended (Hettmansperger and Mckean, 2010). Our methods could easily be adapted for use with skewed data where different score functions would better fit the data (Kloke and McKean, 2012).

A Min-P model selection based procedure was also considered. This approach first fits a number of OLS and R-estimation baseline models. The model with the smallest treatment effect p-value is then chosen. To account for the model selection, nonparametric bootstrap is used for inference. While computationally slow, inference through bootstrap p-values and confidence intervals is distribution free. In contrast, test statistics for OLS or R-estimation baseline models usually assume asymptotic normality. Overall, inference through a nonparametric framework is preferable, as this is robust to distributional assumptions and less sensitive to small sample sizes.

The fact that the Min-P method can adaptively pick among different baseline models is quite advantageous. Many have pointed out that potential efficiency gains from including baselines as covariates ultimately depends on the underlying covariance structure of the baselines and outcomes (Jemielita, Putt, and Mehrotra, 2016; Kenward and Roger, 2010; Mehrotra, 2014). Similarly, Tudor and Koch noted that the magnitude of correlation between the baselines and outcomes play an important role in potential efficiency gains for baseline adjustment in nonparametric models (Tudor and Koch, 1994). This was also illustrated in Section 4.4 for the general LCB linear model. However, this covariance structure is never known and thus the best baseline covariate to include is unknown. For higher order designs, we previously used information criteria to pick the baseline covariate and obtain efficient treatment effect estimates (Jemielita, Putt, and Mehrotra, 2016). The newly proposed Min-P method offers several advantages to our original approach. First, the information criteria based approach assumed that the joint measurements of the baselines and outcomes follows a normal distribution. This is inefficient if the distribution is not normal. Second, the

information criteria based approach makes no adjustment for the model selection. In contrast, the Min-P method makes no distributional assumptions and explicitly accounts for the model selection.

Our methods revolved around regression models where the outcomes were within-subject contrasts chosen to target a specific treatment effect. While this was developed for the general crossover design, there may be better approaches for incomplete block designs. Our work with incomplete block design research (Chapter 3) suggests that a mixed model approach, where within-subject and between-subject information is combined, is more effective than a corresponding within-subject model approach. Analagous to a mixed model, Kloke, Mckean, and Rashid extended R-estimation to a linear model with a dependent error structure (Kloke, Mckean, and Rashid, 2009). This approach could be a possible robust alternative to mixed models. Regardless, our developed methods are quite efficient for uniform designs, such as the 2×2 or 3×3 design.

Overall, given a general loss function, an optimal linear combination of baselines can be found. This allows us to consider robust nonparametric regression models in conjunction with efficient baseline utilization. Moreover, by using a Min-P model selection procedure, we are not restricted to pre-specifying a specific model. More importantly, our proposed data-driven methods yield efficient estimates of treatment effect under a variety of covariances and distributions. Further, inference is nonparametric and yields valid inference in small samples. For data that are not normally distributed, R-estimation baseline models are an attractive alternative to previously recommended nonparametric baseline adjusted or unadjusted models. For pratical use, the Min-P models can be utilized. Compared to standard methods, such as CFB or the simple XDIFF OLS model, the Min-P aproach can yield substantial gains in power across a range of covariances and distributions.

CHAPTER 5

CONCLUSION

The goal of this research was to efficiently incorporate baselines into the analysis of a crossover design. While there has been a variety of proposed baseline models in the literature, our methods are the first to explicitly leverage the relationship between the baselines and outcomes to deliver a more efficient treatment effect estimate. Specifically, linear combinations of baselines (LCBs) are incorporated in a regression model such that the efficiency of a pairwise treatment effect estimate is increased. Given the typically small sample sizes of crossover designs, combining all available baseline information into a single metric is especially advantageous since the number of covariates in a regression model can significantly impact the degrees of freedom in the analysis.

Chapter 2 discussed baseline utilization for uniform designs under an assumption of normality. Uniform designs are highly efficient and all estimation of a treatment effect estimate can come from within-subject contrasts. Exploiting this, the optimal LCB minimizes the conditional variance corresponding to the within-subject contrasts related to some pairwise treatment effect estimate. In this setting, we showed that the optimal LCB depends on the joint covariance structure of the baselines and outcomes. In practice, the covariance is unknown and thus an information criteria based adaptive approach was proposed. This adaptive approach selected the LCB baseline model corresponding to the most likely underlying covariance structure. Based on simulation studies and real data examples, this adaptive method proved especially effective in higher order designs, such as the 3×3 and 4×4 design. For the 2×2 design, there was little efficiency gain by adaptively picking the most likely LCB model and the simple XDIFF (OLS) (or $X_A - X_B$) baseline model was efficient across a range of scenarios.

In Chapter 3, under the assumption of normality, various baseline methods are proposed and evaluated for incomplete block crossover designs. For mixed effects models, in which the treatment effect estimate is a weighted function of within-subject and between-subject information, period-specific LCB models were proposed. Relative to commonly used models, such as change from baseline, these proposed models proved highly efficient across a range of covariance structures. Since mixed effects models require estimation of a covariance matrix, we further stressed the importance

of using a Kenward-Roger degrees of freedom adjustment. This adjustment is necessary to maintain the nominal type I error rate and yield valid inference. However, even with this adjustment, the type I error rate is inflated at small sample sizes. To handle this, we proposed a simpler fixed effects analysis involving only within-subject contrasts. In this setting, efficient baseline utilization is identical to the uniform design.

Chapter 4 describes the more general setting where the optimal LCB depends on some known distribution. In a practical setting, the distribution is unknown and the optimal LCB is estimated under some regression model or loss function. For non-normal data, we proposed using R-estimation, a rank-based nonparametric regression model. Relative to previously suggested nonparametric models, such as unadjusted rank-based statistics (Koch, 1972; Ohrvik, 1998) and two-step nonparametric covariance adjustment (Tudor and Koch, 1994), R-estimation baseline models proved efficient across a range of scenarios. As far as we are aware, this was the first study to explicitly demonstrate the advantage of adjusting for baseline covariates in nonparametric models. A data-driven Min-P model selection procedure was also proposed. This approach chooses the best fitting model among a set of OLS and R-estimation baseline models with nonparametric bootstrapping to obtain valid inference. Relative to using only OLS or only R-estimation, this Min-P approach uniformly increased power across a range of distributions.

Overall, we have proposed a wide range of efficient baseline models for crossover designs. A key point has been that efficient baseline utilization requires knowledge of the underlying joint distribution of the baselines and outcomes. In practice, the joint distribution is unknown and we can never know a-priori the best analytical approach. To deal with this, information criteria and Min-P data-driven approaches were developed. The advantage here is that the estimation of a treatment effect is not dependent on some pre-specified individual model, which may be inefficient. Relative to commonly used baseline models, such as change from baseline, our methods consistently demonstrated improved power across a range of crossover designs, covariance structures, and distributions.

APPENDIX A

NOTATION

Table A.1: Notation

Notation	Description
$\mathbf{bY}_{ik} = Y_{iAk} - Y_{iBk}$	Treatment-Ordered Contrast (A vs B)
$\boldsymbol{a}^T\boldsymbol{X}_{ik} = \{\sum a_d X_{idk}$ OR $\sum a_j X_{ijk}\}$	Linear Combination of Baselines (LCB), by period or by treatment
$\boldsymbol{a}_\star^T\boldsymbol{X}_{ik}$	Optimal LCB; sequence invariant
$\boldsymbol{a}_{i\star}^T\boldsymbol{X}_{ik}$	Sequence specific Optimal LCB
$\boldsymbol{a}_{j\star}^T\boldsymbol{X}_{ik}$	Period specific Optimal LCB
$\boldsymbol{W}_{ik} = (X_{i1k}, Y_{i1k}, ..., X_{ipk}, Y_{ipk})^T$	Temporally ordered measurements
$\mathbf{b}_i\boldsymbol{Y}_{ik}$	Period ordered within-subject contrast of interest
Acronyms	
LCB	Linear Combination of Baselines
AICC	Akaike Information Criterion corrected; small sample correction
AR(1) (Σ_{AR})	Autoregressive(1) covariance structure
CS (Σ_{CS})	Compound Symmetry covariance structure
DCS (Σ_{DCS})	Double Compound Symmetry covariance structure
EP (Σ_{EP})	Equipredictability covariance structure
UN (Σ_{UN})	Unstructured covariance structure
OLS	Ordinary Least Squares
GLS	Generalized Least Squares
ANCOVA	Analysis of Covariance
REML	Restricted Maximum Likelihood
KR DF	Kenward Roger Degrees of Freedom
XDIFF	$\mathbf{b}_i\boldsymbol{X}_{ik}$; Same contrast as chosen within-subject contrasts ($\mathbf{b}_i\boldsymbol{Y}_{ik}$)
Chapter 2 Methods	
No Baselines	OLS model 2.16 with no LCB
Change from Baseline (CFB)	Mixed Model with change scores ($Y_{ijk} - X_{ijk}$); Equation 3.9, no LCB
$X_A - X_B$	Fits Equation 2.16 with $X_{iAk} - X_{iBk}$ as the LCB
AR(1) covariate	Fits Equation 2.17 with the LCBs in Table 2.2 under AR(1)
Adaptive	Information Criteria Based Method (Section 2.5.3)
Chapter 3 Methods	
No Baselines	Mixed Model with no LCB; Eqn 3.8
Change from Baseline (CFB)	Mixed Model with change scores ($Y_{ijk} - X_{ijk}$); Equation 3.9, no LCB
$\mathbf{Y}\|\mathbf{X}, \bar{\mathbf{X}}$	Regress outcomes against baselines in mixed model; Equation 3.10
$\mathbf{Y}\|\mathbf{Xa}_\star, \overline{\mathbf{Xa}}_\star$	Regress outcomes against LCBs in mixed model; Equation 3.11
WS No Baselines	Fits WS contrast model (3.18) with no LCB
WS $X_1 - X_2$	Fits WS contrast model (3.18) with $X_1 - X_2$ as covariate
WS Adaptive	WS Model information criteria based approach (Section 3.3.2)
Chapter 4 Methods	
No Baselines (OLS, R-est)	Equation 4.5 with no LCB; OLS or R-estimation
XDIFF (OLS, R-est)	Equation 4.5 with XDIFF covariate; OLS or R-estimation
LCB (OLS, R-est)	Estimate the optimal LCB 4.6; OLS or R-estimation
Min-P1	Choose between XDIFF (OLS, R-est) models
Min-P2	Choose between No Baselines and XDIFF (OLS, R-est) models
Min-P3	Choose between No Baselines, XDIFF, and LCB (OLS, R-est) models

APPENDIX B

CHAPTER 2, UNIFORM DESIGNS

B.1. Optimal LCBs under CS, DCS, EP, and AR(1)

Here we solve for the various optimal LCBs under CS, DCS, EP, and AR(1) for the 2×2, 3×3, and 4×4 designs. Under CS, EP, and DCS, the optimal LCB ($\mathbf{a}_\star^T \mathbf{X}_{ik}$) can be conveniently solved under the treatment ordering. Here:

$$\boldsymbol{a}_\star^T = (a_1^\star, ..., a_Z^\star) = \underset{\boldsymbol{a}^T}{\text{Argmin}} V(Y_{iAk} - Y_{iBk} | \boldsymbol{a}^T \boldsymbol{X}_{ik})$$

Under CS, it follows that $V(Y_{iAk} - Y_{iBk} | \boldsymbol{a}^T \boldsymbol{X}_{ik}) = V(Y_{iAk} - Y_{iBk})$ for all $\mathbf{a}^T$ and all designs. Thus, no LCB can reduce the variance of the desired treatment effect and the optimal LCB under CS for all designs is to use no baseline covariate. Under a temporally related covariance structure, such as AR(1), a period-ordering of the measurements needs to be considered. Consequently, within-subject contrasts differ by sequence/design and so do the optimal LCBs. In general, we solve:

$$\boldsymbol{a}_{i\star}^T = (a_{i1}^\star, ..., a_{ip}^\star) = \underset{\boldsymbol{a}_i^T}{\text{Argmin}} V(\boldsymbol{b}_i^T \boldsymbol{Y}_{ik} | \boldsymbol{a}_i^T \boldsymbol{X}_{ik})$$

B.1.1. 2×2 *Design*

For the 2×2 design, analytical solutions for the optimal LCB can be found under an unstructured covariance. We can then apply these solutions to the various covariance structures. For convenience, let the measurements be ordered temporally and define $V((X_1, Y_1, X_2, Y_2)^T)$ under an unstructured covariance:

$$\Sigma_{UN} = \begin{bmatrix} \sigma_1 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ & \sigma_2 & \sigma_{23} & \sigma_{24} \\ & & \sigma_3 & \sigma_{34} \\ & & & \sigma_4 \end{bmatrix} \tag{B.1}$$

The conditional variance of interest is:

$$V(Y_1 - Y_2|\mathbf{a}^T\mathbf{X}) = V(Y_1 - Y_2) - \frac{(a_1(\sigma_{12} - \sigma_{14}) + a_2(\sigma_{23} - \sigma_{34}))^2}{V(\mathbf{a}^T\mathbf{X})}$$

Note that the right term is always non-negative and $0 < V(Y_1 - Y_2|\mathbf{a}^T\mathbf{X}) \leq V(Y_1 - Y_2)$. Next, take partial derivatives with respect to a_1 and a_2. Accordingly:

$$\begin{aligned}
\frac{\partial V(Y_1 - Y_2|\mathbf{a}^T\mathbf{X})}{\partial a_1} &= -\frac{2(a_1(\sigma_{12} - \sigma_{14}) + a_2(\sigma_{23} - \sigma_{34}))(\sigma_{12} - \sigma_{14})V(\mathbf{a}^T\mathbf{X}) - (a_1(\sigma_{12} - \sigma_{14}) + a_2(\sigma_{23} - \sigma_{34}))^2(2a_1\sigma_1 + 2a_2\sigma_{13})}{V(\mathbf{a}^T\mathbf{X})^2} \\
\frac{\partial V(Y_1 - Y_2|\mathbf{a}^T\mathbf{X})}{\partial a_2} &= -\frac{2(a_1(\sigma_{12} - \sigma_{14}) + a_2(\sigma_{23} - \sigma_{34}))(\sigma_{23} - \sigma_{34})V(\mathbf{a}^T\mathbf{X}) - (a_1(\sigma_{12} - \sigma_{14}) + a_2(\sigma_{23} - \sigma_{34}))^2(2a_2\sigma_3 + 2a_1\sigma_{13})}{V(\mathbf{a}^T\mathbf{X})^2}
\end{aligned}$$

Let $K = -2(a_1(\sigma_{12} - \sigma_{14}) + a_2(\sigma_{23} - \sigma_{34}))/V(\mathbf{a}^T\mathbf{X})^2$. Then:

$$\begin{aligned}
\frac{\partial V(Y_1 - Y_2|\mathbf{a}^T\mathbf{X})}{\partial a_1} &= K\left((\sigma_{12} - \sigma_{14})V(\mathbf{a}^T\mathbf{X}) - (a_1(\sigma_{12} - \sigma_{14}) + a_2(\sigma_{23} - \sigma_{34}))(a_1\sigma_1 + a_2\sigma_{13})\right) \\
&= K\left((\sigma_{12} - \sigma_{14})(a_2^2\sigma_3 + a_1a_2\sigma_{13}) - a_2(\sigma_{23} - \sigma_{34})(a_1\sigma_1 + a_2\sigma_{13})\right) \\
&= Ka_2\left(a_1(\sigma_{13}(\sigma_{12} - \sigma_{14}) - \sigma_1(\sigma_{23} - \sigma_{34})) + a_2(\sigma_3(\sigma_{12} - \sigma_{14}) - \sigma_{13}(\sigma_{23} - \sigma_{34}))\right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial V(Y_1 - Y_2|\mathbf{a}^T\mathbf{X})}{\partial a_2} &= K\left((\sigma_{23} - \sigma_{34})V(\mathbf{a}^T\mathbf{X}) - (a_1(\sigma_{12} - \sigma_{14}) + a_2(\sigma_{23} - \sigma_{34}))(a_2\sigma_3 + a_1\sigma_{13})\right) \\
&= K\left((\sigma_{23} - \sigma_{34})(a_1^2\sigma_1 + a_1a_2\sigma_{13}) - a_1(\sigma_{12} - \sigma_{14})(a_2\sigma_3 + a_1\sigma_{13})\right) \\
&= Ka_1\left(a_1(\sigma_1(\sigma_{23} - \sigma_{34}) - \sigma_{13}(\sigma_{12} - \sigma_{14})) + a_2(\sigma_{13}(\sigma_{23} - \sigma_{34}) - \sigma_3(\sigma_{12} - \sigma_{14}))\right)
\end{aligned}$$

