ABSTRACT

HOLM, KATHLEEN JENNIFER. Comparison of Optimal Design Methods in Inverse Problems. (Under the direction of H.T. Banks.)

This project was initiated with an investigation by Banks et al [6] into the use of traditional and generalized sensitivity functions for experimental design. The authors developed tools for finding the best duration of the data ($[0, T]$), as well as the regions of importance for collecting data based on information from the sensitivity functions. These investigations led to the proposal of a new optimal design method. This work will provide a necessary testing of this new optimal design method with a comparison to more established optimal design methods.

Typical optimal design methods for inverse or parameter estimation problems are designed to choose optimal sampling distributions through minimization of a specific cost function related to the resulting error in parameter estimates. It is hoped that the inverse problem will produce parameter estimates with increased accuracy using data collected according to the optimal sampling distribution. Motivation of the optimal design cost functionals is presented, drawing from linear algebra, geometry and statistics. We formulate the classical optimal design problem in the context of general optimization problems over distributions of sampling times. We present a new Prohorov metric based theoretical framework that permits one to treat succinctly and rigorously any optimal design criteria based on the Fisher Information Matrix (FIM). A fundamental approximation theory is also included in this framework. A new optimal design, SE-optimal design (*standard error* optimal design), is then introduced in the context of this framework. We compare this new design criteria with the more traditional D-optimal and E-optimal designs. The optimal sampling distributions from each design are used to compute and compare parameter estimates and standard errors. Given an optimal mesh, the standard errors for parameters are computed using asymptotic theory or bootstrapping. We use three examples to illustrate ideas: the Verhulst-Pearl logistic population model [8], the standard harmonic oscillator model [8] and a popular glucose regulation model [11, 17, 32]. A Monte Carlo analysis provides a comparison of the optimal design methods over a broad spectrum of simulated data sets. We present an overview of modeling autocorrelated data and formulate the inverse problem for the estimation of model and variance parameters. The optimal design methods are compared in the context of autocorrelated data. The consequences of making an incorrect assumption about the independence of the data is investigated for the optimal design time points.

The goal of this project is to test this newer optimal design method, SE-optimal, against the more traditional methods, D-optimal and E-optimal. As we conduct this comparison in a variety of examples (in terms of mathematical models, statistical models, and variation on

the constraint implementation) we hope to provide a robust analysis, as well as provide the exposure of optimal design methods to a larger community of researchers. The use of optimal experimental design can increase efficiency and the accuracy of parameter estimates, and can provide more information on the problem being investigated.

Comparison of Optimal Design Methods in Inverse Problems

by
Kathleen Jennifer Holm

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Biomathematics

Raleigh, North Carolina

2011

APPROVED BY:

Marie Davidian	Ralph Smith
Hien Tran	H.T. Banks Chair of Advisory Committee

DEDICATION

This work is dedicated to my parents, Doug and Sandy.
Thanks for inspiring me to be my best.

BIOGRAPHY

The author was born in Tucson, Arizona on June 5th, 1982. She spent the first 18 years of her life in Tucson, attending the local public schools through high school. Upon graduating from University High School, she studied for her undergraduate degree at Colorado State University in Fort Collins, Colorado. She earned a bachelors of science degree in mathematics with a cum laude distinction. After graduation, she moved back to Tucson, Arizona, where she worked at International Business Machines (IBM) with the Storage Systems Performance group. Immediately following her work at IBM, she studied at the University of Arizona earning a masters of science degree in applied mathematics in May 2007. She moved to Raleigh, North Carolina to gain a doctor of philosophy degree at North Carolina State University. She earned a doctor of philosophy degree in biomathematics with a minor in statistics in July 2011. After graduate school the author plans to take a post-doctoral position with the Environmental Protection Agency in Research Triangle Park, North Carolina.

ACKNOWLEDGEMENTS

I am very fortunate for the education I have received from NCSU as a biomath graduate student. Many professors trusted in my abilities and challenged me. Members of the faculty and staff at this university provided me with resources to achieve my goals as a graduate student.

I especially want to express my gratitude to my advisor Dr. Tom Banks. He provided me with guidance both professionally and personally. He also provided me with many opportunities for professional development for which I am grateful. I will always remember the joy he obtained from working with students, his curiosity and enthusiasm for research, and his animated story-telling. I feel honored to be a member of his mathematical family. Dr. Banks and Susie warmly opened their home to graduate students on countless occasions. I greatly appreciate their kindness.

I would also like to thank my thesis committee, Drs. Marie Davidian, Ralph Smith, and Hien Tran, for their advice and guidance in the creation and editing of this dissertation. I am grateful to the professors and graduate students who have collaborated with me on various parts of this dissertation. In particular, I want to thank Franz Kappel, Ariel Cintron-Arias, Danielle Robbins, and Alice Matone.

Through my graduate education I have made many valuable friendships. Many of my math graduate student friends from the University of Arizona helped make my transition to NC State easier including Mia Osorio-Reich, Dan Reich, Serina Diniega, Matej Boguszak, and Suz Ward. I especially appreciate the friendships I have with fellow graduate students at NCSU. Thanks to Ben Wells for all the hiking-adventures and long bike rides. In our mathematical family, Danielle Robbins is my sister. We have experienced so much together. Chad Brown has truly supported me in my final steps of graduate school. In addition, I am grateful to the friendships and bruises provided to me by the Carolina Rollergirls. Roller derby really has been the best graduate school stress-reliever.

Finally, I would like to express my gratitude to my family. My parents have given me encouragement, support and love throughout my life and education. They have been there for me as I faced obstacles and have celebrated with me in my accomplishments. I would not be the person I am today without them. My sister Sarah is one of my closest friends, and I always appreciate the laughter we share. I am so excited about becoming an aunt to your little one! I am lucky to have the best grandparents: Grandpa Dell, Grandma Ethel, and Grandma Wilma. Though they won't be able attend my graduation, I know they will be there in spirit. Thanks to my family and friends for believing in me and giving me strength.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Mathematical models are used to describe dynamics arising from biological, physical and engineering systems. If the parameters in the model are known, the model can be used for simulation, prediction, control design, etc. However, typically one does not have accurate values for the parameters. Instead, one must estimate the parameters using experimental data. The simulation and predictive capabilities of the model depend on the accuracy of the parameter estimates. A major question that experimentalists and inverse problem investigators alike often face is how to best collect the data to enable one to efficiently and accurately estimate model parameters. This is the well-known and widely studied *optimal design* problem.

Traditional optimal design methods (D-optimal, E-optimal, c-optimal) [1, 2, 9, 20, 21] use information from the model to find the sampling distribution or mesh for the observation times (and/or locations in spatially distributed problems) that minimizes a design criterion, quite often a function of the Fisher Information Matrix (FIM). Experimental data taken on this optimal mesh is then expected to result in accurate parameter estimates.

A new optimal design, SE-optimal design (*standard error* optimal design), is introduced and compared with the more traditional D-optimal and E-optimal designs. SE-optimal design was first introduced in [6]. The goal of SE-optimal design is to find the observation times $\tau = \{t_i\}$ that minimize the sum of squared normalized standard errors of the estimated parameters as defined by asymptotic distribution results from statistical theories [5, 7, 15, 30]. D-optimal and E-optimal design methods minimize functions of the covariance in the parameter estimates [2, 9, 21]. D-optimal design finds the mesh that minimizes the volume of the confidence ellipsoid of the asymptotic covariance matrix. E-optimal design minimizes the largest principle axis of the confidence ellipsoid of the asymptotic covariance matrix.

We consider the performance of these three different optimal design methods for three different dynamical systems: the Verhulst-Pearl logistic population model, a harmonic oscillator model and a simple glucose regulation model. The three examples we use were chosen primarily

because they offer qualitatively different solutions. The Verhulst-Pearl logistic population model is a very well studied model [6, 7, 8] whose solution is a monotone increasing trajectory with a steady state. The harmonic oscillator model [8], whose solution oscillations which decay over time may provide increased challenges to the optimal design methods. The glucose regulation model [10, 11, 17, 25] is a more complex system of nonlinear differential equations with a vector observation or model output. The glucose regulation model also has a clinical application to the diagnosis of diabetes which could potentially benefit in the information obtained using optimal design methods.

In an effort to provide a reasonably fair comparison, for each optimal design method, standard errors are computed by several methods using the optimal mesh. The optimal design methods are compared based on these standard errors. The goal of this dissertation is not just to test the newer optimal design method, SE-optimal, against the more traditional optimal design methods, D-optimal and E-optimal. We also hope to motivate the use of optimal design methods in a range of practical applications where experimentalists and modelers have common goals to increase efficiency and accuracy in parameter estimates.

The organization of this dissertation is as follows. Chapter 2 carefully motivates the cost functionals for each of the optimal design methods and outlines their relationship to the confidence ellipsoid. To motivate the optimal design cost functions we must bring together ideas from linear algebra, geometry and statistics. In Chapter 3 we formulate the classical optimal design problem in the context of general optimization problems over distributions of sampling times. We present a new Prohorov metric based theoretical framework that allows one to treat succinctly and rigorously any optimal design criteria based on the Fisher Information Matrix (FIM). A fundamental approximation theory is also included in this framework. We also outline how we implement the optimal design methods to actually solve for the optimal design points in this chapter. Our formulation of the inverse problem is given in Chapter 4. Additionally, different methods for the computation of standard errors are given corresponding to variations on the mathematical model and statistical models. The mathematical models used in this optimal design comparison are carefully defined in Chapter 5. Also contained in this chapter are the results of our comparison in terms of parameter estimates and standard errors. We discuss which optimal design methods perform the best based on different criteria. Chapter 6 contains our Monte Carlo analysis ensuring that our comparison is robust to the variation in parameter estimates resulting from many different (simulated) data sets corresponding to the optimal design points. Finally, in Chapter 7, we introduce statistical models which account for autocorrelated data. We carefully explain how autocorrelation can be detected in the data, how it can be modeled, and then how to estimate autocorrelation parameters along with model parameters. The optimal design methods are compared in the context of autocorrelated data.

Chapter 2

Motivation for Optimal Design Methods

2.1 Introduction

Optimal design methods use information from the model to find the sampling distribution or mesh for the observation times (and/or locations in spatially distributed problems) that minimizes a design criterion, quite often a function of the Fisher information matrix (FIM). Experimental data taken on this optimal mesh are then expected to result in accurate parameter estimates.

In this chapter we will give the motivation behind the three optimal design methods we consider in this comparison. First we must briefly introduce the setting in which we will employ these optimal design methods. Underlying our considerations is a *mathematical model*, which may describe some physical, sociological or biological phenomenon, e.g.,

$$\begin{aligned} \dot{x}(t) &= g(t, x(t), q), \\ x(0) &= x_0, \\ f(t, \theta) &= \mathcal{C}(x(t, \theta)), \quad t \in [0, T], \end{aligned} \tag{2.1}$$

where $x(t) \in \mathbb{R}^n$ is the vector of state variables of the system, $f(t, \theta) \in \mathbb{R}^m$ is the vector of observable or measurable outputs, $q \in \mathbb{R}^r$ are the system parameters, $\theta = (q, x_0) \in \mathbb{R}^p, p = r + n$ is the vector of system parameters plus initial conditions x_0, while g and $\mathcal{C}$ are mappings $\mathbb{R}^{1+n+r} \to \mathbb{R}^n$ and $\mathbb{R}^n \to \mathbb{R}^m$, respectively. To consider measures of uncertainty in estimated parameters [5], one also requires a *statistical model*. Our statistical model is given by the stochastic process

$$Y(t) = f(t, \theta_0) + \mathcal{E}(t). \tag{2.2}$$

Here $\mathcal{E}$ is a noisy random process representing measurement errors and, as usual in statistical formulations [5, 6, 30], θ_0 is a hypothesized "true" value of the unknown parameters. Let $\sigma(t)^2 = \mathrm{Var}(\mathcal{E}(t))$. We make the following standard assumptions on the random variable $\mathcal{E}(t)$ throughout our analysis in Chapters 5 and 6:

$$\begin{aligned} \mathrm{E}(\mathcal{E}(t)) &= 0, \quad t \in [0,T], \\ \mathrm{Var}(\mathcal{E}(t)) &= \sigma_0^2, \quad t \in [0,T], \\ \mathrm{Cov}(\mathcal{E}(t)\mathcal{E}(s)) &= \sigma_0^2 \delta(t-s), \quad t,s \in [0,T], \end{aligned}$$

where $\delta(s) = 1$ for $s = 0$ and $\delta(s) = 0$ for $s \neq 0$. A realization of the observation process is given by

$$y(t) = f(t,\theta_0) + \varepsilon(t), \ t \in [0,T],$$

where the measurement error $\varepsilon(t)$ is a realization of $\mathcal{E}(t)$.

Usually our observed data is discrete, $\{t_j, y_j\}_{j=1}^N$. Optimal design methods seek the sampling distribution of times, $\tau = \{t_j\}_{j=1}^N$, that will minimize some function of the Fisher information matrix, $F = \Sigma^{-1}$, where Σ is the covariance matrix

$$\Sigma = \sigma_0^2 \left[\chi^T(\theta_0)\chi(\theta_0)\right]^{-1},$$

where χ is the $N \times p$ (number of time points: N, and number of parameters: p) sensitivity matrix such that $\chi_{i,j} = \frac{\partial f(t_i,\theta)}{\partial \theta_j}$ for $i = 1,\ldots,N$ and $j = 1,\ldots,p$, θ_0 is the vector of true parameters, and σ_0^2 is the true constant variance of the model residuals.

Each optimal design method proposes its own cost functional, $\mathcal{J}(F)$. To obtain the optimal design points, $\tau^* = \{t_i^*\}$, $i = 1,\ldots,N$, one must perform the following minimization.

$$\mathcal{J}(F(\tau^*,\theta_0,)) = \min_{\tau \in \mathcal{T}} \mathcal{J}(F(\tau,\theta_0)),$$

where $\mathcal{T}$ is the set of all time meshes such that $0 \leq t_1 \leq \cdots \leq t_i \leq t_{i+1} \leq \cdots \leq t_N \leq T$.

In general, the cost function $\mathcal{J}(F)$ is a metric that attempts to quantify our uncertainty in the parameters. Thereby minimizing the cost function often leads to small variance/covariance in the parameter estimates. Solution of the inverse problem on the optimal sampling distribution should produce parameter estimates with improved accuracy. In this chapter, we will offer motivation behind D-optimal, E-optimal and SE-optimal design. In order to do this, we first need to explore the properties of the Fisher information matrix and of linear algebra.

2.2 Properties of the Fisher information matrix

The discrete asymptotic Fisher information matrix (FIM) is defined as

$$F = \frac{1}{\sigma_0^2}\left[\chi^T(\theta_0)\chi(\theta_0)\right].$$

Note: For any $m \times n$ matrix A of (column) rank r, the product $A^T A$ is a symmetric matrix and its rank is also r [31]. This implies that if the sensitivity matrix, χ, is not full rank, then the Fisher information matrix will also not be full rank, and therefore not invertible (note: $\Sigma = F^{-1}$).

Definition: *Positive definite matrix*
A $p \times p$ real-valued matrix A, is called positive definite if $x^T Ax > 0$ for all non-zero column vectors $x \in \mathbb{R}^p$.

Definition: *Non-negative definite matrix, or Positive semi-definite*
A $p \times p$ real-valued matrix A, is called non-negative definite if $x^T Ax \geq 0$ for all non-zero column vectors $x \in \mathbb{R}^p$.

The Fisher information and covariance matrices are symmetric ($\Sigma^T = \Sigma$) and non-negative definite. The covariance matrix has variance of the parameters along its diagonal and $\mathrm{Var}(\theta) \geq 0$. Additionally, the elements of the covariance are real-valued. By definition $\mathrm{Cov}(x, y) = \mathrm{Cov}(y, x)$, making the covariance matrix symmetric.