Critical values correspond to the values of $\mathbf{a}^T = (a_1, a_2)$ such that both partial derivatives equals zero. Note that both partials equal 0 when $K = 0$. If $K = 0$, then $a_1 = -(\sigma_{23} - \sigma_{34})$ and $a_2 = \sigma_{12} - \sigma_{14}$ and consequently $V(Y_1 - Y_2|\mathbf{a}^T\mathbf{X}) = V(Y_1 - Y_2)$. Thus, $\mathbf{a}^T = \{a_1 = -(\sigma_{23} - \sigma_{34}), a_2 = \sigma_{12} - \sigma_{14}\}$ maximizes the conditional variance. However, both partials also equal zero for:

$$\begin{aligned}
\Big\{a_1^\star &= \sigma_3(\sigma_{12} - \sigma_{14}) - \sigma_{13}(\sigma_{23} - \sigma_{34}); \quad a_2^\star = \sigma_1(\sigma_{23} - \sigma_{34}) - \sigma_{13}(\sigma_{12} - \sigma_{14})\Big\} \\
\Big\{a_1^\star &= V(X_2)(\mathsf{cov}(X_1, Y_1) - \mathsf{cov}(X_1, Y_2)) - \mathsf{cov}(X_1, X_2)(\mathsf{cov}(X_2, Y_1) - \mathsf{cov}(X_2, Y_2)); \\
a_2^\star &= V(X_1)(\mathsf{cov}(X_2, Y_1) - \mathsf{cov}(X_2, Y_2)) - \mathsf{cov}(X_1, X_2)(\mathsf{cov}(X_1, Y_1) - \mathsf{cov}(X_1, Y_2))\Big\} \quad \text{(B.2)}
\end{aligned}$$

To verify that the above critical values correspond to the minimum of the conditional variance with respect to $\mathbf{a}^T$, second partial derivatives typically need to be taken. However, we know that $V(Y_1 - Y_2|\mathbf{a}^T\mathbf{X}) > 0$ and know that the other critical value, $\mathbf{a}^T = \{a_1 = -(\sigma_{23} - \sigma_{34}), a_2 = \sigma_{12} - \sigma_{14}\}$, corresponds to the maximum. Accordingly, (B.2) must correspond to the minimum and the optimal

LCB under a period-ordering setting is $\mathbf{a}_\star^T\mathbf{X} = a_1^\star X_1 + a_2^\star X_2$. While this derivation was under a period-ordering, the optimal LCB under a treatment ordering is found simply by letting the unstructured covariance (B.1) correspond to $V((X_A, Y_A, X_B, Y_B)^T)$. This yields:

$$\Big\{a_1^\star = V(X_B)(\text{cov}(X_A, Y_A) - \text{cov}(X_A, Y_B)) - \text{cov}(X_A, X_B)(\text{cov}(X_B, Y_A) - \text{cov}(X_B, Y_B));$$
$$a_2^\star = V(X_A)(\text{cov}(X_B, Y_A) - \text{cov}(X_B, Y_B)) - \text{cov}(X_A, X_B)(\text{cov}(X_A, Y_A) - \text{cov}(X_A, Y_B))\Big\}$$

- **Under EP**: Here, $V(Y_A) = V(Y_B) = V(X_A) = V(X_B) = \sigma^2$, $\text{corr}(Y_A, Y_B) = \text{cov}(X_A, X_B) = \rho_2$, $\text{cov}(Y_A, X_A) = \text{cov}(X_B, X_B) = \rho_1$, and $\text{cov}(X_B, Y_A) = \text{cov}(X_A, Y_B) = \rho_3$. Accordingly:

$$\Big\{a_1^\star \;\; = \;\; \sigma^4(\rho_1 - \rho_3) - \sigma^4\rho_2(\rho_3 - \rho_1); \;\; a_2^\star = -\sigma^4(\rho_1 - \rho_3) + \sigma^4\rho_2(\rho_3 - \rho_1)\Big\}$$

  Thus, $a_1^\star = -a_2^\star$ and an optimal LCB is $X_A - X_B$.

- **Under DCS**: Here, $V(Y_A) = V(Y_B) = V(X_A) = V(X_B) = \sigma^2$, $\text{cov}(Y_A, Y_B) = \text{cov}(X_A, X_B) = \text{cov}(X_B, Y_A) = \text{cov}(X_A, Y_B) = \rho_2$, and $\text{cov}(Y_A, X_A) = \text{cov}(X_B, X_B) = \rho_1$. Accordingly:

$$\Big\{a_1^\star \;\; = \;\; \sigma^4(\rho_1 - \rho_2) - \sigma^4\rho_2(\rho_2 - \rho_1); \;\; a_2^\star = -\sigma^4(\rho_1 - \rho_2) + \sigma^4\rho_2(\rho_2 - \rho_1)\Big\}$$

  Thus, $a_1^\star = -a_2^\star$ and an optimal LCB is $X_A - X_B$.

- **AR(1)**: Here, $V(Y_1) = V(Y_2) = V(X_1) = V(X_2) = \sigma^2$, $\text{cov}(X_1, Y_1) = \text{cov}(X_2, Y_1) = \text{cov}(X_2, Y_2) = \rho\sigma^2$, $\text{cov}(X_1, X_2) = \text{cov}(Y_1, Y_2) = \rho^2\sigma^2$, and $\text{cov}(X_1, Y_2) = \rho^3\sigma^2$. Thus:

$$\Big\{a_1^\star \;\; = \;\; \sigma^4(\rho - \rho^3); \;\; a_2^\star = -\sigma^4\rho^2(\rho - \rho^3)\Big\}$$

  Thus, $a_1^\star/a_2^\star = \frac{-1}{\rho^2}$ corresponds to the optimal LCB. Subsequently, $\rho^{-2}X_1 - X_2$ is an optimal LCB under AR(1).

B.1.2. 3×3 *Design*

- **EP**: The conditional variance of interest under EP is

$$V(Y_A - Y_B|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{EP}) = \sigma^2(1 - \rho_1) - \frac{\sigma^4(a_1 - a_2)^2(\rho_1 - \rho_3)^2}{V(\boldsymbol{a}^T\boldsymbol{X})}$$

Next, we take partial derivatives of this conditional variance with respect to $\mathbf{a}^T = (a_1, a_2, a_3)$.

$$\begin{aligned}
\frac{\partial V(Y_A - Y_B|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{EP})}{\partial a_1} &= \frac{-2\sigma^6(\rho_1-\rho_3)^2(a_1-a_2)}{V(\mathbf{a}^T\mathbf{X})^2}[a_3(a_3+3\rho_2 a_2+\rho_1 a_3)+a_2(1+\rho_2)(a_1+a_2)] \\
\frac{\partial V(Y_A - Y_B|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{EP})}{\partial a_2} &= \frac{2\sigma^6(\rho_1-\rho_3)^2(a_1-a_2)}{V(\mathbf{a}^T\mathbf{X})^2}[a_3(a_3+\rho_2 a_2+3\rho_1 a_3)+a_1(1+\rho_2)(a_1+a_2)] \\
\frac{\partial V(Y_A - Y_B|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{EP})}{\partial a_3} &= \frac{2\sigma^6(\rho_1-\rho_3)^2}{V(\mathbf{a}^T\mathbf{X})^2}[(a_1-a_2)^2(a_3+\rho_2(a_1+a_2))]
\end{aligned}$$

Note that there are two critical values that result in each partial equaling zero. The maximum is when $a_1 = a_2$, $a_3 = a_3$ since this reduces $V(Y_A - Y_B|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{EP})$ to $V(Y_A - Y_B|\Sigma_{EP})$. The second critical point is when $a_1 = -a_2$, $a_3 = 0$. Using similar logic as in the 2×2 design, this means that an optimal LCB under EP for a 3×3 design is $\mathbf{a}_\star^T\mathbf{X} = (1, -1, 0) * \mathbf{X} = X_A - X_B$.

- **DCS**: The conditional variance of interest under DCS is

$$V(Y_A - Y_B|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{EP}) = \sigma^2(1-\rho_1) - \frac{\sigma^4(a_1-a_2)^2(\rho_1-\rho_2)^2}{V(\boldsymbol{a}^T\boldsymbol{X})}$$

This conditional variance and the corresponding partial derivatives are almost identical to the EP case. Consequently, the optimal LCB is $X_A - X_B$.

- **AR(1)**: The following conditional variances are of interest

$$\begin{aligned}
V(Y_1 - Y_2|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR}) &= 2\sigma^2(1-\rho^2) - \frac{\sigma^4(a_1-a_3)^2(\rho-\rho^3)^2}{V(\mathbf{a}^T\mathbf{X})} \\
V(Y_2 - Y_3|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR}) &= 2\sigma^2(1-\rho^2) - \frac{\sigma^4(a_1\rho^2+a_2)^2(\rho-\rho^3)^2}{V(\mathbf{a}^T\mathbf{X})} \\
V(Y_1 - Y_3|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR}) &= 2\sigma^2(1-\rho^4) - \frac{\sigma^4(\rho-\rho^3)^2(a_1(1+\rho^2)+a_2-a_3)^2}{V(\mathbf{a}^T\mathbf{X})}
\end{aligned}$$

  - $V(Y_1 - Y_2|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})$ corresponds to sequences ABC/BAC. The corresponding partials with respect to $\mathbf{a}^T$ are:

$$\begin{aligned}
\frac{\partial V(Y_1 - Y_2|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_1} &= \frac{-2\sigma^6(\rho-\rho^3)^2(a_1-a_3)}{V(\mathbf{a}^T\mathbf{X})^2}[a_2(a_2+a_1\rho^2+3a_3\rho^2)+a_3(1+\rho^4)(a_1+a_3)] \\
\frac{\partial V(Y_1 - Y_2|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_2} &= \frac{-2\sigma^6(\rho-\rho^3)^2}{V(\mathbf{a}^T\mathbf{X})^2}[(a_1-a_3)^2(a_2+\rho^2(a_1+a_3)] \\
\frac{\partial V(Y_1 - Y_2|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_3} &= \frac{-2\sigma^6(\rho-\rho^3)^2(a_1-a_3)}{V(\mathbf{a}^T\mathbf{X})^2}[a_2(a_2+3a_1\rho^2+a_3\rho^2)+a_1(1+\rho^4)(a_1+a_3)]
\end{aligned}$$

    All three partials equal zero for (1) $a_1 = a_3$, $a_2 = a_2$ and (2) $a_1 = -a_3$, $a_2 = 0$. The maximum is at $\mathbf{a}^T = (a_1 = a_3, a_2 = a_2)$ since this reduces the conditional variance to

$V(Y_1 - Y_2|\Sigma_{AR})$. The minimum is achieved at $a_1 = -a_3$, $a_2 = 0$. Thus, an optimal LCB for sequences ABC/BAC is $X_1 - X_3$.

– $V(Y_2 - Y_3|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})$ corresponds to sequences CAB/CBA. The corresponding partials with respect to $\mathbf{a}^T$ are:

$$\begin{aligned}
\frac{\partial V(Y_2 - Y_3|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_1} &= \frac{-2\sigma^6(\rho - \rho^3)^2(a_1\rho^2 + a_2)}{V(\mathbf{a}^T\mathbf{X})^2}[a_1(a_2\rho^4 - a_3\rho^6 - a_2) + a_3\rho^2(a_3 + a_2\rho^2)] \\
\frac{\partial V(Y_2 - Y_3|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_2} &= \frac{-2\sigma^6(\rho - \rho^3)^2(a_1\rho^2 + a_2)}{V(\mathbf{a}^T\mathbf{X})^2}[a_1(a_1 + a_3\rho^4 - a_1\rho^4) + a_3(a_3 + a_2\rho^2)] \\
\frac{\partial V(Y_2 - Y_3|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_3} &= \frac{2\sigma^6(\rho - \rho^3)^2(a_1\rho^2 + a_2)^2}{V(\mathbf{a}^T\mathbf{X})^2}[a_1\rho^4 + a_3 + a_2\rho^2]
\end{aligned}$$

All three partials equal zero for (1) $a_1\rho^2 = -a_2$, $a_3 = a_3$ and (2) $a_3 = -a_2\rho^2$, $a_1 = 0$. The first corresponds to the maximum (yields $V(Y_2 - Y_3|\Sigma_{AR})$) while the second corresponds to the minimum. Thus, an optimal LCB for sequences CAB/CBA is $\rho^{-2}X_2 - X_3$.

– $V(Y_1 - Y_3|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})$ corresponds to sequences ACB/BCA. Let:

$$K = \frac{-2\sigma^6(\rho - \rho^3)^2(a_1(1 + \rho^2) + a_2 - a_3)}{V(\mathbf{a}^T\mathbf{X})^2}$$

Then the corresponding partials with respect to $\mathbf{a}^T$ are:

$$\begin{aligned}
\frac{\partial V(Y_1 - Y_3|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_1} &= K[a_3^2(1 + \rho^2 + \rho^4) + a_2(a_2 + 3a_3\rho^2 + a_3\rho^4) + a_1(a_2\rho^2 + a_3\rho^2 + a_3\rho^6 - a_2 + a_3)] \\
\frac{\partial V(Y_1 - Y_3|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_2} &= K[a_1^2(1 - \rho^2 - \rho^4) + a_3^2(1 + \rho^2) + a_1a_3\rho^4 + a_2(a_3\rho^2 - a_1 + a_3)] \\
\frac{\partial V(Y_1 - Y_3|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_3} &= K[a_1^2(1 + \rho^4 + \rho^6) + (1 + \rho^2)(a_2^2 + a_2a_3) + a_1a_2(3\rho^2 + 2\rho^4) + a_1a_3(1 + \rho^2 + \rho^4)]
\end{aligned}$$

Unlike the previous cases, it is not explicitly clear what values of $\mathbf{a}^T = (a_1, a_2, a_3)$ correspond to the minimum of the conditional variance. Using numerical methods (such as PROC NLP or optim), there was a suggestion that $a_1 = a_2$ corresponds to the minimum critical values. Following this idea, set $a = b = 1$ for simplicity. Then, plugging this into $\frac{\partial V(Y_1 - Y_3|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_2}$, we get:

$$a_3 = \frac{-(1 + \rho^2 + \rho^4) - \sqrt{(1 + \rho^2 + \rho^4)^2 + 4\rho^2(1 + \rho^2)^2}}{2(1 + \rho^2)}$$

These values do in fact set the other partials to zero. Further, these values are identical to the values that minimize the conditional variance under numerical methods, giving strong evidence that these values of $\mathbf{a}^T = (a_1, a_2, a_3)$ minimize the conditional variance.

Overall, the optimal LCBs under an AR(1) structure are (1) $X_1 - X_3$ for sequences ABC/BAC; (2) $\rho^{-2}X_2 - X_3$ for sequenes CAB/CBA; (3) $X_1 + X_2 + \frac{-(1+\rho^2+\rho^4)-\sqrt{(1+\rho^2+\rho^4)^2+4\rho^2(1+\rho^2)^2}}{2(1+\rho^2)}X_3$ for sequences ACB/BCA.

B.1.3. 4×4 *Design*

- **EP**: The conditional variance of interest under EP is

$$V(Y_A - Y_B|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{EP}) = \sigma^2(1-\rho_1) - \frac{\sigma^4(a_1-a_2)^2(\rho_1-\rho_3)^2}{V(\boldsymbol{a}^T\boldsymbol{X})}$$

Next, we take partial derivatives of this conditional variance with respect to $\mathbf{a}^T = (a_1, a_2, a_3, a_4)$.

$$\begin{aligned}
\frac{\partial V(Y_A - Y_B|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{EP})}{\partial a_1} &= \frac{-2\sigma^6(\rho_1-\rho_3)^2(a_1-a_2)}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * \\
&\quad (a_3(a_3 + a_1\rho_2 + 3a_2\rho_2 + 2a_4\rho_2) + a_4(a_4 + a_1\rho_2 + 3a_2\rho_2) + a_2(1+\rho_2)(a_1+a_2)) \\
\frac{\partial V(Y_A - Y_B|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{EP})}{\partial a_2} &= \frac{2\sigma^2(\rho_1-\rho_3)^2(a_1-a_2)}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * \\
&\quad (a_3(a_3 + 3a_1\rho_2 + a_2\rho_2 + 2a_4\rho_2) + a_4(a_4 + 3a_1\rho_2 + a_2\rho_2) + a_1(1+\rho_2)(a_1+a_2)) \\
\frac{\partial V(Y_A - Y_B|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{EP})}{\partial a_3} &= \frac{2\sigma^6(\rho_1-\rho_3)^2(a_1-a_2)^2}{V(\boldsymbol{a}^T\boldsymbol{X})^2}(a_3 + a_4\rho_2 + \rho_2(a_1+a_2)) \\
\frac{\partial V(Y_A - Y_B|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{EP})}{\partial a_4} &= \frac{2\sigma^6(\rho_1-\rho_3)^2(a_1-a_2)^2}{V(\boldsymbol{a}^T\boldsymbol{X})^2}(a_4 + a_3\rho_2 + \rho_2(a_1+a_2))
\end{aligned}$$

All partials equal zero for (1) $a_1 = a_2, a_3 = a_3, a_4 = a_4$ and (2) $a_1 = -a_2, a_3 = a_4 = 0$. The maximum corresponds to $\mathbf{a}^T = (a_1 = a_2, a_3 = a_3, a_4 = a_4)$, as this reduces the conditional variance to $V(Y_A - Y_B|\Sigma_{EP})$. The minimum is obtained at $\mathbf{a}^T = (a_1 = -a_2, a_3 = a_4 = 0)$. Thus, an optimal LCB is $X_A - X_B$.

- **DCS**: The conditional variance of interest under DCS is

$$V(Y_A - Y_B|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{EP}) = \sigma^2(1-\rho_1) - \frac{\sigma^4(a_1-a_2)^2(\rho_1-\rho_2)^2}{V(\boldsymbol{a}^T\boldsymbol{X})}$$

Given the almost identical variance formulas, as in the 3×3 design, the optimal LCB under DCS is the same as under EP. Consequently, the optimal LCB is $X_A - X_B$.

- **AR(1)**: The following conditional variances are of interest

$$
\begin{aligned}
V(Y_1 - Y_2|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR(1)}) &= 2\sigma(1-\rho^2) - \frac{\sigma^4(\rho-\rho^3)^2(a_1 - a_3 - \rho^2 a_4)^2}{V(\boldsymbol{a}^T\boldsymbol{X})} \\
V(Y_1 - Y_3|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR(1)}) &= 2\sigma(1-\rho^4) - \frac{\sigma^4((a_1 - a_4)(\rho-\rho^5) + (a_2 - a_3)(\rho-\rho^3))^2}{V(\boldsymbol{a}^T\boldsymbol{X})} \\
V(Y_2 - Y_4|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR(1)}) &= 2\sigma(1-\rho^4) - \frac{\sigma^4((\rho^2 a_1 + a_2)(\rho-\rho^5) + (a_3 - a_4)(\rho-\rho^3))^2}{V(\boldsymbol{a}^T\boldsymbol{X})} \\
V(Y_3 - Y_4|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR(1)}) &= 2\sigma(1-\rho^2) - \frac{\sigma^4(\rho-\rho^3)^2(\rho^4 a_1 + \rho^2 a_2 + a_3)^2}{V(\boldsymbol{a}^T\boldsymbol{X})}
\end{aligned}
$$

  - $V(Y_1 - Y_2|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})$ corresponds to sequences ABCD. The corresponding partials with respect to $\mathbf{a}^T$ are:

$$
\begin{aligned}
\frac{\partial V(Y_1 - Y_2|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR(1)})}{\partial a_1} &= \frac{-2\sigma^6(\rho-\rho^3)^2(a_1 - a_3 - \rho^2 a_4)}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * \\
&\quad (a_2(a_2 + a_1\rho^2 + 3a_3\rho^2 + 3a_4\rho^4) + a_4(a_4 + a_1\rho^6 + 2a_3\rho^2 - a_3\rho^6 - a_1\rho^2 - a_3\rho^6 - a_4\rho^8) + \\
&\quad a_3(1+\rho^4)(a_1 + a_3)) \\
\frac{\partial V(Y_1 - Y_2|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR(1)})}{\partial a_2} &= \frac{2\sigma^6(\rho-\rho^3)^2(a_1 - a_3 - \rho^2 a_4)}{V(\boldsymbol{a}^T\boldsymbol{X})^2}(\rho^2(a_1 + a_3) + a_2 + a_4\rho^4) \\
\frac{\partial V(Y_1 - Y_2|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR(1)})}{\partial a_3} &= \frac{2\sigma^6(\rho-\rho^3)^2(a_1 - a_3 - \rho^2 a_4)}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * \\
&\quad (a_2(a_2 + 3a_1\rho^2 + a_3\rho^2 + a_4\rho^4) + a_4(a_4 + a_1\rho^6 + a_3\rho^2 + a_3\rho^2 + a_1\rho^2 - a_3(\rho^2 + \rho^4)) + \\
&\quad a_1(1+\rho^4)(a_1 + a_3)) \\
\frac{\partial V(Y_1 - Y_2|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR(1)})}{\partial a_4} &= \frac{2\sigma^6(\rho-\rho^3)^2(a_1 - a_3 - \rho^2 a_4)}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * \\
&\quad (a_2(a_2\rho^2 + 3a_1\rho^4 + a_3\rho^4 + a_4\rho^6) + a_4(a_1\rho^8 + a_3\rho^4 + a_1 - a_3) + a_1\rho^2(1+\rho^4)(a_1 + a_3))
\end{aligned}
$$

There are two critical values: (1) $\mathbf{a}^T = (a_1 - a_3 - \rho^2 a_4 = 0)$ and (2) $\mathbf{a}^T = (a_1 = -a_3, a_2 = a_4 = 0)$. The first is the maximum by inpection meaning that $\mathbf{a}^T = (a_1 = -a_3, a_2 = a_4 = 0)$ minimizes the conditional variance. Thus, an optimal LCB for sequence ABCD is $X_1 - X_3$.

– $V(Y_1 - Y_3|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})$ corresponds to sequences BDAC. The corresponding partials are:

$$\begin{aligned}
\frac{\partial V(Y_1 - Y_3|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_1} &= \frac{-2\sigma^6((a_1 - a_4)(\rho - \rho^5) + (a_2 - a_3)(\rho - \rho^3))}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * \\
&\quad [(\rho - \rho^5)(V(\boldsymbol{a}^T\boldsymbol{X}) - (a_1 - a_4)(a_1 + a_2\rho^2 + a_3\rho^4 + a_4\rho^6)) - \\
&\quad (\rho - \rho^3)(a_2 - a_3)(a_1 + a_2\rho^2 + a_3\rho^4 + a_4\rho^6)] \\
\frac{\partial V(Y_1 - Y_3|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_2} &= \frac{-2\sigma^6((a_1 - a_4)(\rho - \rho^5) + (a_2 - a_3)(\rho - \rho^3))}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * \\
&\quad [(\rho - \rho^3)(V(\boldsymbol{a}^T\boldsymbol{X}) - (a_2 - a_3)(a_2 + a_1\rho^2 + a_3\rho^2 + a_4\rho^4)) - \\
&\quad (\rho - \rho^5)(a_1 - a_4)(a_2 + a_1\rho^2 + a_3\rho^2 + a_4\rho^4)] \\
\frac{\partial V(Y_1 - Y_3|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_3} &= \frac{2\sigma^6((a_1 - a_4)(\rho - \rho^5) + (a_2 - a_3)(\rho - \rho^3))}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * \\
&\quad [(\rho - \rho^3)(V(\boldsymbol{a}^T\boldsymbol{X}) + (a_2 - a_3)(a_3 + a_1\rho^4 + a_2\rho^2 + a_4\rho^2)) + \\
&\quad (\rho - \rho^5)(a_1 - a_4)(a_3 + a_1\rho^4 + a_3\rho^2 + a_4\rho^2)] \\
\frac{\partial V(Y_1 - Y_3|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_4} &= \frac{2\sigma^6((a_1 - a_4)(\rho - \rho^5) + (a_2 - a_3)(\rho - \rho^3))}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * \\
&\quad [(\rho - \rho^5)(V(\boldsymbol{a}^T\boldsymbol{X}) + (a_1 - a_4)(a_4 + a_1\rho^6 + a_2\rho^4 + a_3\rho^2)) + \\
&\quad (\rho - \rho^3)(a_2 - a_3)(a_4 + a_1\rho^6 + a_2\rho^4 + a_3\rho^2)]
\end{aligned}$$

The maximum occurs at $\mathbf{a}^T = (a_1 = a_2 = a_3 = a_4)$. From these formulas, it is unclear where the minimum occurs. However, through numerical integration (Newton-Raphson), there is evidence that $\mathbf{a}^T = (a_1 = a_2 = -a_3 = -a_4)$ minimizes the conditional variance. WLOG, let $a_1 = a_2 = 1$, $a_3 = a_4 = -1$. After some algebra:

$$\begin{aligned}
\frac{\partial V(Y_1 - Y_3|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_1} &= \frac{-2\sigma^6((a_1 - a_4)(\rho - \rho^5) + (a_2 - a_3)(\rho - \rho^3))}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * \\
&\quad (2(\rho - \rho^5)(1 - \rho^4) - 2(\rho - \rho^3)(1 + \rho^2 - \rho - \rho^6)) \\
&= \frac{-2\sigma^6((a_1 - a_4)(\rho - \rho^5) + (a_2 - a_3)(\rho - \rho^3))}{V(\boldsymbol{a}^T\boldsymbol{X})^2}(2(\rho - 2\rho^5 + \rho^9 - (\rho - 2\rho^5 + \rho^9))) \\
&= 0 \\
\frac{\partial V(Y_1 - Y_3|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_2} &= \frac{-2\sigma^6((a_1 - a_4)(\rho - \rho^5) + (a_2 - a_3)(\rho - \rho^3))}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * \\
&\quad (2(\rho - \rho^3)(1 + \rho^2 - \rho^4 - \rho^6) - 2(\rho - \rho^5)(1 - \rho^4)) \\
&= 0 \\
\frac{\partial V(Y_1 - Y_3|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_3} &= \frac{2\sigma^6((a_1 - a_4)(\rho - \rho^5) + (a_2 - a_3)(\rho - \rho^3))}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * \\
&\quad (2(\rho - \rho^3)(1 + \rho^2 - \rho^4 - \rho^6) + 2(\rho - \rho^5)(\rho^4 - 1)) \\
&= 0 \\
\frac{\partial V(Y_1 - Y_3|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_4} &= \frac{2\sigma^6((a_1 - a_4)(\rho - \rho^5) + (a_2 - a_3)(\rho - \rho^3))}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * \\
&\quad (2(\rho - \rho^5)(1 - \rho^4) + 2(\rho - \rho^3)(-1 + \rho^6 + \rho^4 - \rho^2)) \\
&= 0
\end{aligned}$$

Thus, for sequence BDAC, an optimal LCB is $X_1 + X_2 - X_3 - X_4$.

– $V(Y_2 - Y_4|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})$ corresponds to sequences CADB. The corresponding partials are:

$$\begin{aligned}
\frac{\partial V(Y_2 - Y_4|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_1} &= \frac{-2\sigma^6((\rho^2 a_1 + a_2)(\rho - \rho^5) + (a_3 - a_4)(\rho - \rho^3))}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * \\
&\quad [(\rho - \rho^5)(\rho^2 V(\boldsymbol{a}^T\boldsymbol{X}) - (\rho^2 a_1 + a_2)(a_1 + a_2\rho^2 + a_3\rho^4 + a_4\rho^6)) - \\
&\quad (\rho - \rho^3)(a_3 - a_4)(a_1 + a_2\rho^2 + a_3\rho^4 + a_4\rho^6)] \\
\frac{\partial V(Y_2 - Y_4|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_2} &= \frac{-2\sigma^6((\rho^2 a_1 + a_2)(\rho - \rho^5) + (a_3 - a_4)(\rho - \rho^3))}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * \\
&\quad [(\rho - \rho^5)(V(\boldsymbol{a}^T\boldsymbol{X}) - (\rho^2 a_1 + a_2)(a_2 + a_1\rho^2 + a_3\rho^2 + a_4\rho^4)) - \\
&\quad (\rho - \rho^3)(a_3 - a_4)(a_2 + a_1\rho^2 + a_3\rho^2 + a_4\rho^4)] \\
\frac{\partial V(Y_2 - Y_4|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_3} &= \frac{-2\sigma^6((\rho^2 a_1 + a_2)(\rho - \rho^5) + (a_3 - a_4)(\rho - \rho^3))}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * \\
&\quad [(\rho - \rho^3)(V(\boldsymbol{a}^T\boldsymbol{X}) - (a_3 - a_4)(a_3 + a_1\rho^4 + a_2\rho^2 + a_4\rho^2)) - \\
&\quad (\rho - \rho^5)(\rho^2 a_1 + a_2)(a_3 + a_1\rho^4 + a_2\rho^2 + a_4\rho^2)] \\
\frac{\partial V(Y_2 - Y_4|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_4} &= \frac{2\sigma^6((\rho^2 a_1 + a_2)(\rho - \rho^5) + (a_3 - a_4)(\rho - \rho^3))}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * \\
&\quad [(\rho - \rho^3)(V(\boldsymbol{a}^T\boldsymbol{X}) + (a_3 - a_4)(a_4 + a_1\rho^6 + a_2\rho^4 + a_3\rho^2)) + \\
&\quad (\rho - \rho^3)(\rho^2 a_1 + a_2)(a_4 + a_1\rho^6 + a_2\rho^4 + a_3\rho^2)]
\end{aligned}$$

Each partial equals zero $\mathbf{a}^T = (\rho^2 a_1 + a_2 = 0, a_3 = a_4)$. This corresponds to the maximum. Again, it's unclear where the minimum occurs. However, numerical integration (Newton-Raphson) indicates that the minimum occurs at $\mathbf{a}^T = (a_1 = 0, a_2 = a_3, a_4 = -(1 + \rho^2)a_3)$. WLOG, let $a_2 = a_3 = 1$, $a_4 = -(1 + \rho^2)$. After some algebra:

$$\begin{aligned}
\frac{\partial V(Y_2 - Y_4|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_1} &= \frac{-2\sigma^6((\rho^2 a_1 + a_2)(\rho - \rho^5) + (a_3 - a_4)(\rho - \rho^3))}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * \\
&\quad ((\rho - \rho^5)(2\rho^2 + \rho^4 - 2\rho^6 - \rho^8) - (\rho - \rho^3)(2\rho^2 + 3\rho^4 - \rho^6 - 3\rho^8 - \rho^{10})) \\
&= 0 \\
\frac{\partial V(Y_2 - Y_4|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_2} &= \frac{-2\sigma^6((\rho^2 a_1 + a_2)(\rho - \rho^5) + (a_3 - a_4)(\rho - \rho^3))}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * \\
&\quad ((\rho - \rho^5)(2 + \rho^2 - 2\rho^4 - \rho^6) - (\rho - \rho^3)(2 + 3\rho^2 - \rho^4 - 3\rho^6 - \rho^8)) \\
&= 0 \\
\frac{\partial V(Y_2 - Y_4|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_3} &= \frac{-2\sigma^6((\rho^2 a_1 + a_2)(\rho - \rho^5) + (a_3 - a_4)(\rho - \rho^3))}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * \\
&\quad ((\rho - \rho^3)(1 + \rho^2 - \rho^4 - \rho^6) - (\rho - \rho^5)(1 - \rho^4)) \\
&= 0 \\
\frac{\partial V(Y_2 - Y_4|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_4} &= \frac{2\sigma^6((\rho^2 a_1 + a_2)(\rho - \rho^5) + (a_3 - a_4)(\rho - \rho^3))}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * \\
&\quad ((\rho - \rho^3)(1 + \rho^2 - \rho^4 - \rho^6) + (\rho - \rho^5)(\rho^4 - 1)) \\
&= 0
\end{aligned}$$

Thus, for sequence BDAC, an optimal LCB is $X_2 + X_3 - (1 + \rho^2)X_4$.

– $V(Y_3 - Y_4|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})$ corresponds to sequences DCBA. The corresponding partials are:

$$\frac{\partial V(Y_3 - Y_4|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_1} = \frac{-2\sigma(\rho - \rho^3)^2(a_1\rho^4 + a_3\rho^2 + a_3)}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * (a_2(a_1\rho^6 + a_3\rho^6 + a_4\rho^8 + a_1\rho^2 + a_3\rho^2) + a_1(a_3\rho^8 + a_4\rho^{10} + a_3) + a_4\rho^4(a_4 + a_3\rho^2))$$

$$\frac{\partial V(Y_3 - Y_4|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_2} = \frac{-2\sigma(\rho - \rho^3)^2(a_1\rho^4 + a_3\rho^2 + a_3)}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * (a_1(a_1(\rho^2 - \rho^6) + a_2\rho^4 + a_3\rho^6 + a_4\rho^8 - a_3\rho^2) + a_2(a_3\rho^4 + a_4\rho^6 - a_3) + a_4\rho^2(a_4 + a_3\rho^2))$$

$$\frac{\partial V(Y_3 - Y_4|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_3} = \frac{-2\sigma(\rho - \rho^3)^2(a_1\rho^4 + a_3\rho^2 + a_3)}{V(\boldsymbol{a}^T\boldsymbol{X})^2} * (a_1(a_1 + a_2\rho^2 + a_4\rho^6 - a_1\rho^8 - 2a_2\rho^6) + a_2(a_2 + a_4\rho^4 - \rho^4 a_2) + a_4(a_4 + a_3\rho^2))$$

$$\frac{\partial V(Y_3 - Y_4|\boldsymbol{a}^T\boldsymbol{X}, \Sigma_{AR})}{\partial a_4} = \frac{2\sigma(\rho - \rho^3)^2(a_1\rho^4 + a_3\rho^2 + a_3)}{V(\boldsymbol{a}^T\boldsymbol{X})^2}(a_1\rho^6 + a_2\rho^4 + a_3\rho^2 + a_4)$$

Each partial equals zero for (1) $a_1\rho^4 + a_3\rho^2 + a_3 = 0$ and (2) $a_1 = a_2 = 0$, $a_4 = -\rho^2 a_3$. The first corresponds to the maximum and reduces the conditional variance to $V(Y_3 - Y_4|\mathbf{a}^T\mathbf{X})$. The later, $\mathbf{a}^T = (a_1 = a_2 = 0, a_4 = -\rho^2 a_3)$ corresponds to the minimum. Thus, for sequence DCBA, an optimal LCB is $X_3 - \rho^{-2}X_4$.