The following theorems will help us prove properties of the confidence ellipsoid.

Theorem (3.25, [29])
If the $p \times p$ matrix A is symmetric, then
(i) A is positive definite if and only if the eigenvalues of A, $\lambda_i > 0$ for all $i = 1, \ldots, p$.
(ii) A is non-negative definite if and only if $\lambda_i \geq 0$ for all $i = 1, \ldots, p$ and $\lambda_i = 0$ for at least one i.

Theorem (3.11, [29])
If a $p \times p$ matrix is symmetric, then it is possible to construct a set of eigenvectors of A such that the set is orthonormal.

Theorem
Let A be a symmetric positive definite matrix, and x a vector. Then

$$x^T Ax = 1$$

defines an ellipsoid, where the eigenvectors of A define the principle directions of the ellipsoid

and $\lambda_i^{-\frac{1}{2}}$ define the half lengths of the principle axes, where λ_i are the corresponding eigenvalues.

Proof
Let A be a $p \times p$ matrix, with eigenvalues λ_i for $i = 1 \dots p$.
Let x be a vector in the column space of A. Let S be a $p \times p$ matrix that contains the orthonormal set of eigenvectors of A; note that S is an orthonormal matrix ($S^{-1} = S^T$). Let Λ be the corresponding $p \times p$ diagonal matrix that contains the corresponding eigenvalues of A. The spectral decomposition of A is [31]

$$A = S^{-1}\Lambda S.$$

We now have a linear transformation from the column space of A into the orthonormal space spanned by the eigenvectors of A

$$\Lambda = SAS^{-1}.$$

The ellipsoid defined by $x^T Ax = 1$, where $x \in \mathbb{R}^p$, is often tilted. To show that $x^T Ax = 1$ truly gives rise to an ellipse, we want to rotate the ellipse using a linear transformation such that we can write the ellipsoid in standard form:

$$\left(\frac{x_1}{r_1}\right)^2 + \ldots + \left(\frac{x_p}{r_p}\right)^2 = 1,$$

where $x^T = (x_1, x_2, \ldots, x_p) \in \mathbb{R}^p$, and $\{r_1, \ldots, r_p\}$ are the elliptical radii or half lengths of the principle axes. The transformation into the orthonormal space spanned by the eigenvectors of A will transform $x^T Ax = 1$ into standard form. Consider the transformation $A = S^{-1}\Lambda S$ of $x^T Ax = 1$

$$x^T Ax = x^T S^{-1}\Lambda Sx = x^T S^T \Lambda Sx = (Sx)^T \Lambda Sx = v^T \Lambda v = 1,$$

where $v := Sx$. Now in the v coordinates - the space spanned by the eigenvectors of A, we can write the ellipse in standard form, let $v^T = (v_1, \ldots, v_p)$

$$\begin{aligned} v^T \Lambda v &= 1, \\ [v_1, \ldots, v_p] \begin{pmatrix} \lambda_1 & 0 & \cdots \\ 0 & \lambda_2 & \ddots \\ \ddots & \ddots & \ddots \\ \cdots & 0 & \lambda_p \end{pmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_p \end{bmatrix} &= 1, \\ [\lambda_1 v_1, \ldots, \lambda_p v_p] \begin{bmatrix} v_1 \\ \vdots \\ v_p \end{bmatrix} &= 1, \end{aligned}$$

$$\lambda_1 v_1^2 + \ldots + \lambda_p v_p^2 = 1,$$
$$\left(\frac{v_1}{\lambda_1^{-\frac{1}{2}}}\right)^2 + \ldots + \left(\frac{v_p}{\lambda_p^{-\frac{1}{2}}}\right)^2 = 1.$$

This orthonormal eigenvector transformation results in an ellipsoid with minor axes lengths equal to $\lambda_i^{-\frac{1}{2}}$, for $i = 1 \ldots p$. The axes are aligned with v, the space spanned by the orthogonal eigenvectors of A. Therefore the original $x^T Ax = 1$ gives the equation for an ellipse with principle axis corresponding to the orthonormal eigenvectors of A with minor axes of length $\lambda_i^{-\frac{1}{2}}$, for $i = 1 \ldots p$, where λ_i, $i = 1 \ldots p$, are the eigenvalues of A. □

Example: Let A be defined as follows

$$A = \begin{pmatrix} 3 & -2 \\ -2 & 4 \end{pmatrix}.$$

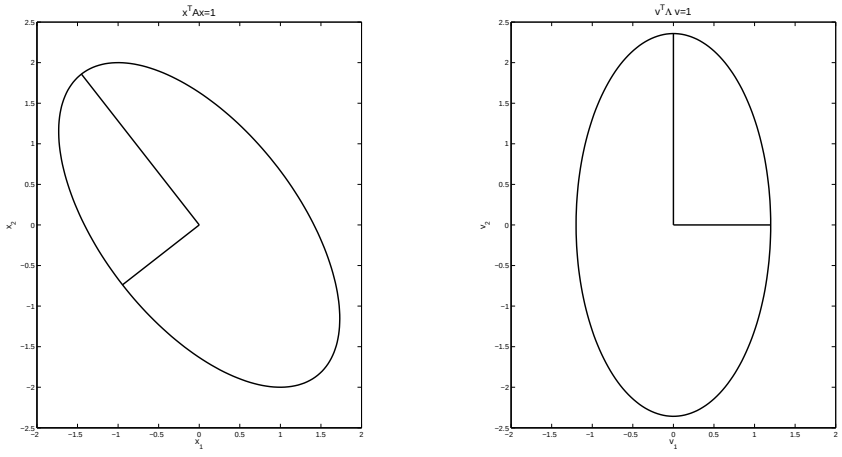
Note that A is symmetric positive definite. We are interested in the ellipse formed by $x^T Ax = 1$. To demonstrate the theorem above we will plot the ellipse formed by matrix A. We will also show how the spectral decomposition of A will form an equivalent ellipse rotated into the space spanned by the eigenvectors of A. The spectral decomposition is defined by

$$A = S^{-1} \Lambda S.$$

For this example, the spectral decomposition of A is given by

$$A = \begin{pmatrix} -.7882 & -.6154 \\ -.6154 & .7882 \end{pmatrix} \begin{pmatrix} 1.4384 & 0 \\ 0 & 5.5616 \end{pmatrix} \begin{pmatrix} -.7882 & -.6154 \\ -.6154 & .7882 \end{pmatrix}.$$

In Fig. 2.1 are the ellipses formed by A and the transformation derived above: $v^T \Lambda v = 1$. The ellipse formed by A is tilted with axes in the direction of its two eigenvectors (with slopes $\approx$ 1.28, and -0.78). The half lengths of the two axis are the same in both plots and are equal to $\lambda_i^{-\frac{1}{2}} \approx (1.4384^{-\frac{1}{2}}, 5.5616^{-\frac{1}{2}}) = (.8338, .4240)$. The plot on the right is just a rotation of the plot on the left.



(a) (b)

Figure 2.1: Ellipses corresponding to symmetric positive-definite matrices, $x^T Ax = 1$ (panel (a)), and linear transformation into orthonormal set of eigenvectors of A: $(x^T S^{-1})\Lambda(Sx) = v^T \Lambda v = 1$ (panel (b)). Both plots have the same axes bounds, tick marks and dimensions, though plotted in different variables (x: panel (a), v: panel (b)).

2.3 Relation to Confidence Ellipsoid

The confidence ellipsoid is a visualization of the confidence intervals for multiple parameter estimates all in the common parameter space. The confidence ellipsoid is defined as

$$(\theta - \hat{\theta})^T \widehat{F}(\theta - \hat{\theta}) = m,$$

where the parameter estimate is $\hat{\theta}$, the estimated Fisher information matrix is $\widehat{F} = \widehat{\Sigma}^{-1}$, and m represents the test statistic associated with the confidence level, α, for the $100\%(1-\alpha)$ confidence ellipsoid and the sample size, N. The definition of the specific test statistic requires assumptions about the distribution of the residuals. A common distributional assumption is to take $\mathcal{E} \sim \mathcal{N}(0, \sigma_0^2)$. This assumption results in the $100(1-\alpha)\%$ confidence ellipsoid being defined by [30]

$$(\theta - \hat{\theta})^T \widehat{F}(\theta - \hat{\theta}) = p\mathrm{F}^{\alpha}_{p,N-p} \iff \frac{(\theta - \hat{\theta})^T [\chi^T(\hat{\theta})\chi(\hat{\theta})](\theta - \hat{\theta})}{p\hat{\sigma}^2} = \mathrm{F}^{\alpha}_{p,N-p},$$

where $\mathrm{F}^{\alpha}_{p,N-p}$ denotes the F-distribution with numerator degrees of freedom p and denominator degrees of freedom $(N-p)$ at the α critical level. Note here $m = p\mathrm{F}^{\alpha}_{p,N-p}$.

The confidence ellipsoid will have principle axes along the eigenvectors of F in $\mathbb{R}^p$, possibly tilted when plotted on the $\theta^T = (\theta_1, \ldots, \theta_p)$ axes. The confidence ellipsoid will be centered at the parameter estimates, $\hat{\theta}$. The half axes will be proportional to (due to the constant m) $\lambda_i^{-\frac{1}{2}}$, where λ_i represents the eigenvalues of F $(i = 1 \ldots p)$. Note that $\lambda_i^{-\frac{1}{2}} = (\lambda_i^{-1})^{\frac{1}{2}} = \tilde{\lambda}_i^{\frac{1}{2}}$, where $\tilde{\lambda}_i$ are the eigenvalues of Σ, assuming F is invertible. As the width of the confidence interval is proportional to the standard error in the estimates, the length of the axes in the confidence ellipsoid are also proportional to the standard error. The $100(1-\alpha)\%$ confidence interval is defined as [5]

$$\hat{\theta} \pm t^{N-p}_{1-\alpha/2} SE(\hat{\theta}),$$

where the standard error is given by $SE(\hat{\theta}_i) = \sqrt{\Sigma_{ii}}$ for $i = 1 \ldots p$.

The theorem proved in the previous section holds for symmetric positive definite matrices. The Fisher information matrix is only symmetric non-negative definite. Supposing that the sensitivity matrix is not of full rank, the FIM will also not be full rank, and the FIM matrix may have at least one zero eigenvalue ($\lambda_i = 0$). Then the ellipsoid of $x^T F x = 1$ will only have s principle axes, where s is the number of non-zero eigenvalues ($s < p$), and the ellipsoid will exist in s-dimensional space, $\mathbb{R}^s$.

2.3.1 The Effect of Covariance on the Confidence Ellipsoid

Covariance between parameters is evident in the tilt of the confidence ellipsoid. If there is no covariance between the parameters, the covariance matrix will be diagonal, and the the confidence ellipsoid will be aligned along the standard axes. Positive covariance between parameters will negatively rotate the ellipsoid (in the clockwise direction) proportional to the degree of covariance between the parameters. Negative covariance between parameters will positively rotate the confidence ellipsoid (in the counter-clockwise direction). Whether or not the optimal design methods are taking the covariance between parameters into effect is still under investigation, and is part of our future research.

The goal of D-optimal and E-optimal design is to minimize the confidence ellipsoid in order to have smaller error in the parameter estimates. These optimal design methods take different approaches to this problem. SE-optimal design's cost functional can also be related to the confidence ellipsoid. We will motivate each of of the optimal design methods in the following sections respectively.

2.4 D-optimal

Optimal design methods find the optimal sampling distribution that minimizes some function of the FIM. The cost functional that D-optimal minimizes is

$$\mathcal{J}(F) = \det(F^{-1}).$$

We will show how minimizing this cost functional is equivalent to minimizing the volume of the confidence ellipsoid.

Let λ_i for $i = 1, ..., p$ represent the eigenvalues of the matrix A. Then the following holds [31]:

$$\det A = \prod_{i=1}^{p} \lambda_i.$$

Note that this result does not place any requirements on the matrix A.

Consider the ellipsoid with axes' half lengths equal to $r_1, \ldots, r_p$. The volume of this ellipsoid is proportional to the product of the the half lengths

$$\text{volume(ellipsoid)} = C \prod_{i=1}^{p} r_i,$$

where C depends on p (e.g. $C = \pi$ for $p = 2$, $C = 4\pi/3$ for $p = 3$, $C = \pi^2/2$ for $p = 4$, $C = 8\pi^2/15$ for $p = 5$, etc.).
Now let us consider the volume of the confidence ellipsoid

$$(\theta - \hat{\theta})^T \widehat{F} (\theta - \hat{\theta}) = m.$$

Recall that this ellipsoid has axes with half-lengths proportional to $\lambda_i^{-\frac{1}{2}}$, where λ_i are the eigenvalues of F, or equivalently $\tilde{\lambda}_i^{\frac{1}{2}}$ where $\tilde{\lambda}_i$ are the eigenvalues of $\Sigma = F^{-1}$. Therefore the volume of the ellipsoid is proportional to

$$V_{CE} = \text{volume(Confidence ellipsoid)} \propto \prod_{i=1}^{p} \lambda_i^{-\frac{1}{2}} = \left(\prod_{i=1}^{p} \lambda_i^{-1} \right)^{\frac{1}{2}} = \prod_{i=1}^{p} \tilde{\lambda}_i^{\frac{1}{2}} = \left(\prod_{i=1}^{p} \tilde{\lambda}_i \right)^{\frac{1}{2}}.$$

Since $f(x) = \sqrt{(}x)$, $x \geq 0$, is a monotone increasing function in x, minimizing $f(x)$ is equivalent to minimizing x when $x \geq 0$. Recall that we have F is non-negative definite. Therefore, the

eigenvalues of F are such that $\lambda_i \geq 0$, so the following equivalence also holds

$$\arg\min_{\tau\in\mathcal{T}}(V_{CE}) = \arg\min_{\tau\in\mathcal{T}}\left(\prod_{i=1}^{p}\lambda_i^{-1}\right)^{\frac{1}{2}} = \arg\min_{\tau\in\mathcal{T}}\left(\prod_{i=1}^{p}\lambda_i^{-1}\right) = \arg\min_{\tau\in\mathcal{T}}\det(F(\tau)^{-1}).$$

From this we can see that minimizing the cost functional for D-optimal design is equivalent to minimizing the volume of the confidence ellipsoid.

2.5 E-optimal

E-optimal design seeks the optimal sampling distribution which minimizes the following cost functional

$$\mathcal{J}(F) = \max\frac{1}{\lambda_i},$$

where λ_i are the eigenvalues of F, $i = 1\dots p$. Here we will show how minimizing this cost functional is equivalent to minimizing the largest principle axis of the confidence ellipsoid. Consider the confidence ellipsoid

$$(\theta - \hat{\theta})^T\widehat{F}(\theta - \hat{\theta}) = m.$$

The half-lengths of the principle axes of the confidence ellipsoid are proportional to $\lambda_i^{-\frac{1}{2}}$, where λ_i are the eigenvalues of F. The largest principle axis will have length proportional to $\max_i(\lambda_i^{-1})^{\frac{1}{2}}$

$$\text{Largest principle axis length} \propto \max_i \lambda_i^{-\frac{1}{2}} = (\max_i \lambda_i^{-1})^{\frac{1}{2}}.$$

Again since $f(x) = \sqrt{x}$, for $x \geq 0$, is a monotone increasing function of x, and since $\lambda_i \geq 0$, we have that $\max_i(\lambda_i^{-1})^{\frac{1}{2}}$ is equivalent to $\max_i(\lambda_i^{-1})$

$$\arg\min_{\tau\in\mathcal{T}}(\max_i \lambda_i^{-1})^{\frac{1}{2}} = \arg\min_{\tau\in\mathcal{T}}(\max_i \lambda_i^{-1}).$$

Minimizing the cost functional for E-optimal is equivalent to minimizing the largest principle axis of the confidence ellipsoid.