B.2. Simulation Information: Treatment Effect Sizes and Covariances

Table B.1: Treatment Effect Sizes: 2×2, 3×3 and 4×4 Design

2×2 Design	Σ_{CS}	Σ_{DCS}	Σ_{EP}	Σ_{AR}	Σ_{UN}
N=12	0.80	0.89	0.80	0.87	0.89
N=16	0.67	0.75	0.67	0.73	0.75
N=20	0.59	0.66	0.59	0.64	0.66
N=24	0.54	0.60	0.54	0.58	0.60
N=28	0.49	0.55	0.49	0.53	0.55
N=32	0.46	0.51	0.46	0.49	0.51
3×3 Design	Σ_{CS}	Σ_{DCS}	Σ_{EP}	Σ_{AR}	Σ_{UN}
N=18	0.64	0.68	0.64	0.69	0.67
N=24	0.54	0.57	0.54	0.58	0.56
N=30	0.48	0.51	0.48	0.51	0.50
N=36	0.43	0.46	0.43	0.46	0.45
N=42	0.40	0.42	0.40	0.42	0.41
N=48	0.37	0.39	0.37	0.40	0.38
4×4 Design	Σ_{CS}	Σ_{DCS}	Σ_{EP}	Σ_{AR}	Σ_{UN}
N=16	0.68	0.73	0.69	0.70	0.69
N=20	0.60	0.63	0.60	0.61	0.60
N=24	0.54	0.58	0.54	0.55	0.54
N=28	0.50	0.53	0.50	0.51	0.50
N=32	0.46	0.49	0.46	0.47	0.46
N=36	0.43	0.46	0.43	0.44	0.43

Note: Treatment effect sizes refer to the true value of: $\tau_A - \tau_B$, $\tau_C - \tau_B$, $\tau_D - \tau_B$ as appropriate.

Simulations: Response Vector Parameters under AR(1)

- 2×2 Design: $\mu = 6$, $\gamma_{id[i,1]} = 0$, $\gamma_{id[i,2]} = 1$, $\zeta_{i1} = 0$, $\zeta_{i2} = 1$.

- 3×3 Design: $\mu = 6$, $\gamma_{id[i,1]} = 0$, $\gamma_{id[i,2]} = 1$, $\gamma_{id[i,3]} = 2$, $\zeta_{i1} = 0$, $\zeta_{i2} = 1$, $\zeta_{i3} = 2$.

- 4×4 Design: $\mu = 6$, $\gamma_{id[i,1]} = 0$, $\gamma_{id[i,2]} = 1$, $\gamma_{id[i,3]} = 2$, $\gamma_{id[i,4]} = 2.5$, $\zeta_{i1} = 0$, $\zeta_{i2} = 1$, $\zeta_{i3} = 2$, $\zeta_{i4} = 2.5$.

Simulations: Covariance Parameters

- 2×2 Design: Under CS, $\rho = 0.6$; Under EP, $\rho_1 = 0.60$, $\rho_2 = 0.70$, $\rho_3 = 0.50$; Under DCS, $\rho_1 = 0.55$, $\rho_2 = 0.80$; under AR(1), $\rho = 0.73$; under UN, $\rho_{AB}^{XX} = 0.40$, $\rho_{AB}^{YY} = 0.50$, $\rho_{AA}^{XY} = 0.82$, $\rho_{AB}^{XY} = 0.60$, $\rho_{BA}^{XY} = 0.55$, $\rho_{BB}^{XY} = 0.72$.

- 3×3 Design: Under CS, $\rho = 0.6$; Under EP, $\rho_1 = 0.60$, $\rho_2 = 0.70$, $\rho_3 = 0.50$; Under DCS, $\rho_1 = 0.55$, $\rho_2 = 0.75$; under AR(1), $\rho = 0.789$; and under UN, $\rho_{AB}^{XX} = 0.43$, $\rho_{AC}^{XX} = 0.45$, $\rho_{BC}^{XX} = 0.47$, $\rho_{AB}^{YY} = 0.57$,$\rho_{AC}^{YY} = 0.55$,$\rho_{BC}^{YY} = 0.52$, $\rho_{AA}^{XY} = 0.85$, $\rho_{AB}^{XY} = 0.65$, $\rho_{AC}^{XY} = 0.61$. $\rho_{BA}^{XY} = 0.55$, $\rho_{BB}^{XY} = 0.75$, $\rho_{BC}^{XY} = 0.64$, $\rho_{CA}^{XY} = 0.63$, $\rho_{CB}^{XY} = 0.53$, $\rho_{CC}^{XY} = 0.72$.

- 4×4 Design: Under CS, $\rho = 0.6$; Under EP, $\rho_1 = 0.60$, $\rho_2 = 0.70$, $\rho_3 = 0.50$; Under DCS, $\rho_1 = 0.55$, $\rho_2 = 0.80$; under AR(1), $\rho = 0.83$; and under UN, $\rho_{AB}^{XX} = 0.46$, $\rho_{AC}^{XX} = 0.45$, $\rho_{AD}^{XX} = 0.42$, $\rho_{BC}^{XX} = 0.50$, $\rho_{BD}^{XX} = 0.48$, $\rho_{CD}^{XX} = 0.55$, $\rho_{AB}^{YY} = 0.60$, $\rho_{AC}^{YY} = 0.54$, $\rho_{AD}^{YY} = 0.52$, $\rho_{BC}^{YY} = 0.56$, $\rho_{BD}^{YY} = 0.52$, $\rho_{CD}^{YY} = 0.55$, $\rho_{AA}^{XY} = 0.80$, $\rho_{AB}^{XY} = 0.60$, $\rho_{AC}^{XY} = 0.55$, $\rho_{AD}^{XY} = 0.62$, $\rho_{BA}^{XY} = 0.61$, $\rho_{BB}^{XY} = 0.80$, $\rho_{BC}^{XY} = 0.63$, $\rho_{BD}^{XY} = 0.55$, $\rho_{CA}^{XY} = 0.50$, $\rho_{CB}^{XY} = 0.55$, $\rho_{CC}^{XY} = 0.82$, $\rho_{CD}^{XY} = 0.60$, $\rho_{DA}^{XY} = 0.55$, $\rho_{DB}^{XY} = 0.52$, $\rho_{DC}^{XY} = 0.50$, $\rho_{DD}^{XY} = 0.77$.

B.3. Uniform Design Type I Error Results

Table B.2: 2×2 Simulations: Under the Null Hypothesis, Type I Error

Truth	**Method**	N=12	N=16	N=20	N=24	N=28	N=32
Σ_{CS}	No Baselines	5.17	5.00	4.88	5.00	4.79	5.08
Σ_{CS}	CFB	[5.44]	5.05	4.94	4.94	4.67	4.73
Σ_{CS}	$X_A - X_B$	5.14	5.01	4.79	4.94	4.84	5.05
Σ_{CS}	AR(1)	5.18	4.95	4.80	5.12	4.92	5.08
Σ_{CS}	Adaptive	[[5.96]]	[5.42]	5.26	[5.33]	5.08	[5.36]
Σ_{DCS}	No Baselines	4.95	4.75	5.20	4.88	4.84	4.81
Σ_{DCS}	CFB	4.98	4.72	[5.37]	5.04	4.92	5.01
Σ_{DCS}	$X_A - X_B$	5.03	4.72	5.16	5.07	5.04	4.73
Σ_{DCS}	AR(1)	5.20	5.04	5.27	5.01	5.03	4.75
Σ_{DCS}	Adaptive	[[5.92]]	5.20	[[5.48]]	5.21	5.12	4.75
Σ_{EP}	No Baselines	4.98	4.87	4.96	5.00	4.86	4.99
Σ_{EP}	CFB	5.01	4.79	5.24	4.97	4.90	4.97
Σ_{EP}	$X_A - X_B$	5.07	4.84	5.21	4.96	4.90	4.83
Σ_{EP}	AR(1)	5.09	4.92	5.02	5.08	4.87	4.67
Σ_{EP}	Adaptive	[[6.36]]	[[5.62]]	[[5.78]]	[[5.55]]	5.21	5.09
Σ_{AR}	No Baselines	5.19	5.31	5.20	5.06	4.83	4.89
Σ_{AR}	CFB	[5.53]	[5.35]	[5.38]	5.09	4.97	4.88
Σ_{AR}	$X_A - X_B$	[5.43]	5.25	5.21	4.83	4.91	4.99
Σ_{AR}	AR(1)	[[5.54]]	[[5.51]]	[5.32]	5.05	5.03	4.93
Σ_{AR}	Adaptive	[[6.04]]	[[5.86]]	[[5.48]]	5.16	5.16	5.04
Σ_{UN}	No Baselines	5.28	4.94	4.96	4.83	5.01	4.75
Σ_{UN}	CFB	4.94	4.65	5.21	4.98	4.97	5.01
Σ_{UN}	$X_A - X_B$	5.10	4.79	5.09	4.97	5.05	4.72
Σ_{UN}	AR(1)	5.04	4.99	5.08	5.01	5.00	4.79
Σ_{UN}	Adaptive	[5.41]	5.00	5.22	5.02	5.08	4.75

Notes: Type I error (%) shown. Entries are in brackets/double brackets if the type I error is two/three SE's above 5% ($> 5.31\%$,$> 5.46\%$) based on 20,000 simulations. Method AR(1) refers to the covariates derived in Table 2.2 under AR(1). The adaptive method, based on AICC values, chooses between methods No Baselines, $X_A - X_B$, and Method AR(1). CS = Compound Symmetry; DCS = Double Compound Symmetry; EP = Equipredictability; AR = Auto-regressive (1); UN=Unstructured.

Table B.3: 3×3 Simulations: Under the Null Hypothesis, Type I Error

Truth	**Method**	N=18	N=24	N=30	N=36	N=42	N=48
Σ_{CS}	No Baselines	4.84	4.84	4.87	5.24	5.05	5.07
Σ_{CS}	CFB	5.00	5.27	5.07	4.88	4.93	4.99
Σ_{CS}	$X_A - X_B$	4.90	4.96	4.79	5.04	4.93	5.07
Σ_{CS}	AR(1)	4.84	4.88	4.93	5.17	4.91	5.11
Σ_{CS}	Adaptive	5.23	5.09	5.00	5.27	5.16	5.18
Σ_{DCS}	No Baselines	4.92	4.82	4.83	5.04	5.21	5.07
Σ_{DCS}	CFB	5.17	5.00	5.16	4.91	4.79	4.92
Σ_{DCS}	$X_A - X_B$	4.83	4.90	4.84	5.08	4.92	5.09
Σ_{DCS}	AR(1)	5.00	4.85	4.89	5.08	5.08	5.12
Σ_{DCS}	Adaptive	5.30	5.04	4.96	5.13	4.95	5.11
Σ_{EP}	No Baselines	4.71	4.88	4.82	5.17	5.25	5.09
Σ_{EP}	CFB	4.82	5.13	4.99	4.90	4.74	5.00
Σ_{EP}	$X_A - X_B$	4.91	4.97	4.82	5.01	4.94	5.13
Σ_{EP}	AR(1)	4.76	4.87	4.83	5.08	4.96	5.11
Σ_{EP}	Adaptive	5.25	5.08	4.86	5.01	4.94	5.13
Σ_{AR}	No Baselines	5.01	4.99	4.96	5.29	5.16	5.08
Σ_{AR}	CFB	4.96	4.68	5.03	5.18	5.25	4.81
Σ_{AR}	$X_A - X_B$	5.05	4.79	4.94	5.18	5.30	5.05
Σ_{AR}	AR(1)	5.16	4.86	5.30	5.21	5.21	5.00
Σ_{AR}	Adaptive	[5.34]	4.89	5.30	5.22	5.21	5.01
Σ_{UN}	No Baselines	5.07	4.96	4.75	5.09	5.21	5.06
Σ_{UN}	CFB	4.11	4.14	4.21	3.99	3.81	3.83
Σ_{UN}	$X_A - X_B$	5.26	4.92	4.92	4.99	4.95	4.94
Σ_{UN}	AR(1)	5.08	4.79	4.84	5.01	5.17	4.95
Σ_{UN}	Adaptive	5.26	4.92	4.92	4.99	4.95	4.94

Notes: Type I error (%) shown. Entries are in brackets/double brackets if the type I error is two/three SE's above 5% ($> 5.31\%$,$> 5.46\%$) based on 20,000 simulations. Method AR(1) refers to the covariates derived in Table 2.2 under AR(1). The adaptive method, based on AICC values, chooses between methods No Baselines, $X_A - X_B$, and Method AR(1). CS = Compound Symmetry; DCS = Double Compound Symmetry; EP = Equipredictability; AR = Auto-regressive (1); UN=Unstructured.

Table B.4: 4×4 Simulations: Under the Null Hypothesis, Type I Error

Truth	**Method**	N=16	N=20	N=24	N=28	N=32	N=36
Σ_{CS}	No Baselines	5.08	5.22	5.03	4.91	4.86	4.95
Σ_{CS}	CFB	4.88	5.03	5.02	5.01	4.77	5.00
Σ_{CS}	$X_A - X_B$	5.00	5.02	5.00	4.88	4.90	4.95
Σ_{CS}	AR(1)	5.01	5.13	5.16	4.97	4.93	5.05
Σ_{CS}	Adaptive	5.24	5.28	5.10	4.90	4.88	4.98
Σ_{DCS}	No Baselines	4.93	4.89	5.09	4.93	5.22	4.97
Σ_{DCS}	CFB	5.12	5.01	5.04	4.94	4.83	4.92
Σ_{DCS}	$X_A - X_B$	5.04	4.98	5.02	4.94	4.99	4.90
Σ_{DCS}	AR(1)	5.07	5.10	5.22	4.91	5.25	4.92
Σ_{DCS}	Adaptive	5.20	5.04	5.03	4.96	5.00	4.90
Σ_{EP}	No Baselines	5.04	4.90	5.10	4.92	5.26	5.04
Σ_{EP}	CFB	4.96	5.04	4.92	4.96	4.81	4.89
Σ_{EP}	$X_A - X_B$	5.00	5.17	5.13	4.82	4.84	5.03
Σ_{EP}	AR(1)	5.04	5.12	5.11	5.04	5.30	4.79
Σ_{EP}	Adaptive	5.08	5.19	5.16	4.82	4.84	5.03
Σ_{AR}	No Baselines	4.95	5.21	5.09	5.14	5.09	5.28
Σ_{AR}	CFB	4.92	5.07	4.88	5.17	5.16	5.25
Σ_{AR}	$X_A - X_B$	4.89	5.01	5.00	5.14	5.17	5.16
Σ_{AR}	AR(1)	5.11	5.05	5.04	5.10	5.08	5.22
Σ_{AR}	Adaptive	5.16	5.06	5.04	5.09	5.08	5.22
Σ_{UN}	No Baselines	5.04	4.98	5.03	5.05	5.01	4.86
Σ_{UN}	CFB	[[6.02]]	[[6.00]]	[[5.91]]	[[5.89]]	[[5.79]]	[[5.96]]
Σ_{UN}	$X_A - X_B$	5.01	5.11	5.08	4.92	5.00	4.93
Σ_{UN}	AR(1)	5.18	5.05	5.07	4.99	5.11	4.84
Σ_{UN}	Adaptive	5.04	5.12	5.08	4.92	5.00	4.93

Notes: Type I error (%) shown. Entries are in brackets/double brackets if the type I error is two/three SE's above 5% ($> 5.31\%$,$> 5.46\%$) based on 20,000 simulations. Method AR(1) refers to the covariates derived in Table 2.2 under AR(1). The adaptive method, based on AICC values, chooses between methods No Baselines, $X_A - X_B$, and Method AR(1). CS = Compound Symmetry; DCS = Double Compound Symmetry; EP = Equipredictability; AR = Auto-regressive (1); UN=Unstructured.

APPENDIX C

CHAPTER 3, INCOMPLETE BLOCK DESIGNS

C.1. Mixed Effects Models: Variance Formulas

For the 3×2 design, closed form solutions are derived for $\widehat{\tau_A - \tau_B}$ and $V(\widehat{\tau_A - \tau_B})$. Consider the following mean model:

$$Y^{\star}_{ijk} = \mu + \tau_{d[i,j]} + \pi_j + \epsilon_{ijk}$$

where $Y^{\star}_{ijk}$ depends on the chosen baseline model. In the No Baselines Model, $Y^{\star}_{ijk} = Y_{ijk}$. For the baseline mixed models, such as $\mathbf{Y}|\mathbf{X}, \bar{\mathbf{X}}$, the regressed outcome is used, $Y^{\star}_{ijk} = Y_{ijk} - \beta_j X_{ijk} - \gamma(X_{i1k} + X_{i2k})$. Denote the vector of crossover design parameters as $\boldsymbol{\eta} = (\mu, \pi_1, \tau_A, \tau_B)^T$. In this formulation, π_2 and τ_C are treated as reference paramters. For ease of notation, let $i = 1, 2, 3, 4, 5, 6$ correspond to sequences AB, BA, BC, CB, AC, CA. Sequence specific design matrices corresponding to $\boldsymbol{\eta}$ are then:

$$\begin{aligned}
\mathbf{X}_1 &= \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \mathbf{X}_2 = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}, \mathbf{X}_3 = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} \\
\mathbf{X}_4 &= \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \mathbf{X}_5 = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \mathbf{X}_6 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}
\end{aligned}$$

Next, let $\Sigma_{\star} = V((Y^{\star}_{i1k}, Y^{\star}_{i2k})^T)$. The inverse of this variance is written as:

$$\Sigma^{-1}_{\star} = \frac{1}{V(Y^{\star}_{i1k})V(Y^{\star}_{i2k}) - \text{cov}(Y^{\star}_{i1k}, Y^{\star}_{i2k})^2} \begin{pmatrix} V(Y^{\star}_{i2k}) & -\text{cov}(Y^{\star}_{i1k}, Y^{\star}_{i2k}) \\ & V(Y^{\star}_{i1k}) \end{pmatrix} = \begin{pmatrix} a & c \\ c & b \end{pmatrix} \quad \text{(C.1)}$$

Then, the GLS estimate of $\boldsymbol{\eta}$ and $V(\boldsymbol{\eta})$ are:

$$\begin{aligned}
\hat{\boldsymbol{\eta}} &= (\sum_{i=1}^{s} \sum_{k=1}^{n_i} \mathbf{X}_i^T \Sigma_{\star}^{-1} \mathbf{X}_i)^{-1} (\sum_{i=1}^{s} \sum_{k=1}^{n_i} \mathbf{X}_i^T \Sigma_{\star}^{-1} \mathbf{Y}_{ik}) \\
V(\hat{\boldsymbol{\eta}}) &= (\sum_{i=1}^{s} \sum_{k=1}^{n_i} \mathbf{X}_i^T \Sigma_{\star}^{-1} \mathbf{X}_i)^{-1}
\end{aligned}$$

Next, after some algebra, $V(\hat{\boldsymbol{\eta}})^{-1}$ equals:

$$\begin{pmatrix} \sum_{i=1}^{s} n_i(a+b+2c) & \sum_{i=1}^{s} n_i(a+c) & (n_1+n_5)(a+c)+(n_2+n_6)(b+c) & (n_2+n_3)(a+c)+(n_1+n_4)(b+c) \\ & \sum_{i=1}^{s} n_i a & (n_1+n_5)a+(n_2+n_6)c & (n_2+n_3)a+(n_1+n_4)c \\ & & (n_1+n_5)a+(n_2+n_6)b & (n_1+n_2)c \\ & & & (n_2+n_3)a+(n_1+n_4)b \end{pmatrix}^{-1}$$

For a balanced design, such that $n_i = n$ for all i, this simplifies to:

$$V(\hat{\boldsymbol{\eta}}) = n^{-1} \begin{pmatrix} 6(a+b+2c) & 6(a+c) & 2(a+b+2c) & 2(a+b+2c) \\ & 6a & 2(a+c) & 2(a+c) \\ & & 2(a+b) & 2c \\ & & & 2(a+b) \end{pmatrix}^{-1} = n^{-1} \begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1}$$

Note that the A,B, C, and D represent block matrices of $V(\hat{\boldsymbol{\eta}})^{-1}$. To solve for $\widehat{\tau_A - \tau_B}$ and $V(\widehat{\tau_A - \tau_B})$, we need to solve for D^{-1} and C^{-1}. Using block inversion, we know that $D^{-1} = (D - CA^{-1}B)^{-1}$Then:

$$\begin{aligned} D - CA^{-1}B &= 2\begin{pmatrix} 2(a+b) & 2c \\ 2c & 2(a+b) \end{pmatrix} - \\ &\quad \begin{pmatrix} 2(a+b+2c) & 2(a+c) \\ 2(a+b+2c) & 2(a+c) \end{pmatrix} \frac{1}{36(ab-c^2)} \begin{pmatrix} 6a & -6(a+c) \\ -6(a+c) & 6(a+b+2c) \end{pmatrix} \begin{pmatrix} 2(a+b+2c) & 2(a+b+2c) \\ 2(a+c) & 2(a+c) \end{pmatrix} \\ &= \begin{pmatrix} 2(a+b) & 2c \\ 2c & 2(a+b) \end{pmatrix} - \frac{1}{3}\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 2(a+b+2c) & 2(a+b+2c) \\ 2(a+c) & 2(a+c) \end{pmatrix} \\ &= \begin{pmatrix} 2(a+b) & 2c \\ 2c & 2(a+b) \end{pmatrix} - \frac{2}{3}\begin{pmatrix} a+b+2c & a+b+2c \\ a+b+2c & a+b+2c \end{pmatrix} \\ &= \begin{pmatrix} 4/3(a+b-c) & -2/3(a+b-c) \\ -2/3(a+b-c) & 4/3(a+b-c) \end{pmatrix} \\ D^{-1} = (D - CA^{-1}B)^{-1} &= \frac{1}{16/9(a+b-c)^2 - 4/9(a+b-c)^2} \begin{pmatrix} 4/3(a+b-c) & 2/3(a+b-c) \\ 2/3(a+b-c) & 4/3(a+b-c) \end{pmatrix} \\ &= \frac{1}{a+b-c}\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \end{aligned}$$

Thus:

$$V(\widehat{\tau_A - \tau_B}) = V(\hat{\tau}_A) + V(\hat{\tau}_B) - 2\mathsf{cov}(\hat{\tau}_A, \hat{\tau}_B) = \frac{1}{n(a+b-c)} = \frac{1}{n} \frac{V(Y^\star_{i1k})V(Y^\star_{i2k}) - \mathsf{cov}(Y^\star_{i1k}, Y^\star_{i2k})^2}{V(Y^\star_{i1k}) + V(Y^\star_{i2k}) + \mathsf{cov}(Y^\star_{i1k}, Y^\star_{i2k})} \quad \text{(C.2)}$$

Now that the variance parameters associated with τ_A and τ_B are known, the closed form for the $\widehat{\tau_A - \tau_B}$ can be found. After some work:

$$\begin{aligned}\widehat{\tau_A - \tau_B} = \Big(& \sum_{k=1}^{n_1}((a-c)Y^\star_{11k} - (b-c)Y^\star_{12k}) + \sum_{k=1}^{n_2}((b-c)Y^\star_{22k} - (a-c)Y^\star_{21k}) \\ & - \sum_{k=1}^{n_3}(aY^\star_{31k} + cY^\star_{32k}) - \sum_{k=1}^{n_4}(cY^\star_{41k} + bY^\star_{42k}) \\ & + \sum_{k=1}^{n_5}(aY^\star_{51k} + cY^\star_{52k}) + \sum_{k=1}^{n_6}(cY^\star_{61k} + bY^\star_{62k})\Big)\frac{1}{2n(a+b-c)} \end{aligned} \tag{C.3}$$

Overall, these results show the closed form for the treatment estimate (C.3) and it's corresponding variance (C.2) for any given mixed model. Note that these closed forms largely depend on $\Sigma_\star = V((Y^\star_{i1k}, Y^\star_{i2k})^T)$, which ultimately depends on the chosen mixed model.

C.1.1. Mixed Model: No Baselines

This model uses the unadjusted outcomes, $Y^\star_{ijk} = Y_{ijk}$. Under an unstructured covariance assumption (3.6), $V(Y_{i1k}) = \sigma_2$, $V(Y_{i2k}) = \sigma_4$, and cov$(Y_{i1k}, Y_{i2k}) = \sigma_{24}$. The estimate of $\tau_A - \tau_B$ is found by substituting $Y^\star_{ijk} = Y_{ijk}$, $a = \frac{\sigma_4}{\sigma_2\sigma_4 - \sigma_{24}^2}$, $b = \frac{\sigma_2}{\sigma_2\sigma_4 - \sigma_{24}^2}$, and $c = \frac{-\sigma_{24}}{\sigma_2\sigma_4 - \sigma_{24}^2}$. The variance of $\widehat{\tau_A - \tau_B}$ is:

$$V(\widehat{\tau_A - \tau_B}) = \frac{1}{n}\frac{\sigma_2\sigma_4 - \sigma_{24}^2}{\sigma_4 + \sigma_2 + \sigma_{24}} \tag{C.4}$$

C.1.2. Mixed Model: Change from Baselines (CFB)

For the CFB model, $Y^\star_{ijk} = Y_{ijk} - X_{ijk}$ with $V(Y_{i1k} - X_{i2k}) = \sigma_1 + \sigma_2 - 2\sigma_{12}$, $V(Y_{i2k} - X_{i2k}) = \sigma_3 + \sigma_4 - 2\sigma_{34}$, and cov$(Y_{i1k} - X_{i1k}, Y_{i2k} - X_{i2k}) = \sigma_{24} - \sigma_{23} - \sigma_{14} + \sigma_{13}$. The estimate of $\tau_A - \tau_B$ and it's variance can be calculated by substituting in these variance quantities as appropriate. The variance is:

$$V(\widehat{\tau_A - \tau_B}) = \frac{1}{n}\frac{(\sigma_1 + \sigma_2 - 2\sigma_{12})(\sigma_3 + \sigma_4 - 2\sigma_{34}) - (\sigma_{24} - \sigma_{23} - \sigma_{14} + \sigma_{13})^2}{(\sigma_1 + \sigma_2 - 2\sigma_{12}) + (\sigma_3 + \sigma_4 - 2\sigma_{34}) + (\sigma_{24} - \sigma_{23} - \sigma_{14} + \sigma_{13})} \tag{C.5}$$

C.1.3. Mixed Models: Baseline Adjustment

Here we examine closed form solutions for the baseline mixed models, $\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$ and $\mathbf{Y}|\mathbf{X}\mathbf{a}_\star, \overline{\mathbf{X}\mathbf{a}}_\star$. These solutions will be examined under the period-specific LCB model as the simpler $\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$ is really just a special case of $\mathbf{Y}|\mathbf{X}\mathbf{a}_\star, \overline{\mathbf{X}\mathbf{a}}_\star$. Consider the period-specific LCB model with pre-specified LCBs, $\mathbf{a}_1\mathbf{X}_{ik}$ and $\mathbf{a}_2\mathbf{X}_{ik}$. The outcomes of interest are then:

$$Y^\star_{i1k} = Y_{i1k} - \beta_1(\mathbf{a}_1^T\mathbf{X}_{ik}) - \gamma(\mathbf{a}_1^T\mathbf{X}_{ik} + \mathbf{a}_2^T\mathbf{X}_{ik})$$
$$Y^\star_{i2k} = Y_{i2k} - \beta_2(\mathbf{a}_2^T\mathbf{X}_{ik}) - \gamma(\mathbf{a}_1^T\mathbf{X}_{ik} + \mathbf{a}_2^T\mathbf{X}_{ik})$$

The goal is to find the covariance of these regressed outcomes. This first requires solving for β_1, β_2, and γ. For simplicity and easier derivations, we can assume that $E(Y_{ijk}) = E(\mathbf{a}_j^T\mathbf{X}_{ik}) = 0$. Let the baselines design matrix for sequence i be denoted as:

$$\boldsymbol{B}_i = \begin{pmatrix} \mathbf{a}_1\mathbf{X}_{ik} & 0 & (\mathbf{a}_1\mathbf{X}_{ik} + \mathbf{a}_2\mathbf{X}_{ik}) \\ 0 & \mathbf{a}_2\mathbf{X}_{ik} & (\mathbf{a}_1\mathbf{X}_{ik} + \mathbf{a}_2\mathbf{X}_{ik}) \end{pmatrix}$$

The GLS estimate of $\boldsymbol{\zeta} = (\beta_1, \beta_2, \gamma)^T$ is then:

$$\hat{\boldsymbol{\zeta}} = (\sum_{i=1}^{s}\sum_{k=1}^{n_i}\mathbf{B}_i^T\Sigma_B^{-1}\mathbf{B}_i)^{-1}(\sum_{i=1}^{s}\sum_{k=1}^{n_i}\mathbf{B}_i^T\Sigma_B^{-1}\mathbf{Y}_{ik})$$

In this setting, Σ_B is the covariance for the conditional outcomes, $\mathbf{Y}_{ik}|\mathbf{a}_1\mathbf{X}_{ik}, \mathbf{a}_2\mathbf{X}_{ik}$. This actually equals the covariance for the simpler conditional outcomes, $\mathbf{Y}_{ik}|X_{i1k}, X_{i2k}$. To see this, note that $\mathbf{Y}_{ik}|\mathbf{a}_1\mathbf{X}_{ik}, \mathbf{a}_2\mathbf{X}_{ik}$ implies the following regression model:

$$Y_{ijk} = \lambda_1\mathbf{a}_1\mathbf{X}_{ik} + \lambda_2\mathbf{a}_2\mathbf{X}_{ik} = (\lambda_1 a_{11} + \lambda_2 a_{21})X_{i1k} + (\lambda_1 a_{12} + \lambda_2 a_{22})X_{i2k} = \lambda_1^\star X_{i1k} + \lambda_2^\star X_{i2k}$$

Next, using standard normal theory and the block matrices from (3.2), it follows that:

$$\Sigma_B^{-1} = \left(\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\right)^{-1} = \begin{pmatrix} d & e \\ e & f \end{pmatrix}$$

Given this, it follows that:

$$\sum_{i=1}^{s}\sum_{k=1}^{n_i}\mathbf{B}_i^T\Sigma_B^{-1}\mathbf{B}_i = \begin{pmatrix} d\sum_i^s\sum_k^{n_i}(\mathbf{a}_1\mathbf{X}_{ik})^2 & e\sum_i^s\sum_k^{n_i}(\mathbf{a}_1\mathbf{X}_{ik})(\mathbf{a}_2\mathbf{X}_{ik}) & (d+e)\sum_i^s\sum_k^{n_i}\left((\mathbf{a}_1\mathbf{X}_{ik})^2+(\mathbf{a}_1\mathbf{X}_{ik})(\mathbf{a}_2\mathbf{X}_{ik})\right) \\ & f\sum_i^s\sum_k^{n_i}(\mathbf{a}_2\mathbf{X}_{ik})^2 & (f+e)\sum_i^s\sum_k^{n_i}\left((\mathbf{a}_2\mathbf{X}_{ik})^2+(\mathbf{a}_1\mathbf{X}_{ik})(\mathbf{a}_2\mathbf{X}_{ik})\right) \\ & & (d+f+2e)\sum_i^s\sum_k^{n_i}\left((\mathbf{a}_1\mathbf{X}_{ik})^2+(\mathbf{a}_2\mathbf{X}_{ik})^2+2(\mathbf{a}_1\mathbf{X}_{ik})(\mathbf{a}_2\mathbf{X}_{ik})\right) \end{pmatrix}$$

Additionally:

$$(\sum_{i=1}^{s}\sum_{k=1}^{n_i}\mathbf{B}_i^T\Sigma_B^{-1}\mathbf{Y}_{ik}) = \begin{pmatrix} d\sum_i^s\sum_k^{n_i}((\mathbf{a}_1\mathbf{X}_{ik})Y_{i1k})+e\sum_i^s\sum_k^{n_i}((\mathbf{a}_1\mathbf{X}_{ik})Y_{i2k}) \\ f\sum_i^s\sum_k^{n_i}((\mathbf{a}_2\mathbf{X}_{ik})Y_{i2k})+e\sum_i^s\sum_k^{n_i}((\mathbf{a}_2\mathbf{X}_{ik})Y_{i1k}) \\ (d+e)\sum_i^s\sum_k^{n_i}((\mathbf{a}_1\mathbf{X}_{ik})Y_{i1k}+(\mathbf{a}_2\mathbf{X}_{ik})Y_{i1k})+(f+e)\sum_i^s\sum_k^{n_i}((\mathbf{a}_1\mathbf{X}_{ik})Y_{i2k}+(\mathbf{a}_2\mathbf{X}_{ik})Y_{i2k}) \end{pmatrix}$$