2.6 SE-optimal

SE-optimal design method minimizes the following design cost functional

$$\mathcal{J}(F) = \sum_{i=1}^{p} \frac{1}{\theta_{0,i}^2} (F^{-1})_{ii}.$$

Note that $F_{ii}^{-1} = SE_i(\theta_0)^2$. Therefore, SE-optimal design minimizes the sum of squared normalized standard errors. Recall that the length of the axes in the confidence ellipsoid, $\lambda_i^{-\frac{1}{2}}$, where λ_i represents the eigenvalues of F $(i = 1 \dots p)$, are also proportional to the standard error. SE-optimal is optimizing over a normalized sum of the eigenvalues of the FIM, λ_i, or the sum of squares of the normalized ellipsoidal radii.

2.7 Discussion

The confidence ellipsoid is a useful visualization of variance in parameter estimates in relation to each other. The tilt of the confidence ellipsoid is a way to visualize the covariance between parameters. The cost functionals proposed by D-optimal, E-optimal, and SE-optimal minimize this confidence ellipsoid in different ways. D-optimal minimizes the volume of the confidence ellipsoid. If several points from the sampling distribution are at the same location, the sensitivity matrix may not be full rank, and the FIM may be ill-conditioned and possibly singular. Since the cost functional of D-optimal requires inverting the FIM, we hope that ill-conditioning of the FIM is avoided by the optimization algorithm. Note that we initialize this algorithm with the uniform mesh for every optimal design method when finding the optimal design points.

E-optimal design minimizes the largest principle axis of the confidence ellipsoid. In making one axis smaller the other axes may become larger. E-optimal will try to minimize the maximum principle axis where the parameter with the largest axis may change at every step of the optimization.

SE-optimal design minimizes the sum of normalized standard errors which are proportional to the half-lengths of the confidence ellipsoid axes. By minimizing the sum of the ellipsoidal radii the error in all parameters is simultaneously being minimized.

The theory behind these optimal design methods prefer that the FIM be symmetric and positive definite, having $\lambda_i > 0, \forall i$. However statistical theory only guarantees that the FIM is symmetric and non-negative definite. The possibility that $\lambda_i = 0$ for some eigenvalue of the FIM may cause some issues for these optimal design methods. If this happens, it may mean that the number of sample points can be decreased. If the sensitivity matrix is full rank, and the FIM is invertible then FIM will be positive definite. Ideally, we would want the optimal sampling distribution to correspond to a FIM with full rank and good condition number in addition to

having an optimal confidence ellipsoid.

Now that we have motivated the cost functions for the optimal design methods we will be comparing, in the following chapter, we will more carefully define the theoretical context of our optimal design problem.

Chapter 3

Optimal Design Formulations

Following [6], we introduce a formulation of *ideal* inverse problems in which continuous in time observations are available-while not practical, the associated considerations provide valuable insight. A major question in this context is how to choose sampling distributions in an intelligent manner. Indeed, this is the fundamental question treated in the optimal design literature and methodology.

We introduce a generalized weighted least squares criterion

$$J(y,\theta) = \int_0^T \frac{1}{\sigma(t)^2}\big(y(t) - f(t,\theta)\big)^2\, dP(t), \tag{3.1}$$

where P is a general measure on $[0,T]$. We seek the parameter estimate $\hat{\theta}$ by minimizing $J(y,\theta)$ for θ. Since P represents a weighting of the difference between data and model output, we can, without loss of generality, assume that P is a bounded measure on $[0,T]$.

If, for points $\tau = \{t_i\}$, $t_1 < \cdots < t_N$ in $[0,T]$, we take

$$P_\tau = \sum_{i=1}^{N} \Delta_{t_i}, \tag{3.2}$$

where Δ_a denotes the Dirac delta distribution with atom $\{a\}$, we obtain

$$J_d(y,\theta) = \sum_{i=1}^{N} \frac{1}{\sigma(t_i)^2}\big(y(t_i) - f(t_i,\theta)\big)^2, \tag{3.3}$$

which is the weighted least squares cost functional for the case where we take a finite number of measurements in $[0,T]$. Of course, the introduction of the measure P allows us to change the weights in (3.3) or the weighting function in (3.1). For instance, if P is absolutely continuous with density $m(\cdot)$ the error functional (3.1) is just the weighted L^2-norm of $y(\cdot) - f(\cdot,\theta)$ with

weight $m(\cdot)/\sigma(\cdot)^2$.

To facilitate our discussions we introduce the *Generalized Fisher Information Matrix* (GFIM)

$$F(P,\theta_0) \equiv \int_0^T \frac{1}{\sigma^2(s)} \nabla_\theta^{\mathsf{T}} f(s,\theta_0) \nabla_\theta f(s,\theta_0)\, dP(s), \tag{3.4}$$

where ∇_θ is a row vector given by $(\partial_{\theta_1}, \ldots, \partial_{\theta_p})$ and hence $\nabla_\theta f$ is an $N \times p$ matrix. It follows that the usual discrete FIM corresponding to P_τ as in (3.2) is given by

$$F(\tau) = F(P_\tau, \theta_0) = \sum_{j=1}^{N} \frac{1}{\sigma^2(t_j)} \nabla_\theta f(t_j,\theta_0)^{\mathsf{T}} \nabla_\theta f(t_j,\theta_0). \tag{3.5}$$

Subsequently we simplify notation and use $\tau = \{t_i\}$ to represent the dependence on $P = P_\tau$ when it has the form (3.2). When one chooses P as simple Lebesgue measure then the GFIM reduces to the continuous FIM

$$F_C = \int_0^T \frac{1}{\sigma^2(s)} \nabla_\theta f(s,\theta_0)^{\mathsf{T}} \nabla_\theta f(s,\theta_0)\, ds. \tag{3.6}$$

The major question in optimal design of experiments is how to best choose P in some family $\mathcal{P}(0,T)$ of observation distributions. We observe that one optimal design formulation we might employ is a criterion that chooses the times $\tau = \{t_i\}$ for P_τ in (3.4) so that (3.5) best approximates (3.6)–i.e., one minimizes $|F_C - F(\tau)|$ over τ where $|\cdot|$ is the norm in $\mathbb{R}^{p\times p}$–see [6]. We do not consider this design here, but rather focus on the SE-optimal design also proposed in [6] and its comparison to more traditional designs.

The introduction of the measure P above allows for a unified framework for optimal design criteria which incorporates all the popular design criteria mentioned in the introduction. As already noted, the GFIM $F(P,\theta)$ introduced in (3.4) depends critically on the measure P. We also remark that we can, without loss of generality, further restrict ourselves to probability measures on $[0,T]$. Thus, let $\mathcal{P}(0,T)$ denote the set of all probability measures on $[0,T]$ and assume that a functional $\mathcal{J} : \mathbb{R}^{p\times p} \to \mathbb{R}^+$ of the GFIM is given. The *optimal design problem* associated with $\mathcal{J}$ is one of finding a probability measure $\hat{P} \in \mathcal{P}(0,T)$ such that

$$\mathcal{J}\big(F(\hat{P},\theta_0)\big) = \min_{P\in\mathcal{P}(0,T)} \mathcal{J}\big(F(P,\theta_0)\big). \tag{3.7}$$

A general theoretical framework for existence and approximation in the context of $\mathcal{P}(0,T)$ taken with the Prohorov metric [3, 12, 22, 28] is given for these problems in Section 4 of [6]. In particular, this theory permits development of computational methods using weighted discrete measures (i.e., weighted versions of (3.2)).

3.1 Theoretical Summary

To summarize and further develop the theoretical considerations that are the basis of our efforts here, we first recall that the Prohorov metric ρ on the space $\mathcal{P}(0,T)$ of probability measures on the Borel subsets of $[0,T]$ can be defined [3, 12, 22, 28] in terms of probabilities on closed subsets of $[0,T]$ and their neighborhoods. However for our uses here it is far more useful to work with an equivalent characterization in terms of convergences when viewing the probability measures $\mathcal{P}(0,T)$ as a subset of the topological dual $C_b[0,T]^*$ of the bounded continuous functions on $[0,T]$ taken with the supremum norm. More precisely, ρ-convergence is equivalent to weak*-convergence on $\mathcal{P}(0,T)$ when considering $\mathcal{P}(0,T)$ as a subset of $C_b[0,T]^*$. It is then known that $(\mathcal{P}(0,T),\rho)$ is a complete, compact and separable metric space. (We will hereafter just denote this space by $\mathcal{P}(0,T)$ since the ρ will be understood.)

Our first observation is that the GFIM as defined in (3.4) is ρ continuous on $\mathcal{P}(0,T)$ for problems in which the observation functions $f(\cdot,\theta)$ are continuously differentiable on $[0,T]$. Thus, whenever $\mathcal{J}:\mathbb{R}^{p\times p}\to\mathbb{R}^+$ is continuous we find that $P\to\mathcal{J}(F(P,\theta))$ is continuous from $\mathcal{P}(0,T)$ to $\mathbb{R}^+$. Since $\mathcal{P}(0,T)$ is ρ compact, we obtain immediately the existence of solutions for the optimization problems

$$\hat{P}_{\mathcal{J}} \equiv \text{argmin}_{P\in\mathcal{P}(0,T)}\mathcal{J}(F(P,\theta_0)). \tag{3.8}$$

Our second observation is related to the separability of $\mathcal{P}(0,T)$ and in particular to the density of finite convex combinations over rational coefficients of Dirac measures Δ_a with atoms at a. Specifically, one can prove [3] that the set

$$\mathcal{P}_0(0,T) := \left\{ P\in\mathcal{P}(0,T) \middle| P=\sum_{j=1}^{k} p_j\Delta_{t_j},\ k\in\mathbb{N}^+,\ t_j\in\mathcal{T}_0,\ p_j\geq 0,\ p_j \text{ rational},\ \sum_{j=1}^{k} p_j = 1 \right\}$$

is dense in $\mathcal{P}(0,T)$ in the Prohorov metric ρ. Here $\mathcal{T}_0=\{t_j\}_{j=1}^{\infty}$ is a countable, dense subset of $[0,T]$. In short, the set of $P\in\mathcal{P}(0,T)$ with finite support in $\mathcal{T}_0$ and rational masses is dense in $\mathcal{P}(0,T)$. This leads, for a given choice $\mathcal{J}$, to approximation schemes for $\hat{P}_{\mathcal{J}}$ as defined in (3.8). To implement these for a given choice of $\mathcal{J}$ (examples are discussed below) would require approximation by $P^N_{\{p_j,t_j\}}=\sum_{j=1}^{N} p_j\Delta_{t_j}$ in the GFIM (3.4) and then optimization over appropriate sets of $\{p_j,t_j\}$ in (3.8) with P replaced by $P^N_{\{p_j,t_j\}}$. For a fixed N, existence of minima in these problems follow from the theory outlined above. In standard optimal designs these problems are approximated even further by fixing the weights or masses p_j as $p_j=\frac{T}{N}$ (which then becomes simply a scale factor in the sum) and searching over the $\{t_j\}$. This, of course, is equivalent to replacing the $P^N_{\{p_j,t_j\}}$ by P_τ of (3.2) in (3.4) and searching over the $\tau=\{t_j\}$ for a fixed number N of grid points. This embodies the tacit assumption of equal

value of the observations at each of the times $\{t_j\}$. We observe that weighting of information at each of the observation times is carried out in the inverse problems via the weights $\sigma(t_j)$ for observation variances in (3.3). We further observe that the weights $\{p_j\}$ in $P^N_{\{p_j,t_j\}}$ are related to the value of the observations as a function of the model sensitivities $\nabla_\theta f(t_j, \theta_0)$ in the FIM while the weights $\frac{1}{\sigma(t_j)^2}$ are related to the reliability in the data measurement processes. We note that all of our remarks on theory related to existence above in the general probability measure case also hold for this discrete minimization case.

The formulation (3.8) incorporates all strategies for optimal design which entail optimization of a functional depending continuously on the elements of the Fisher information matrix. In case of the traditional design criteria, $\mathcal{J}$ is the determinant (D-optimal), or the smallest eigenvalue (E-optimal), respectively, of the inverse of the Fisher information matrix. Specifically, the optimal design methods we consider are SE-optimal design, D-optimal design, and E-optimal design. The design cost functional for the SE-optimal design method is given by (see [6])

$$\mathcal{J}(F) = \sum_{i=1}^{p} \frac{1}{\theta_{0,i}^2} (F^{-1})_{ii},$$

where $F = F(\tau)$ is the FIM, defined above in (3.5), θ_0 is the true parameter vector, and p is the number of parameters to be estimated. Note that both inversion and taking the trace of a matrix are continuous operations. We observe that $F^{-1}_{ii} = SE_i(\theta_0)^2$. Therefore, SE-optimal design minimizes the sum of squared normalized standard errors.

D-optimal design minimizes the volume of the confidence interval ellipsoid for the covariance matrix ($\Sigma_0^N = F^{-1}$). The design cost functional for D-optimal design is given by (see [9, 21])

$$\mathcal{J}(F) = \det(F^{-1}).$$

Again we note that taking the determinant is a continuous operation on matrices so that the cost functional for D-optimal is continuous in F as required by the theory.

E-optimal design minimizes the principle axis of the confidence interval ellipsoid of the covariance matrix. The design cost functional for E-optimal design is given by (see [2, 9])

$$\mathcal{J}(F) = \max \frac{1}{\lambda_i},$$

where λ_i, $i = 1 \dots p$ are the eigenvalues of F (which are continuous functions of F). Therefore $\frac{1}{\lambda_i}$, $i = 1 \dots p$, corresponds to the eigenvalues of the asymptotic covariance matrix $\Sigma_0^N = F^{-1}$.

3.2 Constrained Optimization and Implementations

Each optimal design computational method we employ is based on constrained optimization to find the mesh of time points $\tau^* = \{t_i^*\}$, $i = 1, \ldots, N$ that satisfy

$$\mathcal{J}(F(\tau^*, \theta_0,)) = \min_{\tau \in \mathcal{T}} \mathcal{J}(F(\tau, \theta_0)),$$

where $\mathcal{T}$ is the set of all time meshes such that $0 \leq t_1 \leq \cdots \leq t_i \leq t_{i+1} \leq \cdots \leq t_N \leq T$.

These optimal design methods were implemented using constrained optimization algorithms, either MATLAB's *fmincon* or SolvOpt, developed by A. Kuntsevich and F. Kappel [23], with four variations on the constraint implementation.

These constrained optimization algorithms solve the grid selection minimization problem of the form

$$\vec{\nu}^* = \min_{\vec{\nu}} \mathcal{J}(\vec{\nu}),$$

subject to the constraint(s)

$$A\vec{\nu} \leq b, \text{ and/or } A_{eq}\vec{\nu} = b_{eq},$$

where $\vec{\nu}$ is an N-vector, A is an $(N+1) \times N$ matrix, b is an $(N+1)$-vector, A_{eq} is an $N \times N$ matrix, and b_{eq} is a scalar.

For our problem, we have the constraint

$$0 \leq \nu_1 \leq \nu_2 \leq \ldots \leq \nu_N \leq 1,$$

where $\vec{t} = (t_1, ..., t_N) = \vec{\nu}T = (\nu_1 T, ..., \nu_n T)$, then

$$0 \leq t_1 \leq t_2 \leq \ldots \leq t_N \leq T.$$

To express this constraint in the form

$$A\vec{\nu} \leq b, \text{ and/or } A_{eq}\vec{\nu} = b_{eq},$$

we have several options in algebraic formulations. We denote the different *constraint implementations* (which result in different parameter and SE outcomes even in cases where the $\{t_i\}$ are initially required to satisfy similar constraints) by $(C1) - (C4)$. Our four different constraint implementations are detailed below and the differences in the implementations of the constrained optimization algorithm account for the differences in the optimal meshes generated. As is explained, a primary difference in carrying out the optimizations is the number of points over which we optimize (i.e., the number of degrees of freedom in the problem).

Constraint implementation $(C1)$*:*

For this constraint implementation, it differs from the other three in that it is not required that the end points are included in the optimal mesh. For this constraint we define the $(N+1) \times N$ matrix,

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ -1 & 1 & 0 & 0 & \cdots \\ 0 & -1 & 1 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}.$$

We define the $(N+1)$-vector

$$b = [0, \cdots, 0, 1]^{\mathrm{T}}.$$

The constraint $A\nu \leq b$, implies

$$0 \leq \nu_1 \leq \nu_2 \leq \ldots \leq \nu_N \leq 1.$$

Setting $\vec{t} = \vec{\nu} T$, we obtain

$$0 \leq t_1 \leq t_2 \leq \ldots \leq t_N \leq T.$$

In this case we optimize over N points.

Constraint implementation $(C2)$*:*

For this constraint implementation, we require that the end points are included in the optimal mesh. We optimize over the remaining mesh points $(t_2, \ldots t_{N-1})$. For this constraint we define the $(N-1) \times (N-2)$ matrix,

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ -1 & 1 & 0 & 0 & \cdots \\ 0 & -1 & 1 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}.$$

We define the $(N-1)$-vector

$$b = [0, \cdots, 0, 1]^{\mathrm{T}}.$$

The constraint $A\nu \leq b$, implies

$$0 = \nu_1 \leq \nu_2 \leq \nu_2 \leq \ldots \leq \nu_{N-1} \leq \nu_N = 1.$$

Upon setting $\vec{t} = \vec{\nu} T$, we obtain

$$0 = t_1 \leq t_2 \leq \ldots \leq t_{N-1} \leq t_N = T.$$

In this case we optimize over $N-2$ points.

Constraint implementation (C3):

For the third constraint implementation, we include the end points in the optimal mesh. We optimize over the remaining mesh points $(t_2, \ldots t_{N-1})$. For this constraint we define the $(N-1) \times (N-2)$ matrix,

$$A = \begin{pmatrix} -1 & 0 & 0 & 0 & \cdots \\ 0 & -1 & 0 & 0 & \cdots \\ 0 & 0 & -1 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & 0 & -1 \\ 1 & \cdots & 1 & 1 & 1 \end{pmatrix}.$$

We define the $(N-1)$-vector,

$$b = [0, \cdots, 0, T]^{\mathrm{T}}.$$

The constraint $A\nu \leq b$, implies

$$\nu_i \geq 0, \text{ for } i = 2, \ldots, N-1$$

and

$$\nu_2 + \nu_3 \ldots + \nu_{N-1} \leq T.$$

To form $\vec{t}$ from $\vec{\nu}$, we first must define the $(N-2) \times (N-2)$ matrix

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ 1 & 1 & 0 & 0 & \cdots \\ 1 & 1 & 1 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 1 & \cdots & 1 & 1 & 1 \end{pmatrix}.$$

Setting $t_1 = 0$, $t_N = T$ and

$$[t_2, \ldots, t_{N-1}]^{\mathrm{T}} = B[\nu_2, \ldots, \nu_{N-1}]^{\mathrm{T}},$$

which implies that

$$t_k = \sum_{j=2}^{k} \nu_j, \text{ for all } k = 2, \ldots N-1.$$

Then

$$0 = t_1 \leq \nu_2 \leq \nu_2 + \nu_3 \leq \ldots \leq (\nu_2 + \nu_3 + \ldots + \nu_{N-1}) \leq t_N = T,$$

or equivalently,

$$0 = t_1 \leq t_2 \leq \ldots \leq t_{N-1} \leq t_N = T.$$

We again optimize over $N-2$ points.

Constraint implementation (C4):

For the fourth constraint, we include the end points in the optimal mesh. For this constraint we define the $(N-1) \times (N-1)$ matrix,

$$A = \begin{pmatrix} -1 & 0 & 0 & 0 & \cdots \\ 0 & -1 & 0 & 0 & \cdots \\ 0 & 0 & -1 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & 0 & -1 \end{pmatrix}.$$

We define the $(N-1)$-vector,

$$b = [0, \cdots, 0]^{\mathrm{T}}.$$

The constraint $A\nu \leq b$, implies

$$\nu_i \geq 0, \text{ for } i = 2, \ldots, N.$$

In addition, we define the $(N-1)$-row vector

$$A_{eq} = [1, 1, \ldots, 1],$$

and the scalar $b_{eq} = T$. The additional constraint, $A_{eq}\nu = b_{eq}$, implies

$$\sum_{j=2}^{N} \nu_j = T$$

To form $\vec{t}$ from $\vec{\nu}$, we first must define the $(N-1)\times(N-1)$ matrix

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ 1 & 1 & 0 & 0 & \cdots \\ 1 & 1 & 1 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 1 & \cdots & 1 & 1 & 1 \end{pmatrix}.$$

Setting $t_1 = 0$ and

$$[t_2, \ldots, t_N]^{\mathrm{T}} = B[\nu_2, \ldots, \nu_N]^{\mathrm{T}},$$

which implies that

$$t_k = \sum_{j=2}^{k} \nu_j, \text{ for all } k = 2, \ldots N.$$

Then

$$0 = t_1 \leq \nu_2 \leq \nu_2 + \nu_3 \leq \ldots \leq (\nu_2 + \nu_3 + \ldots + \nu_N) = t_N = T,$$

or equivalently,

$$0 = t_1 \leq t_2 \leq \ldots \leq t_{N-1} \leq t_N = T.$$

In this algorithm we again effectively optimize over $N-2$ points.

Chapter 4

Inverse Problem and Standard Error Methodology

We begin by finding the optimal discrete sampling distribution of time points $\tau = \{t_i\}_{i=1}^N$, for a fixed number N of points in a fixed interval $[0, T]$, using one of three optimal design methods described above in Chapters 2 and 3. These three optimal design methods are then compared based on the standard errors computed for parameters using these sampling times. Since there are different ways to compute standard errors, we will compare the optimal design method using different techniques for computing the standard errors.

We assume statistical models with constant variance for most of this work, but for completeness we also define our standard error methodology for statistical models with non-constant variance (relative error). In addition throughout this chapter independence of the errors is assumed. In a later chapter we will discuss the necessary modifications for a statistical model with correlated errors (Chapter 7).

4.1 Standard Error Methodology for Constant Variance Data

In the following sections we will describe the methods for computing standard errors. First we consider the scalar observation case ($m = 1$).

Assume we are given experimental data $(y_1, t_1), \ldots, (y_N, t_N)$, corresponding to our optimal time points $\tau = \{t_i\}_{i=1}^N$, from the following underlying observation process

$$Y_j = f(t_j, \theta_0) + \mathcal{E}_j,$$

where $j = 1, \ldots, N$ and the $\mathcal{E}_j$ are *iid* with constant variance. Note that $E(Y_j) = f(t_j, \theta_0)$ and

$\text{Var}(Y_j) = \sigma_0^2$, with associated corresponding realizations of Y_j given by

$$y_j = f(t_j, \theta_0) + \epsilon_j.$$

Once we have an optimal distribution of time points we will obtain data or simulated data, $\{y_i\}_{i=1}^N$, a realization of the random process $\{Y_i\}_{i=1}^N$, corresponding to the optimal time points, $\tau = \{t_i\}_{i=1}^N$. Parameters are then estimated using inverse problem formulations as described in [5]. Since the variance $\text{Var}(\mathcal{E}(t)) = \sigma_0^2$ is assumed to be constant, the inverse problem is formulated using ordinary least squares (OLS). The OLS estimator is defined by

$$\Theta_{\text{OLS}} = \Theta_{\text{OLS}}^N = \arg\min_\theta \sum_{j=1}^N [Y_j - f(t_j, \theta)]^2. \tag{4.1}$$

The estimate $\hat{\theta}_{\text{OLS}}$ is defined as

$$\hat{\theta}_{\text{OLS}} = \hat{\theta}_{\text{OLS}}^N = \arg\min_\theta \sum_{j=1}^N [y_j - f(t_j, \theta)]^2.$$

The OLS cost function can be optimized to get the estimate $\hat{\theta}$ corresponding to data from the optimal time points. This nonlinear optimization is implemented in MATLAB using *fminsearch*, using simulated data, with initial parameter guess $\theta^0 = 1.4\theta_0$, where θ_0 represents the vector of true parameter values. Specifics about our simulated data is given for each example.

4.1.1 Asymptotic Theory for Computing Standard Errors

To compute the standard errors of the estimated parameters, we first must compute the sensitivity matrix

$$\chi_{j,k} = \frac{\partial(\mathcal{C}x(t_j))}{\partial\theta_k} = \frac{\partial f(t_j, \theta)}{\partial\theta_k}, \text{ for } j = 1, \ldots N, k = 1, \ldots p.$$

Note that $\chi = \chi^N$ is an $N \times p$ matrix. The true constant variance

$$\sigma_0^2 = \frac{1}{N}\text{E}\left[\sum_{j=1}^N [Y_j - f(t_j, \theta_0)]^2\right],$$

can be estimated by

$$\hat{\sigma}_{\text{OLS}}^2 = \frac{1}{N-p}\sum_{j=1}^N [y_j - f(t_j, \hat{\theta}_{\text{OLS}})]^2.$$

The true covariance matrix is approximately (asymptotically as $N \to \infty$) given by,

$$\Sigma_0^N = \sigma_0^2[\chi^T(\theta_0)\chi(\theta_0)]^{-1}.$$

Note that the approximate Fisher Information Matrix (FIM) is defined by

$$F = F(\tau) = F(\tau, \theta_0) = (\Sigma_0^N)^{-1}, \tag{4.2}$$

and is explicitly dependent on the sampling times τ.

When the true values, θ_0 and σ_0^2, are unknown, the covariance matrix is estimated by

$$\Sigma_0^N \approx \hat{\Sigma}^N(\hat{\theta}_{\text{OLS}}) = \hat{\sigma}^2_{\text{OLS}}[\chi^T(\hat{\theta}_{\text{OLS}})\chi(\hat{\theta}_{\text{OLS}})]^{-1}. \tag{4.3}$$

The corresponding FIM can be estimated by

$$\hat{F}(\tau) = \hat{F}(\tau, \hat{\theta}_{\text{OLS}}) = (\hat{\Sigma}^N(\hat{\theta}_{\text{OLS}}))^{-1}. \tag{4.4}$$

The asymptotic standard errors are given by

$$SE_k(\theta_0) = \sqrt{(\Sigma_0^N)_{kk}},\ k = 1, \ldots, p. \tag{4.5}$$

These standard errors are estimated in practice (when θ_0 and σ_0 are not known) by

$$SE_k(\hat{\theta}_{\text{OLS}}) = \sqrt{(\hat{\Sigma}^N(\hat{\theta}_{\text{OLS}}))_{kk}},\ k = 1, \ldots, p. \tag{4.6}$$

It can be shown, under certain conditions, for $N \to \infty$, that the estimator Θ^N_{OLS} is asymptotically normal [30]; i.e., for N large

$$\Theta^N_{\text{OLS}} \sim \mathcal{N}_p(\theta_0, \Sigma_0^N). \tag{4.7}$$

4.1.2 Monte Carlo Method for Asymptotic Standard Errors

To account for the variability in the asymptotic standard errors due to the variability in the residual errors in the simulated data, we use Monte Carlo trials to examine the average behavior. For a single Monte Carlo trial, we generate simulated data on the optimal mesh $\{t_j\}_{j=1}^N$,

$$y_j = f(t_j, \theta_0) + \epsilon_j,\ j = 1, \ldots N,$$

where ϵ_j are realizations of $\mathcal{E}_j \sim \mathcal{N}(0, \sigma^2)$ for $j = 1, \ldots, N$. Parameters are estimated using OLS with initial parameter guess $\theta^0 = 1.4\theta_0$, where θ_0 are the true parameters. Standard errors are estimated using asymptotic theory (4.6). The parameter estimates and their estimated standard

errors are stored, and the process is repeated with new simulated data corresponding to the optimal mesh for M Monte Carlo trials (e.g., $M = 1000$). The average of the M parameter estimates and standard errors are used to compare the optimal design methods in one of our examples (Section 5.1).

4.1.3 The Bootstrapping Method

An alternative way of computing parameter estimates and standard errors uses the bootstrapping method [7]. Again we outline this for the case of scalar ($m = 1$) observations.

As in the previous section, assume we are given experimental data $(y_1, t_1), \ldots, (y_N, t_N)$ from the following underlying observation process

$$Y_j = f(t_j, \theta_0) + \mathcal{E}_j, \tag{4.8}$$

where $j = 1, \ldots, N$ and the $\mathcal{E}_j$ are independent identically distributed *(iid)* from a distribution $\mathcal{F}$ with mean zero ($\mathrm{E}(\mathcal{E}_j) = 0$) and constant variance σ_0^2, and θ_0 is the "true" parameter value. Associated corresponding realizations of Y_j are given by

$$y_j = f(t_j, \theta_0) + \epsilon_j.$$

The bootstrapping algorithm is presented for sample points corresponding to the t_j, $j = 1 \ldots N$. To compare the optimal design methods based on their bootstrapping standard errors, we will take our sample points corresponding to the optimal time distribution ($\tau = \{t_i\}_{i=1}^N$).

The following algorithm [7] can be used to compute the bootstrapping estimate $\hat{\theta}_{boot}$ of θ_0 and its empirical distribution.

1. First estimate $\hat{\theta}^0$ from the entire sample, using OLS.

2. Using this estimate define the standardized residuals:

$$\bar{r}_j = \sqrt{\frac{N}{(N-p)}} \left(y_j - f(t_j, \hat{\theta}^0) \right)$$

   for $j = 1, \ldots, N$. Then $\{\bar{r}_1, \ldots, \bar{r}_N\}$ are realizations of *iid* random variables $\bar{R}_j$ from the empirical distribution $\mathcal{F}_N$, and p for the number of parameters. Observe that

$$E(\bar{r}_j | \mathcal{F}_N) = N^{-1} \sum_{j=1}^{N} \bar{r}_j = 0, \quad \mathrm{Var}(\bar{r}_j | \mathcal{F}_N) = N^{-1} \sum_{j=1}^{N} \bar{r}_j^2 = \hat{\sigma}^2.$$

   Set $m = 0$. Note that in this algorithm, m represents the m^{th} bootstrap sample, and

unrelated to the number of states (scalar or vector).

3. Create a bootstrap sample of size N using random sampling with replacement from the data (realizations) $\{\bar{r}_1,\ldots,\bar{r}_N\}$ to form a bootstrap sample $\{r_1^m,\ldots,r_N^m\}$.

4. Create bootstrap sample points

$$y_j^m = f(t_j,\hat{\theta}^0) + r_j^m,$$

where $j = 1,\ldots,N$.

5. Obtain a new estimate $\hat{\theta}^{m+1}$ from the bootstrap sample $\{y_j^m\}$ using OLS. Add $\hat{\theta}^{m+1}$ into the vector Θ, where Θ is a vector of length Mp (M is the number of bootstrap samples) which stores the bootstrap estimates.

6. Set $m = m + 1$ and repeat steps 3–5.

7. Carry out the above iterative process M times where M is large (e.g., M=1000), resulting in a vector Θ of length Mp.

8. We then calculate the mean and standard error from the vector Θ using the formulae

$$\begin{aligned} \hat{\theta}_{boot} &= \tfrac{1}{M}\textstyle\sum_{m=1}^{M}\hat{\theta}^m, \\ \mathrm{Cov}(\hat{\theta}_{boot}) &= \tfrac{1}{M-1}\textstyle\sum_{m=1}^{M}(\hat{\theta}^m - \hat{\theta}_{boot})^T(\hat{\theta}^m - \hat{\theta}_{boot}), \\ SE_k(\hat{\theta}_{boot}) &= \sqrt{\mathrm{Cov}(\hat{\theta}_{boot})_{kk}}. \end{aligned} \tag{4.9}$$

4.2 Standard Error Methodology for a Vector System

Though the vector methodology is similar to that in the scalar case, for completeness we outline it here for a system of differential equations such as the simple glucose regulation model (see Section 5.3).