Putting this all together, we note that:

$$\hat{\boldsymbol{\zeta}} \xrightarrow{p} \begin{pmatrix} (d)V(\mathbf{a}_1\mathbf{X}_{ik}) & (e)\mathsf{cov}(\mathbf{a}_1\mathbf{X}_{ik},\mathbf{a}_2\mathbf{X}_{ik}) & (d+e)(V(\mathbf{a}_1\mathbf{X}_{ik})+\mathsf{cov}(\mathbf{a}_1\mathbf{X}_{ik},\mathbf{a}_2\mathbf{X}_{ik})) \\ & (f)V(\mathbf{a}_2\mathbf{X}_{ik}) & (f+e)(V(\mathbf{a}_2\mathbf{X}_{ik})+\mathsf{cov}(\mathbf{a}_1\mathbf{X}_{ik},\mathbf{a}_2\mathbf{X}_{ik})) \\ & & (d+f+2e)\left(V(\mathbf{a}_1\mathbf{X}_{ik})+V(\mathbf{a}_2\mathbf{X}_{ik})+2\mathsf{cov}(\mathbf{a}_1\mathbf{X}_{ik},\mathbf{a}_2\mathbf{X}_{ik})\right) \end{pmatrix}^{-1}$$
$$* \begin{pmatrix} (d)\mathsf{cov}(Y_{i1k},\mathbf{a}_1\mathbf{X}_{ik})+(e)\mathsf{cov}(Y_{i2k},\mathbf{a}_1\mathbf{X}_{ik}) \\ (f)\mathsf{cov}(Y_{i2k},\mathbf{a}_2\mathbf{X}_{ik})+(e)\mathsf{cov}(Y_{i2k},\mathbf{a}_2\mathbf{X}_{ik}) \\ (d+e)\left(\mathsf{cov}(Y_{i1k},\mathbf{a}_1\mathbf{X}_{ik})+\mathsf{cov}(Y_{i1k},\mathbf{a}_2\mathbf{X}_{ik})\right)+(f+e)\left(\mathsf{cov}(Y_{i2k},\mathbf{a}_1\mathbf{X}_{ik})+\mathsf{cov}(Y_{i2k},\mathbf{a}_2\mathbf{X}_{ik})\right) \end{pmatrix} = (\beta_1,\beta_2,\gamma)^T$$

Thus, (β_1,β_2,γ) can be determined directly through the joint covariance of the baselines and outcomes (3.2). The variance of the regressed outcomes can now be determined. Note that:

$$\begin{aligned} V(Y_{ijk}^\star) &= V(Y_{ijk}-\beta_j(\mathbf{a}_j\mathbf{X}_{ik})-\gamma(\mathbf{a}_1\mathbf{X}_{ik}+\mathbf{a}_2\mathbf{X}_{ik})) \\ &= V(Y_{ijk})+\beta_j^2V(\mathbf{a}_j\mathbf{X}_{ik})+\gamma^2V(\mathbf{a}_1\mathbf{X}_{ik}+\mathbf{a}_2\mathbf{X}_{ik}) \\ &\quad -2\beta_j\mathsf{cov}(Y_{ijk},\mathbf{a}_j\mathbf{X}_{ik})-2\gamma\mathsf{cov}(Y_{ijk},\mathbf{a}_j\mathbf{X}_{ik})+2\beta_j\gamma(V(\mathbf{a}_j\mathbf{X}_{ik})+\mathsf{cov}(\mathbf{a}_1\mathbf{X}_{ik},\mathbf{a}_2\mathbf{X}_{ik})) \\ \mathsf{cov}(Y_{i1k}^\star,Y_{i2k}^\star) &= \mathsf{cov}(Y_{i1k}-\beta_1\mathbf{a}_1\mathbf{X}_{ik}-\gamma(\mathbf{a}_1\mathbf{X}_{ik}+\mathbf{a}_2\mathbf{X}_{ik}),Y_{i2k}-\beta_2\mathbf{a}_2\mathbf{X}_{ik}-\gamma(\mathbf{a}_1\mathbf{X}_{ik}+\mathbf{a}_2\mathbf{X}_{ik}) \\ &= \mathsf{cov}(Y_{i1k},Y_{i2k})-\beta_2\mathsf{cov}(Y_{i1k},\mathbf{a}_2\mathbf{X}_{ik})-\gamma\mathsf{cov}(Y_{i1k},\mathbf{a}_1\mathbf{X}_{ik}+\mathbf{a}_2\mathbf{X}_{ik})-\beta_1\mathsf{cov}(Y_{i2k},\mathbf{a}_1\mathbf{X}_{ik}) \\ &\quad +\beta_1\beta_2\mathsf{cov}(\mathbf{a}_1\mathbf{X}_{ik},\mathbf{a}_2\mathbf{X}_{ik})+\beta_1\gamma(V(\mathbf{a}_1\mathbf{X}_{ik})+\mathsf{cov}(\mathbf{a}_1\mathbf{X}_{ik},\mathbf{a}_2\mathbf{X}_{ik}))-\gamma\mathsf{cov}(Y_{i2k},\mathbf{a}_1\mathbf{X}_{ik}+\mathbf{a}_2\mathbf{X}_{ik}) \\ &\quad +\beta_2\gamma(V(\mathbf{a}_2\mathbf{X}_{ik})+\mathsf{cov}(\mathbf{a}_1\mathbf{X}_{ik},\mathbf{a}_2\mathbf{X}_{ik}))+\gamma V(\mathbf{a}_1\mathbf{X}_{ik}+\mathbf{a}_2\mathbf{X}_{ik}) \end{aligned}$$

Based on this, we now have $\Sigma_\star = V((Y_{i1k}^\star,Y_{i2k}^\star)^T)$ and can now estimate $\tau_A-\tau_B$ (C.3) and it's variance (C.2). Notably, this is in a very general form and rather complex. Without any further assumptions, it is difficult to directly compare the variances of the various mixed models.

C.2. WS Model: Variance Formula

Estimation is mode straightforward in the WS Model since it proceeds through standard OLS. For $n = n_i$, after some algebra, it can be shown that:

$$V(\widehat{\tau_A - \tau_B})_{WS} = \frac{1}{3n}\left(V(Y_{i1k} - Y_{i2k}) - \frac{\text{cov}(Y_{i1k} - Y_{i2k}, \mathbf{a}^T\mathbf{X}_{ik})^2}{V(\mathbf{a}^T\mathbf{X}_{ik})}\right) \quad \text{(C.6)}$$

For WS No baselines, set $a_1 = a_2 = 0$. For WS $X_1 - X_2$, set $a_1 = 1$ and $a_2 = -1$.

C.3. Comparison of Variances under CS, EP, AR(1)

Variance formulas under CS, EP, and AR(1) are now considered for all models, except for the period-specific LCB model. Even under these constrained covariance structures, the variance formula for the period-specific LCB model is too complex and doesn't simplify down to any nice form. For the model $\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$:

- Under CS, $\beta_1 = \beta_2 = 0$ and $\gamma = \rho/(\rho + 1)$. This result simplifies the baseline model to using a single period-invariant covariate, $X_{i1k} + X_{i2k}$. After some algebra, it follows that:

$$V(\widehat{\tau_A - \tau_B}) = \frac{1}{n}\frac{\sigma(1 + 2\rho - 3\rho^2)}{5\rho + 2}$$

- Under EP, $\beta_1 = \beta_2 = -(\rho_2 - \rho_3)/(\rho_1 - 1)$ and $\gamma = -(\rho_3 - \rho_1\rho_2)/(\rho_1^2 - 1)$. Although not as concise as the CS case, the variance of interest equals:

$$V(\widehat{\tau_A - \tau_B}) = \frac{1}{n}\frac{-\sigma(-\rho_1^4 + 2\rho_1^2\rho_2^2 + 2\rho_1^2\rho_3^2 + 2\rho_1^2 - 8\rho_1\rho_2\rho_3 - \rho_2^4 + 2\rho_2^2\rho_3^2 + 2\rho_2^2 - \rho_3^4 + 2\rho_3^2 - 1)}{-\rho_1^3 - 2\rho_1^2 + \rho_1\rho_2^2 + 4\rho_1\rho_2\rho_3 + \rho_1\rho_3^2 + \rho_1 - 2\rho_2^2 - 2\rho_2\rho_3 - 2\rho_3^2 + 2}$$

- Under AR(1), $\beta_1 = \beta_2 = \gamma = \rho/(\rho^2 + 2)$ such that:

$$V(\widehat{\tau_A - \tau_B}) = \frac{1}{n}\frac{\sigma(\rho^{10} - 3\rho^8 - 11\rho^6 - 11\rho^4 + 8\rho^2 + 16)}{(\rho^2 + 2)^2(3\rho^4 + 6\rho^2 + 8)}$$

For the No Baselines/CFB mixed models and WS Baseline models, the variance formulas are much simpler. After some algebra we get:

Method	Σ_{CS}	Σ_{EP}	Σ_{AR}
Mixed Model, No Baselines	$\sigma\frac{(1-\rho)(1+\rho)}{2+\rho}$	$\sigma\frac{1-\rho_2}{2+\rho_2}$	$\frac{\sigma(1-\rho^2)}{2+\rho^4}$
Mixed Model, CFB	$\sigma(1-\rho)$	$2\sigma\frac{(1-\rho_1)^2-(\rho_2-\rho_3)^2}{2(1-\rho_1)+(\rho_2-\rho_3)}$	$\sigma\frac{4(1-\rho)^2-(2\rho^2-\rho-\rho^3)^2}{4(1-\rho)+(2\rho^2-\rho-\rho^3)}$
WS $X_1 - X_2$	$\frac{2}{3}\sigma(1-\rho)$	$\frac{2}{3}\sigma\left((1-\rho_1)-\frac{(\rho_2-\rho_3)^2}{1-\rho_1}\right)$	$\frac{2}{3}\sigma\left((1-\rho^2)-\frac{1}{2}\rho^2(1-\rho^2)\right)$
WS No Baselines	$\frac{2}{3}\sigma(1-\rho)$	$\frac{2}{3}\sigma(1-\rho_1)$	$\frac{2}{3}\sigma\left((1-\rho^2)-\frac{1}{2}\rho^2(1-\rho^2)\right)$

The ratio of variances between the No Baselines mixed model and CFB mixed model is:

$$\sigma\frac{(1-\rho)(1+\rho)}{2+\rho}\frac{1}{\sigma(1-\rho)} = \frac{1+p}{2+p}$$

This ratio is always less than one and thus the No Baselines mixed model is more efficient than CFB. Similarly, the No Baselines mixed model is always more efficient than either of the WS models. Next, the ratio of variances between the No Baselines mixed model and $\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$ is:

$$\sigma\frac{(1-\rho)(1+\rho)}{2+\rho}\frac{5\rho+2}{\sigma(1+2\rho-3\rho^2)} = \frac{5\rho^2+7\rho+2}{3\rho^2+7\rho+2}$$

Based this, it is clear that for a positive ρ, the above ratio is greater than 1 and thus $\mathbf{Y}|\mathbf{X}, \overline{\mathbf{X}}$ is more efficient than the No Baselines mixed model.

APPENDIX D

CHAPTER 4, ROBUST AND NONPARAMETRIC ESTIMATION

D.1. Optimal LCB under OLS

As in Section 4.4, let the LCB regression parameter vary by sequence. Under a squared-loss, the goal is to solve:

$$(\hat{\boldsymbol{\gamma}}, \hat{\beta}, \boldsymbol{a}_\star) = \operatorname{Argmin} \sum_i^s \sum_k^{n_i} \epsilon_{ik}^2 = \operatorname*{Argmin}_{\boldsymbol{\gamma}, \beta, \boldsymbol{a}} \sum_i^s \sum_k^{n_i} (Y_{ik}^\star - \mathbf{W}_{ik}\boldsymbol{\gamma} + \beta_i \mathbf{a}^T \mathbf{X}_{ik})^2$$

Expanding out the formula we get:

$$\sum_i^s \sum_k^{n_i} \epsilon_{ik}^2 = \sum_i^s \sum_k^{n_i} \left((Y_{ik}^\star - \mathbf{W}_{ik}\boldsymbol{\gamma})^2 - 2\beta_i \boldsymbol{a}^T \mathbf{X}_{ik} (Y_{ik}^\star - \mathbf{W}_{ik}\boldsymbol{\gamma}) + \beta_i^2 \boldsymbol{a}^T \mathbf{X}_{ik} \mathbf{X}_{ik}^T \boldsymbol{a} \right) \tag{D.1}$$

Next, assume that $E(\mathbf{a}^T \mathbf{X}_{ik}) = 0$. This done for convenience and in data analysis is achieved by centering the covariate by the sample mean. Then, noting $E(Y_{ik}^\star) = \mathbf{W}_{ik}\boldsymbol{\gamma}$, it follows that:

$$E\left(\sum_i^s \sum_k^{n_i} \epsilon_{ik}^2 \right) = \sum_i^s \sum_k^{n_i} E(\epsilon_{ik}^2) = \sum_i^s \sum_k^{n_i} \left(V(Y_{ik}^\star) - 2\beta_i \mathsf{cov}(Y_{ik}^\star, \boldsymbol{a}^T \mathbf{X}_{ik}) + V(\boldsymbol{a}^T \mathbf{X}_{ik}) \right)$$

Next, given our linear regression model, it follows that:

$$\mathsf{cov}(Y_{ik}^\star, \mathbf{a}^T \mathbf{X}_{ik}) = \mathsf{cov}(\mathbf{W}_{ik}\boldsymbol{\gamma}, \mathbf{a}^T \mathbf{X}_{ik}) + \beta_i \mathsf{cov}(\mathbf{a}^T \mathbf{X}_{ik}, \mathbf{a}^T \mathbf{X}_{ik}) + \mathsf{cov}(\epsilon_{ik}, \mathbf{a}^T \mathbf{X}_{ik})$$

As discussed in Section 4.4, it follows that $\beta_i = \frac{\mathsf{cov}(Y_{ik}^\star, \mathbf{a}^T \mathbf{X}_{ik})}{V(\boldsymbol{a}^T \mathbf{X}_{ik})}$. Putting this all together:

$$\begin{aligned} E\left(\sum_i^s \sum_k^{n_i} \epsilon_{ik}^2 \right) &= \sum_i^s \sum_k^{n_i} \left(V(Y_{ik}^\star) - 2\frac{\mathsf{cov}(Y_{ik}^\star, \boldsymbol{a}^T \mathbf{X}_{ik})^2}{V(\boldsymbol{a}^T \mathbf{X}_{ik})} + \frac{\mathsf{cov}(Y_{ik}^\star, \boldsymbol{a}^T \mathbf{X}_{ik})^2}{V(\boldsymbol{a}^T \mathbf{X}_{ik})^2} V(\boldsymbol{a}^T \mathbf{X}_{ik}) \right) \\ &= \sum_i^s \sum_k^{n_i} \left(V(Y_{ik}^\star) - \frac{\mathsf{cov}(Y_{ik}^\star, \boldsymbol{a}^T \mathbf{X}_{ik})^2}{V(\boldsymbol{a}^T \mathbf{X}_{ik})} \right) \end{aligned}$$

Under normality, $V(Y_{ik}^\star | \mathbf{a}^T \mathbf{X}_{ik}) = V(Y_{ik}^\star) - \frac{\mathsf{cov}(Y_{ik}^\star, \boldsymbol{a}^T \mathbf{X}_{ik})^2}{V(\boldsymbol{a}^T \mathbf{X}_{ik})}$. As expected, the optimal LCB through OLS is the same as under normality.

D.2. Optimal LCB: Asymptotic Equivalence between Normality and the T-Distribution

Without loss of generality, let $n = n_i$ for all i. For $n \to \infty$, by the weak law of large numbers (WLLN) note first that:

$$\frac{1}{n}\sum_{i}^{s}\sum_{k=1}^{n_i}(\mathbf{a}\mathbf{X}_{ik} - E(\mathbf{a}^T\mathbf{X}_{ik}))^2 \xrightarrow{p} E\left((\mathbf{a}\mathbf{X}_{ik} - E(\mathbf{a}^T\mathbf{X}_{ik}))^2\right) = V(\mathbf{a}\mathbf{X}_{ik})$$

Notably, the WLLN requires that $V(\mathbf{a}\mathbf{X}_{ik})$ exists. For a multivariate T-distribution with v degrees of freedom, $V(\mathbf{a}\mathbf{X}_{ik})$ exists for $v > 2$. Consider the sample mean for the conditional variance under a multivariate T-distribution with $v > 2$ degrees of freedom.

$$\frac{1}{n}\sum_{i}^{s}\sum_{k=1}^{n_i}V(Y_{ik}^{\star}|\mathbf{a}^T\mathbf{X}_{ik})_T = \frac{1}{n}\sum_{i}^{s}\sum_{k=1}^{n_i}\frac{v + (\mathbf{a}\mathbf{X}_{ik} - E(\mathbf{a}^T\mathbf{X}_{ik}))^2V(\mathbf{a}^T\mathbf{X}_{ik})^{-1}}{v+1}\left(V(Y_{ik}^{\star}) - \frac{\text{cov}(Y_{ik}^{\star}, \mathbf{a}^T\mathbf{X}_{ik})^2}{V(\mathbf{a}^T\mathbf{X}_{ik})}\right)$$

Note that v, $\text{cov}(Y_{ik}^{\star}, \mathbf{a}^T\mathbf{X}_{ik})$, $V(\mathbf{a}^T\mathbf{X}_{ik})$, and $V(Y_{ik}^{\star})$ are constants. Next, the continuous mapping theorem, or Mann-Wald mapping theorem, states that for $X_n \xrightarrow{p} \mu$, $g(X_n) \xrightarrow{p} g(\mu)$ for some continuous function g. Specifically, $g : R \to R'$ is a Borel function who set of D discontinuities is such that $\{w : X(w) \in D\} \in F$ and $P(X \in D) = 0$ where F refers to the sample space. In other words, if a sequence of random variables (X_n) converge in probability to some mean μ, then the transformed sequence of random variables $(g(X_n))$ converges in probability to $g(\mu)$. Thus, it follows that:

$$\begin{aligned}\frac{1}{n}\sum_{i}^{s}\sum_{k=1}^{n_i}V(Y_{ik}^{\star}|\mathbf{a}^T\mathbf{X}_{ik})_T &\xrightarrow{p} \frac{1}{n}\sum_{i}^{s}\sum_{k=1}^{n_i}\frac{v + V(\mathbf{a}^T\mathbf{X}_{ik})V(\mathbf{a}^T\mathbf{X}_{ik})^{-1}}{v+1}\left(V(Y_{ik}^{\star}) - \frac{\text{cov}(Y_{ik}^{\star}, \mathbf{a}^T\mathbf{X}_{ik})^2}{V(\mathbf{a}^T\mathbf{X}_{ik})}\right) \\ &= \frac{1}{n}\sum_{i}^{s}\sum_{k=1}^{n_i}\left(V(Y_{ik}^{\star}) - \frac{\text{cov}(Y_{ik}^{\star}, \mathbf{a}^T\mathbf{X}_{ik})^2}{V(\mathbf{a}^T\mathbf{X}_{ik})}\right)\end{aligned}$$

Thus, the conditional variances for normality and a T-distribution are asymptotically equivalent and the optimal LCBs for both distributions are asymptotically identical.

D.3. Optimal LCB under EP

Here we find the optimal LCB under the general covariance assumption of $\text{cov}(X_{ijk}, Y_{ijk}) = \sigma_1^2\rho_1^{\star}$ for $j = j$ and $\text{cov}(X_{ijk}, Y_{ij'k}) = \sigma_2^2\rho_2^{\star}$ for $j \neq j'$. One example of this is the EP covariance structure.

Without loss of generality, assume that $Y_{ik}^{\star} = Y_{i1k} - Y_{i2k}$. Then:

$$\beta_i = \frac{\text{cov}(Y_{i1k} - Y_{i2k}, \mathbf{a}^T\mathbf{X}_{ik})}{V(\mathbf{a}^T\mathbf{X}_{ik})} = \frac{\sum_j^p a_j \text{cov}(Y_{i1k}, X_{ijk}) - \sum_j^p a_j \text{cov}(Y_{i2k}, X_{ijk})}{V(\mathbf{a}^T\mathbf{X}_{ik})} \tag{D.2}$$

$$= \frac{a_1\sigma_1^2\rho_1^{\star} + a_2\sigma_2^2\rho_2^{\star} - a_1\sigma_2^2\rho_2^{\star} - a_2\sigma_1^2\rho_1^{\star}}{V(\mathbf{a}^T\mathbf{X}_{ik})} \tag{D.3}$$

$$= \frac{\text{cov}(Y_{i1k} - Y_{i2k}, a_1X_{i1k} + a_2X_{i2k})}{V(\mathbf{a}^T\mathbf{X}_{ik})} = \frac{(a_1 - a_2)(\sigma_1^2\rho_1^{\star} - \sigma_2^2\rho_2^{\star})}{V(\mathbf{a}^T\mathbf{X}_{ik})} \tag{D.4}$$

First, since $\text{cov}(Y_{i1k} - Y_{i2k}, \mathbf{a}^T\mathbf{X}_{ik}) = \text{cov}(Y_{i1k} - Y_{i2k}, a_1X_{i1k} + a_2X_{i2k})$, only the baselines corresponding to the within-subject contrast contribute information. Second, this general covariance scenario means that $\beta_i = \beta$ for all i. Third, $|\text{cov}(Y_{i1k} - Y_{i2k}, \mathbf{a}^T\mathbf{X}_{ik})|$ is maximized at $a_1 = -a_2$. This indicates that the LCB $X_{i1k} - X_{i2k}$ offers the most information. Overall, this means that for some within subject contrast $Y_{ik}^{\star} = \mathbf{b}_i\mathbf{Y}_{ik}$, the optimal LCB is $\mathbf{b}_i\mathbf{X}_{ik}$, or XDIFF. In the case where $\mathbf{b}_i\mathbf{Y}_{ik} = Y_{i1k} - Y_{i2k}$, the optimal LCB is $X_{i1k} - X_{i2k}$.

D.4. Model Selection and Bootstrapping

Here we describe further details on bootstrapping for our Min-P model selection procedure. For a set of $m = 1, ..., M$ models, obtain a set of estimates $\{\hat{\tau}_1, ..., \hat{\tau}_M\}$ with estimated standard errors $\{SE(\hat{\tau}_1), ..., SE(\hat{\tau}_M\})$. Under the null hypothesis of $H_0 : \tau_m = 0$, obtain p-values $\{p(\hat{\tau}_1), ..., p(\hat{\tau}_M)\}$. The model and corresponding estimate with the smallest p-value is then chosen. Formally, the Min-P estimate is defined as:

$$\hat{\theta} = \{\hat{\tau}_m : p(\hat{\tau_m}) = \text{min}\{p(\hat{\tau}_1), ..., p(\hat{\tau}_M)\}\} \tag{D.5}$$

For crossover designs, bootstrap resampling is done within sequence groups and at the subject level. For example, in the 2×2 design, subjects are resampled with replacement within sequence AB and BA separately. This is similar to the two-group bootstrap resampling described by Hesterberg (Hesterberg, 2015). Specific details on the resampling for the 2×2 and the 3×3 designs are given in Section 4.6.

For the remainder of this section, bootstrap resampling is done within sequence groups and $b = 1, ..., B$ bootstrap resamples are generated. Define $\hat{\theta}_b$ as the Min-P estimate ($\hat{\theta}$) for bootstrap re-

sample b. The bagged or smoothed estimator of the Min-P estimate is:

$$\hat{\theta}_S = \frac{1}{B}\sum_{1}^{B}\hat{\theta}_b$$

Bootstrap Standard Errors

The standard error of the Min-P estimate can be estimated using two different variants of bootstrap resampling. First, the standard bootstrap approach:

$$SE(\hat{\theta})_B = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}_b - \hat{\theta}_S)^2} \tag{D.6}$$

Alternatively, as Efron showed, we can work directly with the smoothed estimator $\hat{\theta}_S$ (Efron, 2014). Efron used the nonparametric delta method approximation to estimate the standard error of the smoothed estimate, SE($\hat{\theta}_S$). Generally, using a smoothed or bagged estimate reduces the variability (Friedman and Hall, 2007). Indeed, Efron showed analytically that SE($\hat{\theta}$)$_S$ is more efficient than SE($\hat{\theta}_S$) (Efron, 2014).

Bootstrap Confidence Intervals

There are a number of methods for obtaining confidence intervals under bootstrap resampling. Hesterberg and Efron separately give a good overview for various methods (Efron, 1987; Hesterberg, 2015). Assume the goal is to estimate a two-sided CI of the Min-P estimate $\hat{\theta}$ at the $(1-\alpha)$ level. Let the nonparametric bootstrap cumulative distribution function be defined as:

$$\hat{G}(s) = \#\{\hat{\theta}_b < s\}/B \tag{D.7}$$

The bootstrap percentile CI is then defined as::

$$CI_{pct} = [\hat{G}^{-1}(\alpha/2), \hat{G}^{-1}(1-\alpha/2)] \tag{D.8}$$

Note that $\hat{G}^{-1}(\alpha/2)$ is simply the $(\alpha/2)$-quantile, for example 2.5%, of the empirical bootstrap distribution of $\hat{\theta}_b$. The Bias-Corrected and Accelerated (BCa) CI improves on the percentile approach by correcting for potential bias in the parameter estimate ($\hat{\theta}$) and acceleration. The acceleration

describes how much the standard error of a parameter estimate changes with respect to changes in the parameter estimate. Explicitly:

$$CI_{BCa} = [\hat{G}^{-1}(\Phi(z(\alpha/2))), \hat{G}^{-1}(\Phi(z(1-\alpha/2)))] \tag{D.9}$$

where $z[\alpha/2] = z_0 + (z_0 + z^{\alpha/2})/(1 + a(z_0 + z^{\alpha/2}))$. $z_0 = \Phi^{-1}(\hat{G}(\hat{\theta}))$ is the bias, a is the acceleration, and $z^{\alpha} = \Phi^{-1}(\alpha)$. Note that for $a = z_0 = 0$ (no bias, no acceleration), the BCa CI is identical to the percentile CI. Acceleration is approximated by a measure of skewness (Efron, 1979). Further, standard software, such as the boot package in R, can be used to construct BCa intervals. More recently, Efron proposed a BCa interval based on the smoothed bootstrap estimate (Efron, 2014).

The BCa CI has various advantages. First, for bias z_0 and acceleration a, the BCa CI is based on the general assumption that for some monotone function g (Diciccio and Tibshirani, 1987):

$$g(\hat{\theta}) - g(\theta) \approx N(-z_0(1 + ag(\theta)), (1 + ag(\theta))^2)$$

In contrast, a typical CI is based on the more restrictive assumption that $(\hat{\theta} - \theta)/SE(\hat{\theta}) \approx N(0, 1)$. Further, while the BCa CI is rooted in this general transformation assumption, the interval can be constructed without knowing the transformation g (Diciccio and Tibshirani, 1987). Second, the BCa CI is second order accurate, meaning that coverage probabilities differ from the nominal values by $O(n^{-1})$. The percentile method is first order accurate, $O(n^{-1/2})$. This means that the BCa CI will converge to correct coverage faster than the percentile method.

Lastly, for small sample sizes, bootstrap confidence intervals tend to be too narrow (Hesterberg, 2015). To correct this, Hesterberg proposed the Expanded Interval where the critical value α is adjusted. If the bootstrap distribution of $\hat{\theta}$ is normal, then:

$$G^{-1}(\alpha/2) = \hat{\theta} - z_{\alpha/2}\hat{\sigma}$$

where $\hat{\sigma}^2 = \frac{1}{B}\sum_i^B (\hat{\theta}_B - \hat{\theta}_S)^2$. Next, find an adjusted α' such that:

$$\begin{aligned} G^{-1}(\alpha'/2) &\approx \hat{\theta} - z_{\alpha'/2}\hat{\sigma} \\ &= \hat{\theta} - t_{\alpha/2,n-1}s \end{aligned}$$

where $s^2 = \frac{1}{B-1}\sum_i^B(\hat{\theta}_B - \hat{\theta}_S)^2$. Then, since $\frac{s}{\hat{\sigma}} = \sqrt{n/(n-1)}$, we get:

$$\alpha'/2 = \Phi(-\sqrt{n/(n-1)}t_{\alpha/2,n-1})$$

These adjusted critical values are then used to construct percentile and BCa intervals. For small sample sizes, this provides better coverage under normal populations and other distributions by avoiding CIs that are too narrow Hesterberg, 2015.

D.5. Parametric and Nonparametric Comparisons: 2×2 and 3×3 Simulations

Table D.1: Parametric and Nonparametric Comparisons, 2×2 Simulations: Type I Error

		Normal			**T-Dist**		
Truth	**Method**	N=12	N=20	N=32	N=12	N=20	N=32
Σ_{CS}	XDIFF (OLS)	5.08	4.26	4.82	4.60	4.10	4.64
Σ_{CS}	XDIFF (R-est)	5.26	4.48	4.70	4.46	4.38	4.16
Σ_{CS}	Wilcoxon: Ys	4.24	4.92	4.16	3.98	4.82	4.22
Σ_{CS}	Rank ANCOVA: XDIFF	4.30	4.76	4.44	4.28	4.76	4.26
Σ_{CS}	CFB	5.00	4.45	4.70	3.65	3.95	4.50
Σ_{CS}	Min-P1	4.78	4.10	4.80	4.30	4.24	4.74
Σ_{CS}	Min-P2	4.18	3.84	4.40	4.12	3.90	4.28
Σ_{EP}	XDIFF (OLS)	5.30	4.56	5.00	4.46	4.06	4.62
Σ_{EP}	XDIFF (R-est)	5.46	4.48	4.82	4.48	4.00	4.36
Σ_{EP}	Wilcoxon: Ys	4.70	4.72	4.50	4.50	4.96	4.36
Σ_{EP}	Rank ANCOVA: XDIFF	4.40	4.52	4.38	4.10	4.58	4.32
Σ_{EP}	CFB	5.00	4.45	4.70	3.65	3.95	4.50
Σ_{EP}	Min-P1	4.92	4.36	4.78	4.20	4.26	4.74
Σ_{EP}	Min-P2	4.58	4.20	4.66	4.28	4.06	4.66
Σ_{AR}	XDIFF (OLS)	4.90	4.84	4.88	4.38	4.04	4.28
Σ_{AR}	XDIFF (R-est)	5.42	4.98	4.94	4.36	4.68	5.04
Σ_{AR}	Wilcoxon: Ys	4.16	4.86	4.70	3.76	5.02	4.38
Σ_{AR}	Rank ANCOVA: XDIFF	4.12	4.90	4.52	3.94	5.04	4.56
Σ_{AR}	CFB	4.90	4.95	5.30	4.25	3.95	4.50
Σ_{AR}	Min-P1	4.60	4.84	4.98	4.52	4.58	5.08
Σ_{AR}	Min-P2	4.26	4.54	4.92	4.22	4.10	4.62

Notes: Entries are in brackets if the type I error is two SEs above 5% ($> 5.61\%$) based on 5,000 simulations. Wilcoxon refers to the Wilcoxon rank sum test with no baseline adjustment. Rank ANCOVA regresses the ranks of the outcomes against the ranks of XDIFF and then uses the residuals in a Wilcoxon rank sum test. XDIFF (OLS, R-est) uses XDIFF as a covariate in the given regression model. CFB uses change scores with no baseline covariate. Min-P1 chooses between XDIFF (OLS, R-est). Min-P2 chooses between No Baselines and XDIFF (OLS, R-est). All methods are discussed in Section 4.6 and Table 4.2. CS = Compound Symmetry; EP = Equipredictability; AR = Auto-regressive(1). N refers to total sample size.

Table D.2: Parametric and Nonparametric Comparisons, 2×2 Simulations: Power

		Normal			T-Dist		
Truth	**Method**	N=12	N=20	N=32	N=12	N=20	N=32
Σ_{CS}	XDIFF (OLS)	**75.2**	**77.7**	**78.5**	85.8	86.4	86.0
Σ_{CS}	XDIFF (R-est)	70.4	74.3	76.3	**88.3**	92.4	94.9
Σ_{CS}	Wilcoxon: Ys	**74.1**	**78.9**	77.3	83.3	**92.7**	**94.7**
Σ_{CS}	Rank ANCOVA: XDIFF	68.1	75.5	75.5	79.6	91.5	94.2
Σ_{CS}	CFB	50.8	50.7	50.6	65.5	63.1	61.5
Σ_{CS}	Min-P1	70.8	75.6	77.6	87.3	92.0	94.3
Σ_{CS}	Min-P2	73.1	76.7	**78.0**	**89.0**	**93.2**	**94.9**
Σ_{EP}	XDIFF (OLS)	**85.3**	**88.5**	**88.6**	91.4	92.6	92.7
Σ_{EP}	XDIFF (R-est)	81.7	85.8	86.7	**93.9**	**96.8**	**98.2**
Σ_{EP}	Wilcoxon: Ys	74.3	79.1	77.6	82.6	92.7	94.7
Σ_{EP}	Rank ANCOVA: XDIFF	75.6	84.0	84.0	83.8	95.3	97.7
Σ_{EP}	CFB	80.1	80.0	79.8	85.7	85.3	83.8
Σ_{EP}	Min-P1	**82.3**	**87.0**	**88.1**	92.5	96.3	**97.7**
Σ_{EP}	Min-P2	82.1	85.6	87.1	**93.3**	**96.6**	97.6
Σ_{AR}	XDIFF (OLS)	**80.8**	**83.6**	**84.3**	89.2	89.5	89.4
Σ_{AR}	XDIFF (R-est)	76.4	80.5	82.1	**91.2**	**94.8**	**96.3**
Σ_{AR}	Wilcoxon: Ys	74.9	78.6	77.2	82.7	92.5	94.0
Σ_{AR}	Rank ANCOVA: XDIFF	70.6	79.6	80.2	80.3	93.3	96.1
Σ_{AR}	CFB	70.9	70.8	71.2	78.8	78.5	77.8
Σ_{AR}	Min-P1	77.6	**81.8**	**83.7**	89.9	94.2	95.8
Σ_{AR}	Min-P2	**78.0**	81.5	82.9	**91.1**	**94.8**	**96.1**

Note: Values (Power %) are shown in bold if method yields the highest or second highest power in that sample size/covariance structure combination without type I error inflation. Entries are in brackets if under the same scenario, but under the null hypothesis, the type I error is two SEs above 5% ($> 5.61\%$) based on 5,000 simulations. Wilcoxon refers to the Wilcoxon rank sum test with no baseline adjustment. Rank ANCOVA regresses the ranks of the outcomes against the ranks of XDIFF and then uses the residuals in a Wilcoxon rank sum test. XDIFF (OLS, R-est) uses XDIFF as a covariate in the given regression model. CFB uses change scores with no baseline covariate. Min-P1 chooses between XDIFF (OLS, R-est). Min-P2 chooses between No Baselines and XDIFF (OLS, R-est). All methods are discussed in Section 4.6 and Table 4.2. CS = Compound Symmetry; EP = Equipredictability; AR = Auto-regressive(1). N refers to total sample size.