We begin by finding the optimal discrete sampling distribution of time points $\tau = \{t_i\}_{i=1}^N$, for a fixed number of points, N, and a fixed final time, T, using either SE-optimal, D-optimal, or E-optimal. These three optimal design methods are then compared based on their parameter estimates and standard errors using these sampling times. The standard errors can be computed using asymptotic theory or the bootstrapping method.

More specifically, once we have an optimal distribution of time points we will obtain data or simulated data, $\{\vec{y}_i\}_{i=1}^N$, a realization of the random process $\{\vec{Y}_i\}_{i=1}^N = \{(G_i, I_i)^T\}_{i=1}^N$ given by

$$\vec{Y}_i = \vec{f}(t_i,\theta_0) + \vec{\mathcal{E}}_i,$$

corresponding to the optimal time points, $\tau = \{t_i\}_{i=1}^N$, where $\vec{\mathcal{E}}_i = \vec{\mathcal{E}}(t_i)$. Note that here $G(t)$ and $I(t)$ represent the concentration of glucose and insulin, respectively, in the blood at time t, and represent the two compartments in our vector system ($m = 2$). (G_i, I_i) represents the ith observation corresponding to time t_i.
Define $V_0 = \mathrm{Var}(\vec{\mathcal{E}}_i) = \mathrm{diag}(\sigma_{0,G}^2, \sigma_{0,I}^2)$. Though now we have a vector system, we still assume constant variance.

4.2.1 Asymptotic Theory for Computing Standard Errors for a Vector System

When the variance is assumed to be constant, the inverse problem is formulated using ordinary least squares (OLS). The OLS *estimator* for a vector system is defined by

$$\Theta_{\mathrm{OLS}} = \Theta_{\mathrm{OLS}}^N = \arg\min_\theta \sum_{j=1}^N [\vec{Y}_j - \vec{f}(t_j, \theta)]^T V_0^{-1} [\vec{Y}_j - \vec{f}(t_j, \theta)].$$

For a given realization $\{y_j\}$, the OLS *estimate* $\hat{\theta}_{\mathrm{OLS}}$ is defined as

$$\hat{\theta}_{\mathrm{OLS}} = \hat{\theta}_{\mathrm{OLS}}^N = \arg\min_\theta \sum_{j=1}^N [\vec{y}_j - \vec{f}(t_j, \theta)] V_0^{-1} [\vec{y}_j - \vec{f}(t_j, \theta)].$$

The definition of variance gives

$$V_0 = \mathrm{diag\ E}\left(\frac{1}{N} \sum_{j=1}^N [\vec{Y}_j - \vec{f}(t_j, \theta_0)][\vec{Y}_j - \vec{f}(t_j, \theta_0)]^T \right).$$

In the case that V_0 is unknown, an unbiased estimate can be obtained from the realizations $\{\vec{y}_i\}_{i=1}^N$ and $\hat{\theta}$ by

$$V_0 \approx \hat{V} = \mathrm{diag}\left(\frac{1}{N-p} \sum_{j=1}^N [\vec{y}_j - \vec{f}(t_j, \hat{\theta})][\vec{y}_j - \vec{f}(t_j, \hat{\theta})]^T \right),$$

which is solved simultaneously (in an iterative procedure - see [5]) with normal equations for the estimate $\hat{\theta} = \hat{\theta}_{\mathrm{OLS}}$, where p is the number of parameters being estimated.

To compute the standard errors of the estimated parameters, we first must compute the $2 \times p$ sensitivity matrices $D_j(\theta) = D_j^N(\theta)$ which are given by

$$D_j = \begin{pmatrix} \frac{\partial f_1(t_j,\theta)}{\partial \theta_1} & \frac{\partial f_1(t_j,\theta)}{\partial \theta_2} & \cdots & \frac{\partial f_1(t_j,\theta)}{\partial \theta_p} \\ \frac{\partial f_2(t_j,\theta)}{\partial \theta_1} & \frac{\partial f_2(t_j,\theta)}{\partial \theta_2} & \cdots & \frac{\partial f_2(t_j,\theta)}{\partial \theta_p} \end{pmatrix},$$

for $j = 1, \ldots, N$. For this system we can rewrite D_j in terms of $(G(t_j, \theta), I(t_j, \theta))^T$ (since $(f_1(t_j, \theta), f_2(t_j, \theta))^T = (G(t_j, \theta), I(t_j, \theta))^T$). We have

$$D_j = \begin{pmatrix} \frac{\partial G(t_j,\theta)}{\partial \theta_1} & \frac{\partial G(t_j,\theta)}{\partial \theta_2} & \cdots & \frac{\partial G(t_j,\theta)}{\partial \theta_p} \\ \frac{\partial I(t_j,\theta)}{\partial \theta_1} & \frac{\partial I(t_j,\theta)}{\partial \theta_2} & \cdots & \frac{\partial I(t_j,\theta)}{\partial \theta_p} \end{pmatrix}.$$

The true covariance matrix is approximately (asymptotically as $N \to \infty$) given by

$$\Sigma_0^N \approx \left(\sum_{j=1}^{N} D_j^T(\theta_0) V_0^{-1} D_j(\theta_0) \right)^{-1}.$$

When the true values, θ_0 and V_0, are unknown, the covariance matrix is estimated by

$$\Sigma_0^N \approx \hat{\Sigma}^N = \left(\sum_{j=1}^{N} D_j^T(\hat{\theta}_{\text{OLS}}) \hat{V}^{-1} D_j(\hat{\theta}_{\text{OLS}}) \right)^{-1}.$$

The corresponding FIM, asymptotic standard errors and asymptotic distribution are again given by (4.4), (4.5), (4.6), and (4.7), respectively.

4.2.2 The Bootstrap Method for a Vector System

The bootstrap method for a system of differential equations is the same as described in the previous section, except that each state variable has its own residuals that must be separately sampled with replacement. The first four steps of the bootstrap algorithm of Section 4.1.3 modified for a system with vector observations is outlined here for completeness.

1. First estimate $\hat{\theta}^0$ from the entire sample, using OLS.

2. Using this estimate define the standardized residuals:

$$\bar{r}_{G,j} = \sqrt{\frac{N}{(N-p)}} \left(y_{1,j} - f_1(t_j, \hat{\theta}^0) \right),$$

$$\bar{r}_{I,j} = \sqrt{\frac{N}{(N-p)}} \left(y_{2,j} - f_2(t_j, \hat{\theta}^0) \right)$$

   for $j = 1, \ldots, N$. Then $\{\bar{r}_{G,1}, \ldots, \bar{r}_{G,N}\}$,$\{\bar{r}_{I,1}, \ldots, \bar{r}_{I,N}\}$ are realizations of *iid* random variables from the empirical distribution $\mathcal{F}_N$, and p for the number of parameters.

   Set $m = 0$.

3. Create a two different bootstrap samples of size N using random sampling with replacement from the data (realizations) $\{\bar{r}_{G,1},\ldots,\bar{r}_{G,N}\}$ and $\{\bar{r}_{I,1},\ldots,\bar{r}_{I,N}\}$ to form the bootstrap samples $\{r^m_{G,1},\ldots,r^m_{G,N}\}$ and $\{r^m_{I,1},\ldots,r^m_{I,N}\}$.

4. Create bootstrap sample points

$$y^m_{1,j} = f_1(t_j,\hat{\theta}^0) + r^m_{G,j},$$

$$y^m_{2,j} = f_2(t_j,\hat{\theta}^0) + r^m_{I,j},$$

where $j = 1,\ldots,N$.

5. Steps 5-8 are the same as those of the algorithm for scalar observations in Section 4.1.3.

4.3 Standard Error Methodology for Non-constant Variance Data

We present this standard error methodology for the scalar case ($m = 1$), but this could be generalized for a vector case as was shown for the constant variance case (section 4.2).

We suppose now that we are given experimental data $(y_1,t_1),\ldots,(y_N,t_N)$ from the following underlying observation process

$$Y_j = f(t_j,\theta_0)(1+\mathcal{E}_j) \tag{4.10}$$

where $j = 1,\ldots,N$ and the $\mathcal{E}_j$ are *iid* with mean zero and non-constant variance. Note that $E(Y_j) = f(t_j,\theta_0)$ and $\mathrm{Var}(Y_j) = \sigma_0^2 f^2(t_j,\theta_0)$, with associated corresponding realizations of Y_j given by

$$y_j = f(t_j,\theta_0)(1+\epsilon_j).$$

We see that the variance generated in this fashion is model dependent and hence generally is longitudinally non-constant variance. The appropriate method to use to estimate θ_0 and σ_0^2 is a particular form of the Generalized Least Squares (GLS) method [5, 15].

To define the *random variable* θ_{GLS} the following equation must be solved for the estimator θ_{GLS}:

$$\sum_{j=1}^{N} w_j[Y_j - f(t_j,\theta_{\mathrm{GLS}})]\nabla f(t_j,\theta_{\mathrm{GLS}}) = 0, \tag{4.11}$$

where Y_j obeys (4.10) and $w_j = f^{-2}(t_j,\theta_{\mathrm{GLS}})$. We note these are the normal equations (obtained by equating to zero the gradient of the weighted least squares criterion in the case where the weights w_j are not dependent on θ). The quantity θ_{GLS} is a random variable, hence if $\{y_j\}_{j=1}^N$ is

a *realization* of the random process Y_j then solving

$$\sum_{j=1}^{N} f^{-2}(t_j, \widehat{\theta})[y_j - f(t_j, \widehat{\theta})]\nabla f(t_j, \widehat{\theta}) = 0, \tag{4.12}$$

for $\widehat{\theta}$ we obtain an estimate $\widehat{\theta}_{\text{GLS}}$ for θ_{GLS}.

The GLS Algorithm

An estimate $\widehat{\theta}_{\text{GLS}}$ can be solved for iteratively. The iterative procedure as described in [15] is summarized as follows:

1. Estimate $\widehat{\theta}_{\text{GLS}}$ by $\hat{\theta}^{(0)}$ using the OLS equation (4.1). Set $k = 0$.
2. Form the weights $\widehat{w}_j = f^{-2}(t_j, \hat{\theta}^{(k)})$.
3. Re-estimate $\widehat{\theta}$ by solving

$$\hat{\theta}^{(k+1)} = \arg\min_{\theta \in \Theta_{ad}} \sum_{j=1}^{N} \widehat{w}_j \left(y_j - f\big(t_j, \theta\big)\right)^2$$

   to obtain the $k+1$ estimate $\hat{\theta}^{(k+1)}$ for $\widehat{\theta}_{\text{GLS}}$.
4. Set $k = k+1$ and return to 2. Terminate the process when two of the successive estimates for $\widehat{\theta}_{\text{GLS}}$ are sufficiently close.

4.3.1 Asymptotic Theory for Computing Standard Error for Non-Constant Variance Data

Assume we are given experimental data $(y_1, t_1), \ldots, (y_N, t_N)$ from the following underlying observation process

$$Y_j = f(t_j, \theta_0)(1 + \mathcal{E}_j),$$

where $j = 1, \ldots, N$ and the $\mathcal{E}_j$ are *iid* with non-constant variance. Note that $E(Y_j) = f(t_j, \theta_0)$ and $\text{Var}(Y_j) = \sigma_0^2 f^2(t_j, \theta_0)$, with associated corresponding realizations of Y_j given by

$$y_j = f(t_j, \theta_0)(1 + \epsilon_j).$$

When using asymptotic theory [5, 8], we obtain the estimate $\hat{\theta}$ using the Generalized Least Squares (GLS) algorithm. Then σ_0^2 is approximated by

$$\sigma_0^2 \approx \hat{\sigma}_{GLS}^2 = \frac{1}{N-p} \sum_{j=1}^{N} \frac{1}{f^2(t_j, \hat{\theta})} (f(t_j; \hat{\theta}) - y_j)^2.$$

We estimate the covariance matrix using $\hat{\theta}$ and $\hat{\sigma}^2_{GLS}$ by

$$\Sigma_0^N \approx \hat{\Sigma}^N(\hat{\theta}) = \hat{\sigma}^2_{GLS}[\chi^T(\hat{\theta})W(\hat{\theta})\chi(\hat{\theta})]^{-1},$$

where $W^{-1}(\theta) = diag(f^2(t_1,\theta),\ldots,f^2(t_N,\theta))$. We compute the standard error using $\hat{\Sigma}^N(\hat{\theta})$ and

$$SE_k(\hat{\theta}) = \sqrt{\hat{\Sigma}^N_{kk}(\hat{\theta})}.$$

4.3.2 Bootstrapping Algorithm for Computing Standard Error for Non-constant Variance Data

The following algorithm [13, 14, 16, p. 287–290] can be used to compute the *bootstrapping estimate* $\hat{\theta}_{boot}$ of θ_0 and its empirical distribution.

1. First obtain the estimate $\hat{\theta}^0$ from the entire sample, using GLS.

2. ***For the case where*** $f(t_j,\theta_0)$ ***is a linear function of the parameters*** θ_0***:***
   (a) Using this estimate define the standardized residuals:

$$\bar{r}_j = \sqrt{\frac{N}{(N-p)}} \quad \frac{\left(y_j - f(t_j,\hat{\theta}^0)\right)}{f(t_j,\hat{\theta}^0)}$$

   for $j = 1,\ldots,N$. Then $\{\bar{r}_1,\ldots,\bar{r}_N\}$ are realizations of *iid* random variables $\bar{R}_j$, and p for the number of parameters.
   (b) Define $\bar{r}_{avg} = N^{-1}\sum_{j=1}^N \bar{r}_j$.
   (c) Define $\hat{\sigma}^2_{boot} = N^{-1}\sum_{j=1}^N \bar{r}_j^2$.
   (d) Define the non-constant variance standardized residuals:

$$\bar{s}_j = \left(1 - \frac{\bar{r}^2_{avg}}{\hat{\sigma}^2_{boot}}\right)^{-1/2} (\bar{r}_j - \bar{r}_{avg}).$$

   Then $\{\bar{s}_1 \ldots \bar{s}_N\}$ are *iid* from the empirical distribution $\mathcal{F}_N$. We modify the standardized residuals from what we had in the constant variance case, so that the following conditions would continue to hold:

$$E(\bar{s}_j|\mathcal{F}_N) = 0, \quad \mathrm{Var}(\bar{s}_j|\mathcal{F}_N) = \hat{\sigma}^2.$$

   Set $m = 0$.
   For the case where $f(t_j,\theta_0)$ ***is a non-linear function of the parameters*** θ_0***:***

Define the non-constant variance standardized residuals

$$\bar{s}_j = \left(y_j - f(t_j, \hat{\theta}^0)\right) / f(t_j, \hat{\theta}^0).$$

Then $\{\bar{s}_1 \ldots \bar{s}_N\}$ are *iid* from the empirical distribution $\mathcal{F}_N$. Again the standardized residuals have been modified from the linear model case so that the following conditions only approximately hold:

$$E(\bar{s}_j|\mathcal{F}_N) \approx 0, \quad \text{Var}(\bar{s}_j|\mathcal{F}_N) \approx \hat{\sigma}^2.$$

Set $m = 0$.

3. Create a bootstrap sample of size N using random sampling with replacement from the data (realizations) $\{\bar{s}_1,\ldots,\bar{s}_N\}$ to form a bootstrap sample $\{s_1^m, \ldots, s_N^m\}$.

4. Create bootstrap sample points

$$y_j^m = f(t_j, \hat{\theta}^0) + f(t_j, \hat{\theta}^0) s_j^m,$$

where $j = 1,\ldots,N$.

5. Obtain a new estimate $\hat{\theta}^{m+1}$ from the bootstrap sample $\{y_j^m\}$ using GLS. Add $\hat{\theta}^{m+1}$ into the vector Θ, where Θ is a vector of length M which stores the bootstrap estimates.