Table D.3: Parametric and Nonparametric Comparisons, 3×3 Simulations: Type I Error

		Normal			T-Dist		
Truth	**Method**	N=18	N=24	N=30	N=18	N=24	N=30
Σ_{CS}	XDIFF (OLS)	5.56	5.38	5.40	4.56	4.58	4.70
Σ_{CS}	XDIFF (R-est)	[5.82]	5.22	[5.72]	4.74	4.82	4.84
Σ_{CS}	LCB (OLS)	[5.78]	5.00	[5.66]	5.52	4.68	5.36
Σ_{CS}	LCB (R-est)	5.14	4.54	5.22	4.68	3.82	5.18
Σ_{CS}	Ohrvik's Aligned Rank	5.50	5.30	5.30	5.12	4.96	5.52
Σ_{CS}	Rank ANCOVA: XDIFF	5.40	5.60	5.36	5.50	5.46	5.48
Σ_{CS}	CFB	5.46	5.26	4.52	5.32	4.90	4.76
Σ_{CS}	Min-P1	[5.80]	4.92	5.60	5.58	4.70	5.04
Σ_{CS}	Min-P2	5.28	4.74	5.20	4.98	4.32	4.50
Σ_{CS}	Min-P3	4.98	4.26	4.78	4.32	3.54	4.40
Σ_{EP}	XDIFF (OLS)	4.54	4.70	4.82	3.72	3.76	3.84
Σ_{EP}	XDIFF (R-est)	5.34	4.66	4.94	4.40	4.16	4.28
Σ_{EP}	LCB (OLS)	[5.90]	5.10	5.52	5.40	4.66	5.28
Σ_{EP}	LCB (R-est)	4.86	4.68	5.40	5.00	4.44	5.08
Σ_{EP}	Ohrvik's Aligned Rank	5.34	5.24	5.22	5.54	5.04	5.06
Σ_{EP}	Rank ANCOVA: XDIFF	5.24	5.12	4.96	5.04	5.04	4.96
Σ_{EP}	CFB	5.30	5.22	4.90	5.52	5.32	4.74
Σ_{EP}	Min-P1	5.34	4.44	4.74	4.56	3.92	4.18
Σ_{EP}	Min-P2	5.18	4.34	4.74	4.50	3.80	4.20
Σ_{EP}	Min-P3	4.88	4.00	4.70	4.54	3.82	4.22
Σ_{AR}	XDIFF (OLS)	4.62	4.54	4.68	3.92	3.52	3.90
Σ_{AR}	XDIFF (R-est)	5.18	4.68	4.56	4.48	3.76	4.34
Σ_{AR}	LCB (OLS)	[6.20]	5.40	[5.62]	5.60	5.08	5.34
Σ_{AR}	LCB (R-est)	5.50	4.74	5.44	4.90	4.14	5.46
Σ_{AR}	Ohrvik's Aligned Rank	5.54	5.22	5.20	5.30	4.68	5.30
Σ_{AR}	Rank ANCOVA: XDIFF	5.34	4.98	4.62	4.94	4.82	4.68
Σ_{AR}	CFB	[6.06]	5.14	5.06	5.68	5.48	5.34
Σ_{AR}	Min-P1	5.08	4.46	4.62	4.44	3.48	4.14
Σ_{AR}	Min-P2	4.94	4.20	4.50	4.52	3.46	4.06
Σ_{AR}	Min-P3	5.10	3.92	4.84	4.52	3.36	4.28

Note: Entries are in brackets if the type I error is two SEs above 5% ($> 5.61\%$) based on 5,000 simulations. Ohrvik refers to the Ohrvik aligned rank test with no baseline adjustment. Rank ANCOVA regresses the ranks of the aligned outcomes against the ranks of XDIFF and then uses the residuals in a Wilcoxon rank sum test. XDIFF (OLS, R-est) uses XDIFF as a covariate in the given regression model. LCB (OLS, R-est) estimates the optimal LCB in the given regression model. CFB uses change scores with no baseline covariate. Min-P1 chooses between XDIFF (OLS, R-est). Min-P2 chooses between No Baselines and XDIFF (OLS, R-est). Min-P3 chooses between No Baselines, XDIFF, and LCB (OLS, R-est). All methods are discussed in Section 4.6 and Table 4.2. CS = Compound Symmetry; EP = Equipredictability; AR = Auto-regressive(1). N refers to total sample size.

Table D.4: Parametric and Nonparametric Comparisons, 3×3 Simulations: Power

		Normal			T-Dist		
Truth	**Method**	N=18	N=24	N=30	N=18	N=24	N=30
Σ_{CS}	XDIFF (OLS)	**80.3**	**79.3**	**79.4**	87.4	85.8	87.1
Σ_{CS}	XDIFF (R-est)	[77.1]	76.8	[77.4]	92.4	**92.9**	**94.9**
Σ_{CS}	LCB (OLS)	[74.1]	73.9	[75.9]	87.6	86.3	87.2
Σ_{CS}	LCB (R-est)	67.0	68.3	72.1	87.6	88.6	91.9
Σ_{CS}	Ohrvik's Aligned Rank	**82.1**	**80.3**	**79.9**	**95.1**	**94.9**	**95.6**
Σ_{CS}	Rank ANCOVA: XDIFF	78.7	78.0	77.8	91.8	93.0	94.8
Σ_{CS}	CFB	50.6	51.0	51.4	68.2	65.6	64.9
Σ_{CS}	Min-P1	[79.8]	77.1	78.2	92.2	91.9	94.0
Σ_{CS}	Min-P2	79.9	76.7	78.4	**92.8**	92.2	94.0
Σ_{CS}	Min-P3	72.7	72.3	74.9	91.0	90.6	93.2
Σ_{EP}	XDIFF (OLS)	**88.0**	**87.5**	**87.3**	91.6	90.9	92.0
Σ_{EP}	XDIFF (R-est)	84.8	85.3	85.1	95.0	**96.0**	**97.2**
Σ_{EP}	LCB (OLS)	[78.9]	79.0	80.2	90.5	90.0	90.9
Σ_{EP}	LCB (R-est)	72.7	73.9	76.8	90.0	91.8	94.5
Σ_{EP}	Ohrvik's Aligned Rank	82.1	80.1	79.9	**95.1**	**95.3**	95.7
Σ_{EP}	Rank ANCOVA: XDIFF	83.6	83.8	84.1	94.3	**95**.3	**96.8**
Σ_{EP}	CFB	80.2	79.3	78.9	89.6	88.8	88.1
Σ_{EP}	Min-P1	**86.9**	**85.7**	**86.2**	**95.1**	95.2	**96.8**
Σ_{EP}	Min-P2	85.6	83.5	84.6	**95.1**	95.0	96.5
Σ_{EP}	Min-P3	80.1	79.4	81.9	93.6	94.3	95.9
Σ_{AR}	XDIFF (OLS)	**89.0**	**88.0**	**87.5**	92.3	91.7	92.2
Σ_{AR}	XDIFF (R-est)	86.4	85.7	85.5	**96.1**	**96.4**	97.3
Σ_{AR}	LCB (OLS)	[88.0]	**87.9**	[88.2]	95.1	94.6	95.3
Σ_{AR}	LCB (R-est)	82.8	83.6	85.9	95.0	96.3	**97.4**
Σ_{AR}	Ohrvik's Aligned Rank	83.3	81.8	80.6	95.6	96.2	95.9
Σ_{AR}	Rank ANCOVA: XDIFF	85.1	84.6	84.4	94.9	96.2	97.1
Σ_{AR}	CFB	[80.4]	80.6	79.2	90.6	89.1	89.0
Σ_{AR}	Min-P1	**88.7**	86.6	86.2	95.9	95.9	97.0
Σ_{AR}	Min-P2	86.9	84.8	84.9	95.9	95.8	96.5
Σ_{AR}	Min-P3	86.9	86.1	**86.9**	**96.8**	**97.1**	**97.9**

Note: Values (Power %) are shown in bold if method yields the highest or second highest power in that sample size/covariance structure combination without type I error inflation. Entries are in brackets if under the same scenario, but under the null hypothesis, the type I error is two SEs above 5% ($> 5.61\%$) based on 5,000 simulations. Ohrvik refers to the Ohrvik aligned rank test with no baseline adjustment. Rank ANCOVA regresses the ranks of the aligned outcomes against the ranks of XDIFF and then uses the residuals in a Wilcoxon rank sum test. XDIFF (OLS, R-est) uses XDIFF as a covariate in the given regression model. LCB (OLS, R-est) estimates the optimal LCB in the given regression model. CFB uses change scores with no baseline covariate. Min-P1 chooses between XDIFF (OLS, R-est). Min-P2 chooses between No Baselines and XDIFF (OLS, R-est). Min-P3 chooses between No Baselines, XDIFF, and LCB (OLS, R-est). All methods are discussed in Section 4.6 and Table 4.2. CS = Compound Symmetry; EP = Equipredictability; AR = Auto-regressive(1). N refers to total sample size.

APPENDIX E

SOFTWARE

In addition to the code examples throughout the dissertation, code for all simulations and methods can be found in the following GITHUB page:

`https://github.com/thomasjemielita/Crossover-Research`

BIBLIOGRAPHY

Armitage, P and Hills, M (1982). The two-period cross-over trial. *The Statistician* 31, 119–131.

Bellavance, F and Tardif, S (1995). A non-parametric approach to the analysis of three-treatment three-period cross-over designs. *Biometrika* 82, 865–875.

Chen, X, Meng, Z, and Zhang, J (2012). Handling of baseline measurements in the analysis of crossover trials. *Statistics in Medicine* 31, 1791–1803.

Chen, X and Wei, L (2003). A comparison of recent methods for the analysis of small-sample cross-over studies. *Statistics in Medicine* 22, 2821–2833.

Chi, E (1991). Recovery of Inter-block Information in Cross-over Trials. *Statistics in Medicine* 10, 1115–1122.

Chinchilli, V and Esinhart, J (1996). Design and analysis of intra-subject variability in cross-over experiments. *Statistics in Medicine* 15, 1619–16934.

Correa, J and Bellavance, F (2001). Power comparison of robust approximate and non-parametric tests for the analysis of cross-over trials. *Statistics in Medicine* 20, 1185–1196.

Diciccio, T and Tibshirani, R (1987). Bootstrap confidence intervals and bootstrap approximations. *Journal of the American Statistical Association* 82, 163–170.

Ding, P (2016). On the conditional distribution of the multivariate t distribution. *The American Statistician* 70, 293–295.

Efron, B (1979). Another Look at the Jackknife. *The Annals of Statistics* 7, 1–26.

Efron, B (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association* 82, 171–185.

Efron, B (2014). Estimation and Accuracy after Model Selection. *Journal of the American Statistical Association* 109, 991–1007.

Fox, J and Weisberg, H (2011). *An R Companion to Applied Regression*. Thousand Oakes, CA: Sage.

Friedman, J and Hall, P (2007). On bagging and nonlinear estimation. *Journal of Statistical Planning and Inference* 137, 669–683.

Gurka, M, Edwards, L, and Muller, K (2011). Avoiding bias in mixed model inference for fixed effects. *Statistics in Medicine* 30, 2696–2707.

Hesterberg, T (2015). What Teachers Should Know about the Bootstrap: Resampling in the Undergraduate Statistics Curriculum. *The American Statistician* 69, 371–386.

Hettmansperger, T and Mckean, J (2010). *Robust Nonparametric Statistical Methods*. London: Chapman and Hall.

Hettsmansperger, T and Mckean, J (1983). A geometric interpretation of inferences based on ranks in the linear models. *Journal of the American Statistical Association* 78, 885–893.

Hills, M and Armitage, P (1979). The two-period cross-over clinical trial. *British Journal of Clinical Pharmacology* 8, 7–20.

Huber, P (1964). Robust Estimation of a Location Parameter. *Annals of Mathematical Statistics* 35, 73–101.

Huber, P (1973). Robust regression: asymptotics, conjectures, and monte carlo. *Annals of Statistics* 1, 799–821.

Hurvich, C and Tsai, C (1990). The impact of model selection on inference in linear regression. *The American Statistician* 44, 214–217.

Hurvich, C and Tsai, C (1989). Regression and time series model selection in small samples. *Biometrika* 76, 297–307.

Jaeckel, L (1972). Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals. *The Annals of Mathematical Statistics* 43, 1449–1458.

Jemielita, T, Putt, M, and Mehrotra, D (2016). Improved Power in Crossover Designs through Linear Combinations of Baselines. *Statistics in Medicine* 35, 5625–5641.

Jones, B and Kenward, M (2003). *Design and Analysis of Cross-over Trials*. London: Chapman and Hall.

Judkins, D and Porter, K (2015). Robustness of Ordinary Least Squares in Randomized Clinical Trials. *Statistics in Medicine* 35, 1763–1773.

Jureckova, J (1971). Nonparametric Estimate of Regression Coefficients. *The Annals of Mathematical Statistics* 42, 1328–1338.

Kenward, M and Jones, B (1987). The analysis of data from 2x2 cross-over trials with baseline measurements. *Statistics in Medicine* 6, 911–926.

Kenward, M and Roger, J (1997). Small sample inference for fixed effects estimators from restricted maximum likelihood. *Biometrics* 53, 983–997.

Kenward, M and Roger, J (2010). The use of baseline covariates in crossover studies. *Biostatistics* 11, 1–17.

Kloke, J, Mckean, J, and Rashid, M (2009). Rank-based estimation and associated inferences for linear models with cluster correlated errors. *Journal of the American Statistical Association* 104, 384–390.

Kloke, J and McKean, J (2012). Rfit: Rank-based estimation for linear models. *The R Journal* 4, 57–64.

Koch, G (1972). The use of nonparametric methods in the statistical analysis of the two-period change-over design. *Biometrics* 28, 577–584.

Koul, H, Sievers, G, and Mckean, J (1987). An estimator of the scale parameter for the rank analysis of linear models under general score functions. *Scandinavian Journal of Statistics* 14, 131–141.

Lavange, L and Koch, G (2006). Rank score tests. *Circulation* 114, 2528–2533.

Mckean, J and Sheather, S (1991). Small Sample Properties of Robust Analyses of Linear Models Based on R-Estimates: A Survey. *Directions in Robust Statistics and Diagnostics* The IMA Volumes in Mathematics and Its Applications, 1–19.

Mehrotra, D (2014). A recommended analysis for 2x2 Crossover Trials with baseline measurements. *Pharmaceutical Statistics* 13, 376–387.

Metcalfe, C (2010). The analysis of cross-over trials with baseline measurements. *Statistics in Medicine* 29, 3211–3218.

Ohrvik, J (1998). Nonparametric methods in crossover trials. *Biometrical Journal* 40, 771–789.

Putt, M and Chinchilli, V (2004). Nonparametric Approaches to the Analysis of Crossover Studies. *Statistical Science* 19, 712–719.

Quade, D (1967). Rank analysis of covariance. *Journal of the American Statistical Association* 62, 1187–1200.

Schaalje, G, Mcbride, J, and Fellingham, G (2002). Adequacy of approximations to distributions of test statistics in complex mixed linear models. *Journal of Agricultural, Biological and Environmental Statistics* 7, 512–524.

Tsai, K and Patel, H (1996). Robust procedures for analysing a two-period cross-over design with baseline measurements. *Statistics in Medicine* 15, 117–126.

Tudor, G and Koch, G (1994). Review of nonparametric methods for the analysis of crossover studies. *Statistical Methods in Medical Research* 3, 345–381.

Vonesh, E and Chinchilli, V (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. New York: Marcel Dekker.

Yan, Z (2012). The impact of baseline covariates on the efficiency of statistical analyses of crossover designs. *Statistics in Medicine* 32, 956–963.