6. Set $m = m + 1$ and repeat steps 3–5.

7. Carry out the above iterative process M times where M is large (e.g., $M = 1000$), resulting in a vector Θ of length M.

8. We then calculate the mean, standard error, and confidence intervals from the vector Θ using the same formulae (4.9) as before.

If bootstrapping samples $\{y_1^m, ..., y_N^m\}$ resemble the data $\{y_1, ..., y_N\}$ in terms of the empirical error distribution, $\mathcal{F}_N$, then the parameter estimates are expected to be consistent. The modification of the standardized residuals allows each of the bootstrapping samples to have an empirical distribution with the same mean and variance as the original $\mathcal{F}_N$.

4.4 Choosing the Correct Statistical Model: Constant vs Non-constant Variance

In practice it is important to choose the correct statistical model. Residual tests can be useful in determining where a constant variance or non-constant variance assumption should be made. To

test the assumption of constant variance, one can plot the residuals vs time. If the assumption is correct, one should observe no pattern in the residuals across time with relatively constant variance. If the pattern of the residuals over time is non-constant one can propose a model in the form

$$\text{Var}(Y_j) = \sigma^2 g^2(t_j, \theta).$$

If the residuals exhibit relative error: i.e., the amount of variance in the residuals scales with the model's function value, then one can choose $g(t, \theta) = f(t, \theta)$. The choice of $g(t, \theta)$ can be tested using the weighted residual test. Define the weighted residuals as follows

$$\epsilon_{j,GLS} = \frac{y_j - f(t_j, \hat{\theta})}{\hat{\sigma} g(t_j, \hat{\theta})}.$$

If your statistical model is correct, plotting the weighted residuals vs time should exhibit a random pattern. We will discuss testing the assumption of independent errors in Chapter 7.

4.5 Discussion of Asymptotic Theory vs Bootstrapping

A simulation study comparing standard error computations from asymptotic theory and the bootstrapping method is given in [7]. Here we summarize the conclusions from this comparison. In general, the time to compute standard errors from asymptotic theory is shorter compared to that for the bootstrapping method. However, asymptotic theory requires sensitivity equations to be solved analytically or numerically, where as the bootstrapping method does not. Examining the standard errors themselves, they found no advantage in asymptotic theory or bootstrapping for constant variance data. With non-constant data there can be an advantage to using the bootstrapping method. If local variation in the data is present in regions of importance for the estimation of the parameters, then the bootstrapping standard error estimate for those parameters will be larger due to a corrective term. When this corrective term is present, the bootstrapping standard error estimate will be larger yet more accurate and more conservative than the smaller asymptotic theory standard error estimate. If local variance is not sufficiently large in regions of importance for the estimation of a parameter, then there will be no advantageous method for computing the standard error for that parameter.

Chapter 5

Optimal Design Comparison for Three Examples

We will compare the optimal design methods using the standard errors resulting from the optimal time points each method proposes. Since there are different ways to compute the standard errors we will present results for several of these computational methods.

We compare the optimal design methods for mathematical models of increasing complexity. First is the Verhulst-Pearl logistic population model: a first order differential equation whose solution is a monotone increasing function with a steady state, a very well studied model. Second is the harmonic oscillator model: a second order differential equation whose solution exhibits oscillatory behavior. Lastly is a glucose regulation model: a system of nonlinear differential equations.

5.1 The Logistic Growth Example

We first compare the optimal design methods for the logistic example using the Monte Carlo method for asymptotic estimates and standard errors.

5.1.1 Logistic Model

The Verhulst-Pearl logistic population model describes a population that grows at an intrinsic growth rate until it reaches its carrying capacity. It is given by the differential equation:

$$\dot{x}(t) = rx(t)\left(1 - \frac{x(t)}{K}\right),\ x(0) = x_0,$$

where K is the carrying capacity of the population, r is the intrinsic growth rate, and x_0 is the initial population size. The analytical solution to the differential equation above is given by,

$$x(t) = f(t, \theta_0) = \frac{K}{1 + (K/x_0 - 1)e^{-rt}},$$

where $\theta_0 = (K, r, x_0)$ is the true parameter vector.
Our statistical model is given by

$$Y(t) = f(t, \theta_0) + \mathcal{E}(t),$$

where we choose $\mathcal{E} \sim \mathcal{N}(0, \sigma_0^2)$ to generate simulated data (for use in the Monte Carlo calculations). A realization of the observation process is given by

$$y(t) = f(t, \theta_0) + \varepsilon(t),\ t \in [0, T].$$

5.1.2 Logistic Results

For the logistic model, we use SolvOpt to solve for the optimal mesh for each of the optimal design methods (D-optimal, E-optimal and SE-optimal), using the second constraint ($C2$) on the time points: $t_1 \geq 0$, $t_N \leq T$ and $t_i \leq t_{i+1}$, such that the optimal mesh contains 0 and T. For this example, we took $T = 25$ and $N = 10$ or $N = 15$. Figures 5.1 and 5.3 contain plots of the resulting optimal distribution of time points for the different optimal design methods, along with the uniform mesh, plotted on the logistic curve, for $N = 10$ and $N = 15$, respectively.

These optimal design methods are compared based on their average Monte Carlo asymptotic estimates and standard errors (described in Section 4.1.2). The simulated data was generated assuming the true parameter values $\theta_0 = (K, r, x_0) = (17.5, 0.7, 0.1)$, and constant variance $\sigma_0^2 = 0.16$. The average estimates and standard errors are based on $M = 1000$ Monte Carlo trials. Since we obtain histograms of estimates and standard errors from this Monte Carlo analysis, we can also gain information for comparison from the median of these histograms or sampling distributions. Monte Carlo asymptotic estimates and standard errors were also computed on the uniform mesh. We report the average and median estimates and standard errors in Tables 5.1 and 5.2 ($N = 10$, and $N = 15$). Histograms for the Monte Carlo standard errors are given in Figs. 5.2 and 5.4 for $N = 10$ and $N = 15$, respectively.

5.1.3 Discussion of Logistic Results

The average asymptotic estimates from the uniform distribution and each of the optimal design methods are very close to the true values, θ_0. For $N = 10$ (Table 5.1), SE-optimal has the closest average and median estimates, followed by D-optimal (for r and x_0) and E-optimal

Table 5.1: Average and Median estimates and standard errors using SolvOpt, $N = 10$, $M = 1000$, and $\theta_0 = (17.5, 0.7, 0.1)$. Optimization with constraint implementation $(C2)$.

Parameter	Method	Average Est	Median Est	Average SE	Median SE
K	Unif	17.4978	17.4954	1.789×10^{-1}	1.789×10^{-1}
	SE-opt	17.4985	17.4995	2.000×10^{-1}	2.000×10^{-1}
	D-opt	17.5066	17.5024	2.039×10^{-1}	2.038×10^{-1}
	E-opt	17.4959	17.4957	1.512×10^{-1}	1.512×10^{-1}
r	Unif	0.7042	0.6996	5.020×10^{-2}	4.983×10^{-2}
	SE-opt	0.7019	0.7000	3.473×10^{-2}	3.444×10^{-2}
	D-opt	0.7020	0.7029	3.821×10^{-2}	3.816×10^{-2}
	E-opt	0.7139	0.7033	9.696×10^{-2}	9.090×10^{-2}
x_0	Unif	0.1037	0.0999	3.730×10^{-2}	3.696×10^{-2}
	SE-opt	0.1018	0.1002	2.448×10^{-2}	2.432×10^{-2}
	D-opt	0.1025	0.0982	2.947×10^{-2}	2.859×10^{-2}
	E-opt	0.1103	0.0977	6.417×10^{-2}	6.174×10^{-2}

Table 5.2: Average and Median estimates and standard errors using SolvOpt, $N = 15$, $M = 1000$, and $\theta_0 = (17.5, 0.7, 0.1)$. Optimization with constraint implementation $(C2)$.

Parameter	Method	Average Est	Median Est	Average SE	Median SE
K	Unif	17.5004	17.5009	1.467×10^{-1}	1.466×10^{-1}
	SE-opt	17.4941	17.4899	1.633×10^{-1}	1.633×10^{-1}
	D-opt	17.5017	17.4974	1.553×10^{-1}	1.552×10^{-1}
	E-opt	17.5006	17.5006	1.265×10^{-1}	1.265×10^{-1}
r	Unif	0.7018	0.6983	4.118×10^{-2}	4.086×10^{-2}
	SE-opt	0.7008	0.6993	2.739×10^{-2}	2.721×10^{-2}
	D-opt	0.7022	0.7016	3.353×10^{-2}	3.353×10^{-2}
	E-opt	0.7056	0.7004	8.078×10^{-2}	7.799×10^{-2}
x_0	Unif	0.1027	0.1004	3.040×10^{-2}	3.020×10^{-2}
	SE-opt	0.1016	0.0997	1.999×10^{-2}	1.977×10^{-2}
	D-opt	0.1014	0.0992	2.476×10^{-2}	2.440×10^{-2}
	E-opt	0.1078	0.0989	4.920×10^{-2}	4.788×10^{-2}

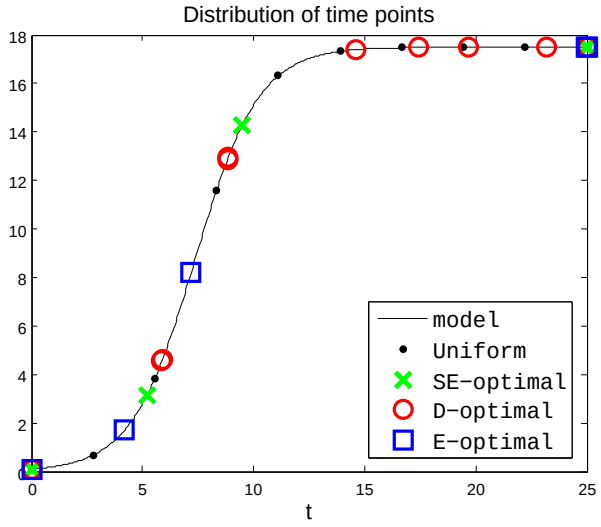
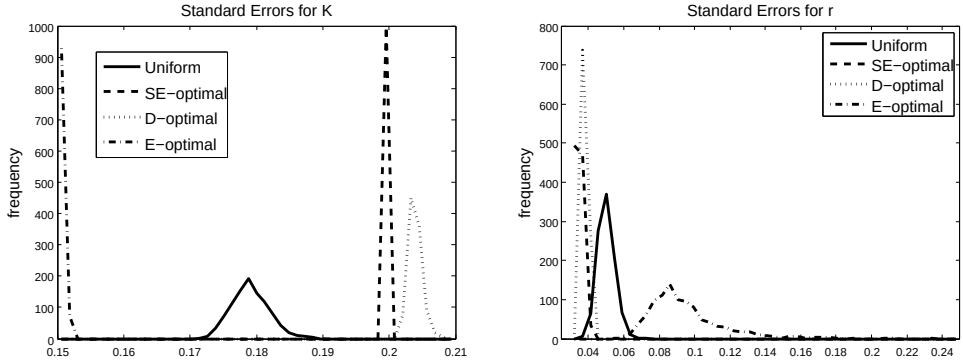


Figure 5.1: The distribution of optimal time points and uniform sampling time points plotted on the logistic curve. Optimal times points obtained using SolvOpt, with $N = 10$, and the optimal design methods SE-optimality, D-optimality, and E-optimality. Optimization with constraint implementation $(C2)$.

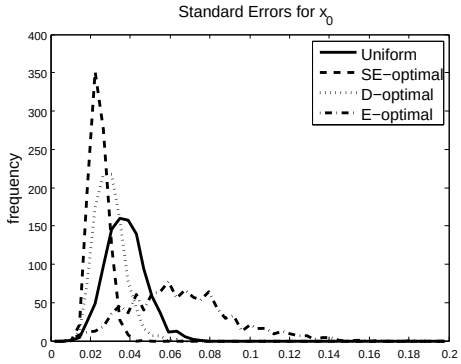
(for K). For $N = 15$ (Table 5.2), the closest average estimate of K came from E-optimal, for r the closest average estimate is from SE-optimal, and for x_0 it was D-optimal. Comparing the average and median estimates, we see that for all cases the averages and medians are very close, indicating that the parameter distributions are symmetric. However, for both $N = 10$ and $N = 15$ the averages were slightly larger than the medians for r and x_0 for all methods, implying that those parameter distributions are slightly skewed to the right (see Tables 5.1 and 5.2).

Comparing the standard errors (Tables 5.1 and 5.2 and Figs. 5.2 and 5.4): For K, we find that E-optimal has the smallest average standard errors, then the uniform grid, then SE-optimal when $N = 10$ or D-optimal when $N = 15$. For r and x_0, SE-optimal has the smallest average standard errors, followed by D-optimal, then the uniform grid. The average and median standard errors are very close. However, E-optimal's distribution of standard errors for r seems to be slightly right-skewed.

In conclusion, all of the optimal design methods produce parameter estimates that are close to the true value. In addition, the standard error estimates are similar when comparing the different optimal design methods. Based on the standard errors, E-optimal is more favorable for



(a) (b)



(c)

Figure 5.2: Using SolvOpt, with $N = 10$, a comparison of optimal design methods using SE-optimality, D-optimality, E-optimality, with a uniform sampling time points in terms of SE_K (panel (a)), SE_r (panel (b)), and SE_{x0} (panel (c)). Optimization with constraint implementation ($C2$).

the accuracy of K, and SE-optimal is more favorable for the accuracy of r and x_0 (followed closely by D-optimal).

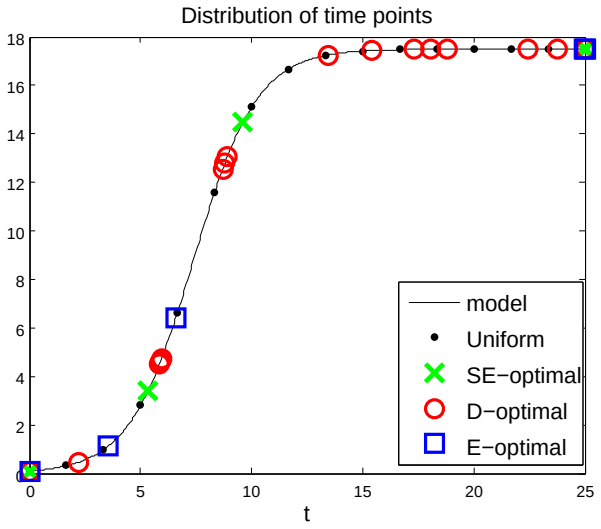
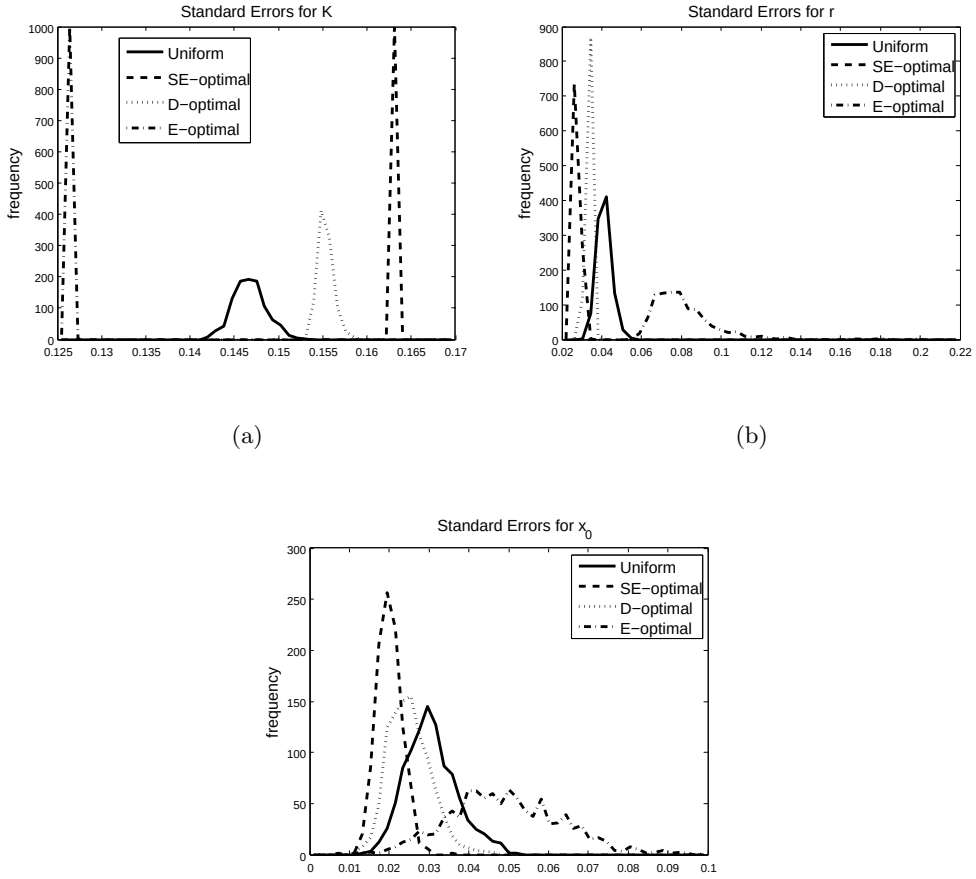


Figure 5.3: The distribution of optimal time points and uniform sampling time points plotted on the logistic curve. Optimal times points obtained using SolvOpt, with $N = 15$, and the optimal design methods SE-optimality, D-optimality, and E-optimality. Optimization with constraint implementation ($C2$).



(a)

(b)

(c)

Figure 5.4: Using SolvOpt, with $N = 15$, a comparison of optimal design methods using SE-optimality, D-optimality, E-optimality, with a uniform sampling time points in terms of SE_K (panel (a)), SE_r (panel (b)), and SE_{x0} (panel (c)). Optimization with constraint implementation ($C2$).

5.2 The Harmonic Oscillator Model

In our next example, we consider the harmonic oscillator, also known as the spring-mass-dashpot model. The model for the harmonic oscillator can be derived using Hooke's Law and mass-balance (see [8]) and is given by

$$m\ddot{x} + c\dot{x} + kx = 0, \quad \dot{x}(0) = x_1, \quad x(0) = x_2.$$

Here, m is mass, c is damping, and k is the spring constant. Dividing through by m, and defining $C = c/m$ and $K = k/m$, we can reduce the number of parameters.

$$\ddot{x} + C\dot{x} + Kx = 0, \quad \dot{x}(0) = x_1, \quad x(0) = x_2.$$

The analytical solution for the position at time t can be obtained and is given by

$$x(t) = e^{-at}\left(C_1 \cos bt + C_2 \sin bt\right),$$

where $C_1 = x_2$, $C_2 = (x_1 + ax_2)/b$, $a = \frac{1}{2}C$, and $b = \sqrt{K - \frac{1}{4}C^2}$. Substituting in C_1 and C_2, we obtain,

$$x(t) = x(t, \theta_0) = f(t, \theta_0) = e^{-at}\left(x_2 \cos bt + \frac{x_1 + ax_2}{b} \sin bt\right), \text{ for } 0 \le t \le T,$$

where for our considerations the true parameter vector is given by $\theta_0 = (C, K, x_1, x_2) = (0.1, 0.2, -1, 0.5)$ in our examples here.

5.2.1 Results for the Oscillator Model

The first way we will compare these optimal design methods, given that we know $\theta_0 = (C, K, x_1, x_2) = (0.1, 0.2, -1, 0.5)$ and $\sigma_0^2 = 0.16$, is to simply use their corresponding standard errors from the asymptotic theory, i.e., the values of $SE(\theta_0)$ given in (4.5). Recall that uncertainty is quantified by constructing confidence intervals using parameter estimates with the asymptotic standard error. Since our main focus here is the width of the confidence intervals, we can forgo the obtaining of the parameter estimates themselves which, for now, we tacitly assume may be similar for the three data sampling distributions we investigate here.

The optimal time points were found using the constrained nonlinear optimization algorithm SolvOpt. The optimal time points for each of the three optimal design methods are plotted with the model for different T and N under the first constraint implementation ($C1$) in Fig. 5.5, the

second constraint implementation ($C2$) in Fig. 5.6, the third constraint implementation ($C3$) in Fig. 5.7, and the last constraint implementation ($C4$) in Fig. 5.8. The standard errors (4.5) from the asymptotic theory corresponding to these optimal meshes are given in Table 5.3-5.6, respectively for the four different constraints.

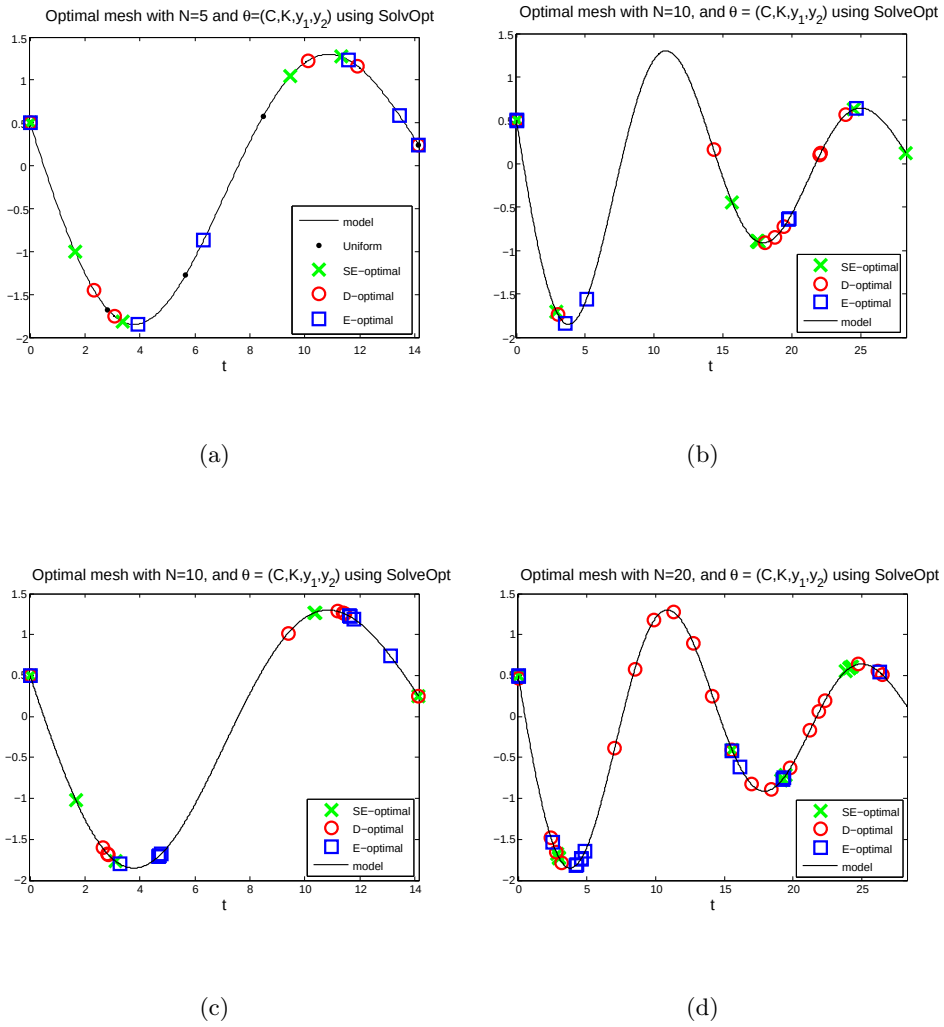


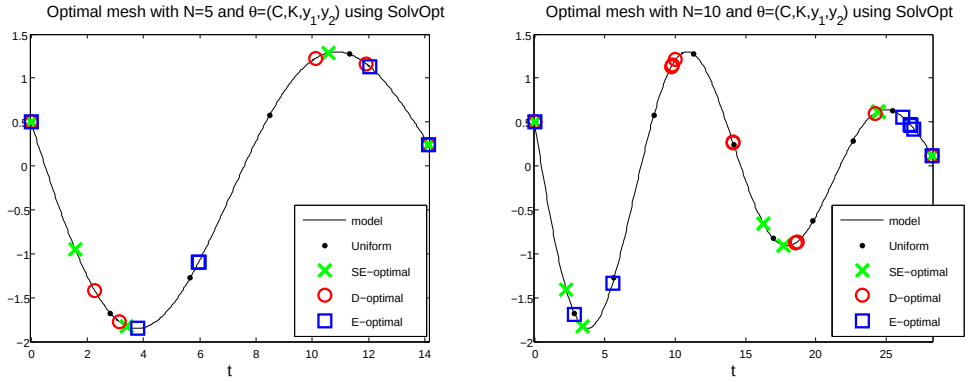
Figure 5.5: Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (C, K, x_1, x_2)$, with $T = 14.14$ (one period) for $N = 5$ (panel (a)) and $N = 10$ (panel (c)) and $T = 28.28$ (two periods) for $N = 10$ (panel (b)) and $N = 20$ (panel (d)). Optimization with constraint implementation ($C1$).

Table 5.3: Approximate asymptotic standard errors from the asymptotic theory (4.5) resulting from different optimal design methods for $\theta_0 = (C, K, x_1, x_2)$, optimization with constraint implementation $(C1)$.

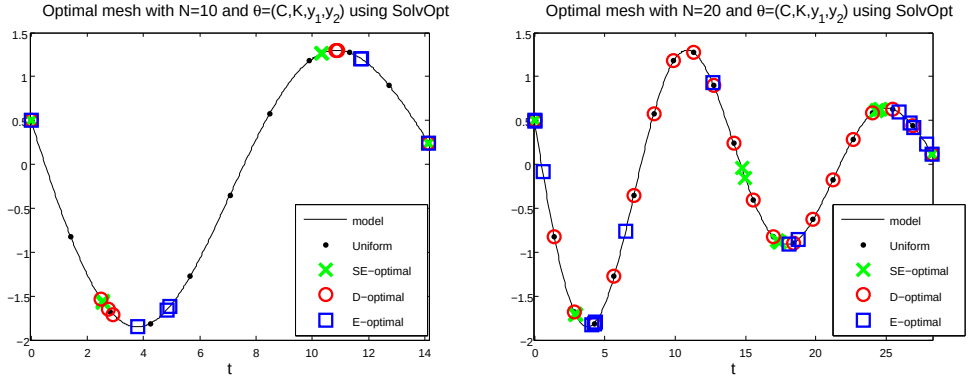
T	N	Method	$SE(C)$	$SE(K)$	$SE(x_1)$	$SE(x_2)$
14.14	5	SE-optimal	7.603×10^{-2}	4.320×10^{-2}	2.869×10^{-1}	3.714×10^{-1}
	5	D-optimal	8.244×10^{-2}	2.539×10^{-2}	2.551×10^{-1}	3.940×10^{-1}
	5	E-optimal	1.243×10^{-1}	2.508×10^{-2}	3.685×10^{-1}	3.815×10^{-1}
14.14	10	SE-optimal	5.527×10^{-2}	2.519×10^{-2}	2.113×10^{-1}	2.716×10^{-1}
	10	D-optimal	5.963×10^{-2}	1.845×10^{-2}	1.949×10^{-1}	2.821×10^{-1}
	10	E-optimal	1.136×10^{-1}	4.187×10^{-2}	2.193×10^{-1}	2.272×10^{-1}
28.28	10	SE-optimal	4.049×10^{-2}	1.980×10^{-2}	2.604×10^{-1}	2.305×10^{-1}
	10	D-optimal	3.919×10^{-2}	1.372×10^{-2}	1.936×10^{-1}	2.816×10^{-1}
	10	E-optimal	7.080×10^{-2}	2.343×10^{-2}	2.242×10^{-1}	2.274×10^{-1}
28.28	20	SE-optimal	2.438×10^{-2}	1.457×10^{-2}	1.517×10^{-1}	1.633×10^{-1}
	20	D-optimal	3.177×10^{-2}	1.102×10^{-2}	1.609×10^{-1}	2.632×10^{-1}
	20	E-optimal	4.422×10^{-2}	1.608×10^{-2}	1.355×10^{-1}	1.385×10^{-1}

Table 5.4: Approximate asymptotic standard errors from the asymptotic theory (4.5) resulting from different optimal design methods for $\theta_0 = (C, K, x_1, x_2)$, optimization with constraint implementation $(C2)$.

T	N	Method	$SE(C)$	$SE(K)$	$SE(x_1)$	$SE(x_2)$
14.14	5	SE-optimal	7.900×10^{-2}	2.657×10^{-2}	2.852×10^{-1}	3.657×10^{-1}
	5	D-optimal	8.251×10^{-2}	2.541×10^{-2}	2.561×10^{-1}	3.921×10^{-1}
	5	E-optimal	1.371×10^{-1}	2.900×10^{-2}	3.583×10^{-1}	3.736×10^{-1}
14.14	10	SE-optimal	5.667×10^{-2}	2.484×10^{-2}	1.964×10^{-1}	2.310×10^{-1}
	10	D-optimal	6.055×10^{-2}	1.648×10^{-2}	1.986×10^{-1}	2.822×10^{-1}
	10	E-optimal	8.507×10^{-2}	2.657×10^{-2}	2.211×10^{-1}	2.283×10^{-1}
28.28	10	SE-optimal	3.430×10^{-2}	2.149×10^{-2}	1.970×10^{-1}	2.274×10^{-1}
	10	D-optimal	7.445×10^{-2}	1.711×10^{-2}	4.314×10^{-1}	3.919×10^{-1}
	10	E-optimal	8.826×10^{-2}	2.532×10^{-2}	2.132×10^{-1}	2.169×10^{-1}
28.28	20	SE-optimal	2.457×10^{-2}	1.500×10^{-2}	1.516×10^{-1}	1.784×10^{-1}
	20	D-optimal	3.254×10^{-2}	1.166×10^{-2}	1.722×10^{-1}	2.867×10^{-1}
	20	E-optimal	5.135×10^{-2}	1.628×10^{-2}	1.451×10^{-1}	1.492×10^{-1}

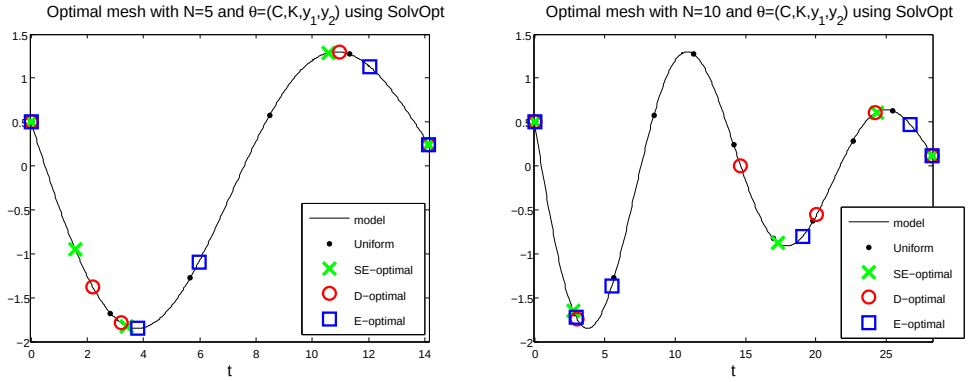


(a) (b)

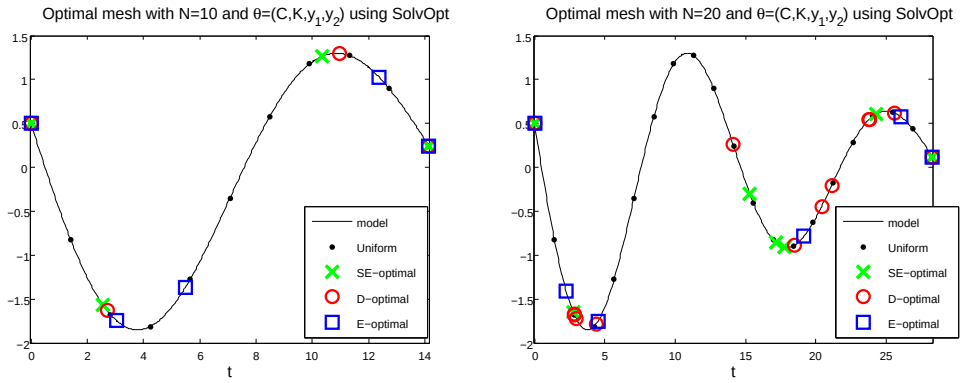


(c) (d)

Figure 5.6: Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (C, K, x_1, x_2)$, with $T = 14.14$ (one period) for $N = 5$ (panel (a)) and $N = 10$ (panel (c)) and $T = 28.28$ (two periods) for $N = 10$ (panel (b)) and $N = 20$ (panel (d)). Optimization with constraint implementation $(C2)$.



(a) (b)



(c) (d)

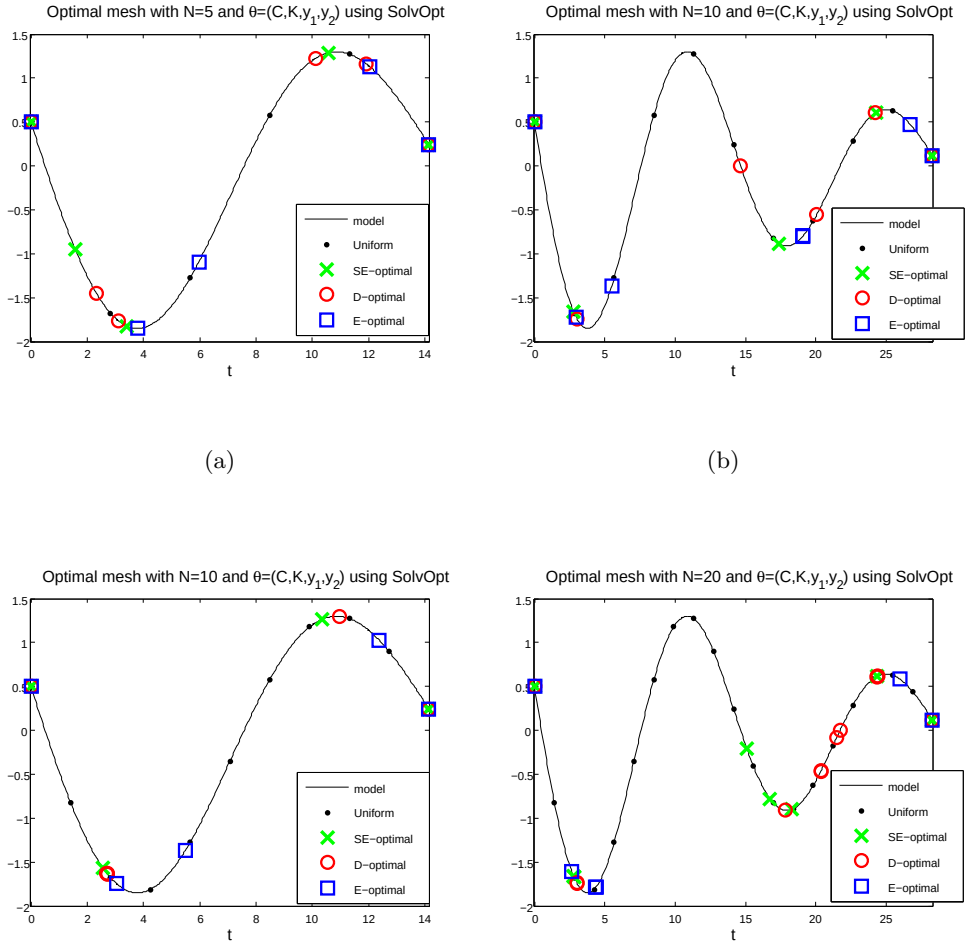
Figure 5.7: Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (C, K, x_1, x_2)$, with $T = 14.14$ (one period) for $N = 5$ (panel (a)) and $N = 10$ (panel (c)) and $T = 28.28$ (two periods) for $N = 10$ (panel (b)) and $N = 20$ (panel (d)). Optimization with constraint implementation $(C3)$.

Table 5.5: Approximate asymptotic standard errors from the asymptotic theory (4.5) resulting from different optimal design methods for $\theta_0 = (C, K, x_1, x_2)$, optimization with constraint implementation ($C3$).

T	N	Method	$SE(C)$	$SE(K)$	$SE(x_1)$	$SE(x_2)$
14.14	5	SE-optimal	7.900×10^{-2}	2.657×10^{-2}	2.852×10^{-1}	3.657×10^{-1}
	5	D-optimal	9.483×10^{-2}	2.106×10^{-2}	2.675×10^{-1}	3.898×10^{-1}
	5	E-optimal	1.371×10^{-1}	2.900×10^{-2}	3.583×10^{-1}	3.736×10^{-1}
14.14	10	SE-optimal	5.666×10^{-2}	2.484×10^{-2}	1.963×10^{-1}	2.309×10^{-1}
	10	D-optimal	6.071×10^{-2}	1.656×10^{-2}	1.978×10^{-1}	2.828×10^{-1}
	10	E-optimal	1.125×10^{-1}	2.838×10^{-2}	2.532×10^{-1}	2.639×10^{-1}
28.28	10	SE-optimal	3.673×10^{-2}	2.399×10^{-2}	1.925×10^{-1}	2.000×10^{-1}
	10	D-optimal	3.764×10^{-2}	1.373×10^{-2}	1.881×10^{-1}	2.812×10^{-1}
	10	E-optimal	7.949×10^{-2}	2.509×10^{-2}	2.154×10^{-1}	2.183×10^{-1}
28.28	20	SE-optimal	2.671×10^{-2}	1.812×10^{-2}	1.368×10^{-1}	1.413×10^{-1}
	20	D-optimal	2.882×10^{-2}	1.057×10^{-2}	1.176×10^{-1}	1.959×10^{-1}
	20	E-optimal	6.467×10^{-2}	2.604×10^{-2}	1.361×10^{-1}	1.376×10^{-1}

Table 5.6: Approximate asymptotic standard errors from the asymptotic theory (4.5) resulting from different optimal design methods for $\theta_0 = (C, K, x_1, x_2)$, optimization with constraint implementation ($C4$).

T	N	Method	$SE(C)$	$SE(K)$	$SE(x_1)$	$SE(x_2)$
14.14	5	SE-optimal	7.900×10^{-2}	2.657×10^{-2}	2.852×10^{-1}	3.657×10^{-1}
	5	D-optimal	8.249×10^{-2}	2.538×10^{-2}	2.553×10^{-1}	3.935×10^{-1}
	5	E-optimal	1.371×10^{-1}	2.900×10^{-2}	3.583×10^{-1}	3.736×10^{-1}
14.14	10	SE-optimal	5.666×10^{-2}	2.484×10^{-2}	1.963×10^{-1}	2.309×10^{-1}
	10	D-optimal	6.073×10^{-2}	1.657×10^{-2}	1.978×10^{-1}	2.828×10^{-1}
	10	E-optimal	1.125×10^{-1}	2.838×10^{-2}	2.532×10^{-1}	2.639×10^{-1}
28.28	10	SE-optimal	3.554×10^{-2}	2.395×10^{-2}	1.906×10^{-1}	2.000×10^{-1}
	10	D-optimal	3.765×10^{-2}	1.373×10^{-2}	1.881×10^{-1}	2.812×10^{-1}
	10	E-optimal	7.948×10^{-2}	2.509×10^{-2}	2.154×10^{-1}	2.183×10^{-1}
28.28	20	SE-optimal	2.512×10^{-2}	1.698×10^{-2}	1.348×10^{-1}	1.510×10^{-1}
	20	D-optimal	2.920×10^{-2}	1.035×10^{-2}	1.221×10^{-1}	1.788×10^{-1}
	20	E-optimal	6.095×10^{-2}	2.597×10^{-2}	1.300×10^{-1}	1.315×10^{-1}



(a) (b)

(c) (d)

Figure 5.8: Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (C, K, x_1, x_2)$, with $T = 14.14$ (one period) for $N = 5$ (panel (a)) and $N = 10$ (panel (c)) and $T = 28.28$ (two periods) for $N = 10$ (panel (b)) and $N = 20$ (panel (d)). Optimization with constraint implementation $(C4)$.

5.2.2 Discussion for the Oscillator Model

The constrained optimization algorithm, SolvOpt, was chosen over MATLAB's *fmincon* for comparisons using the harmonic oscillator example because it overall resulted in more well-behaved standard errors (real and finite values), and *fmincon* often did not.

In most cases, optimal meshes with a larger number of points were nested in the optimal meshes with a reduced the number of points. In some cases for $T = 28.28$ (Figs 5.5 and 5.6) doubling the number of points resulted in extra points being dispersed to otherwise empty regions, while other points were nested in the optimal mesh with fewer points. Often the larger number of points in the optimal mesh resulted in smaller standard errors.

Examining the asymptotic standard errors (Tables 5.3-5.6), different optimal sampling distributions produced the smallest standard errors for different parameters, with no optimal design method having consistently smaller standard errors. For C, most of the time SE-optimal had the smallest standard error, then D-optimal. For K, D-optimal most often had the smallest standard error, followed by SE-optimal. For x_1, D-optimal had the smallest standard errors in most cases. For x_2, either SE-optimal or E-optimal had the smallest standard errors.

The standard errors from the different optimal design methods were usually on the same order of magnitude. No method was always the best while comparing asymptotic standard errors, though for specific parameters some optimal sampling distributions were favorable.

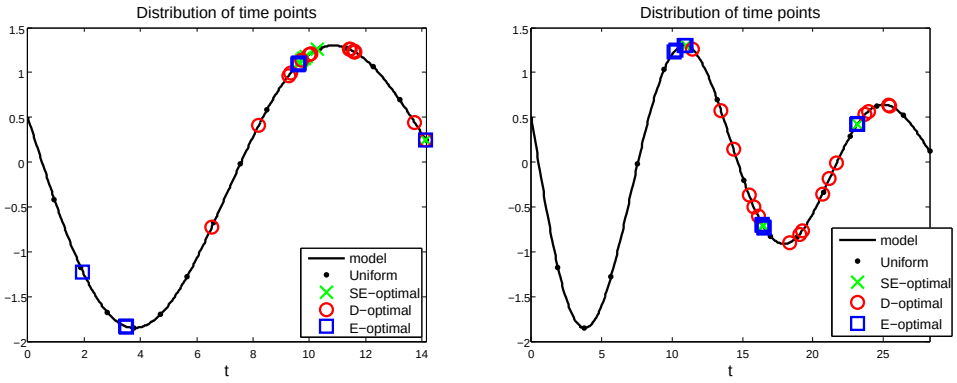
Since the asymptotic standard errors appear explicitly in the cost function we are minimizing for SE-optimal design, it may not be fair to compare these methods based on their asymptotic standard errors. To account for any possible bias in our comparison, we will compare these optimal design methods in the next section using simulated data and the inverse problem to estimate parameters using asymptotic theory and bootstrapping. In these computations, we will compare the optimal design methods based on how close their parameter estimates are to the true parameters, and the values of their estimated standard errors and covariances.

5.2.3 Results for the Oscillator Model - with the Inverse Problem

We solve the inverse problem with the OLS formulation to obtain parameter estimates and standard errors from both asymptotic theory (4.6) and the bootstrapping method (4.9). We create simulated noisy data (in agreement with our statistical model (2.2)) corresponding to the optimal time meshes using true values $\theta_0 = (C, K, x_1, x_2) = (0.1, 0.2, -1, 0.5)$ and *iid* noise with $\mathcal{E}_j \sim \mathcal{N}(0, \sigma_0^2)$ and $\sigma_0^2 = 0.16$. In this section we only estimate a subset of the parameters $\theta = (C, K)$. In addition to the estimates and standard errors, we also report the estimated $\mathrm{Cov}(C, K)$ according to asymptotic theory (4.3) and bootstrapping (4.9). For comparison purposes we also present these results for a uniform grid using the same T and N.

The optimal time points for each of the three optimal design methods are plotted with

the model for $T = 14.14$ and $T = 28.28$ for $N = 15$ under the first constraint implementation $(C1)$ in Fig. 5.9, the second constraint implementation $(C2)$ in Fig. 5.10, the third constraint implementation $(C3)$ in Fig. 5.11, and the last constraint implementation $(C4)$ in Fig. 5.12. The estimates, standard errors, and covariance between parameters as estimated from the asymptotic theory (4.6) corresponding to these optimal meshes are given in Table 5.7, 5.9, 5.11, and 5.13, respectively for the four different constraint implementations. The estimates, standard errors, and covariance between parameters when estimated from the bootstrapping method (4.9) corresponding to these optimal meshes are given in Table 5.8, 5.10, 5.12, and 5.14, respectively for the four different constraints. In each of the tables are also results on the uniform grid of time points for the same T and N. Since this is unaffected by constraints, the results for the uniform grid are repeated in the tables.



(a) (b)

Figure 5.9: Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (C, K)$, $N = 15$, with $T = 14.14$ (one period) (panel (a)) and $T = 28.28$ (two periods) (panel (b)). Optimization with constraint implementation $(C1)$.

5.2.4 Discussion of Oscillator Results with the Inverse Problem

The simulated data was created using the "true" parameter values $\theta_0 = (C, K) = (0.1, 0.2)$. So we can compare the optimal design methods based on how close the parameter estimates are as well as how large the estimates of the standard errors and covariances are.

Table 5.7: Estimates and standard errors from the asymptotic theory (4.6) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (C, K) = (0.1, 0.2)$ and $N = 15$, optimization with constraint implementation $(C1)$.

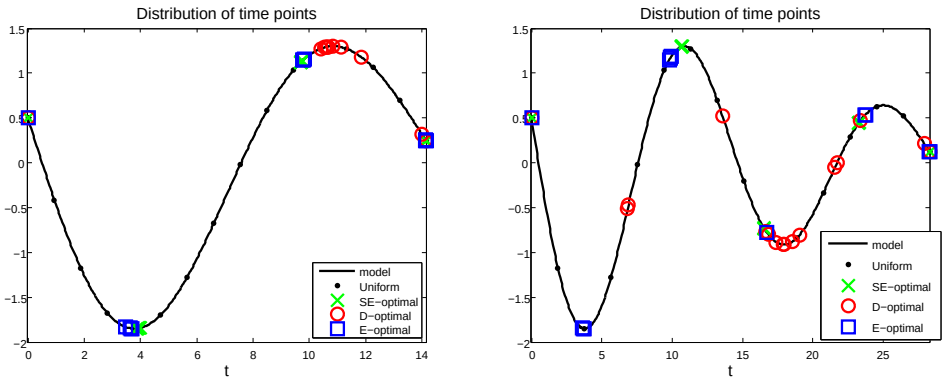
T	Method	$\widehat{C}_{asy}$	$\widehat{SE}(\widehat{C}_{asy})$	$\widehat{K}_{asy}$	$\widehat{SE}(\widehat{K}_{asy})$	$\widehat{\mathrm{Cov}}(\widehat{C}_{asy}, \widehat{K}_{asy})$
14.14	SE-optimal	0.0865	1.369×10^{-2}	0.1979	1.165×10^{-2}	-3.597×10^{-5}
14.14	D-optimal	0.1112	2.104×10^{-2}	0.2038	8.974×10^{-3}	-1.027×10^{-4}
14.14	E-optimal	0.0592	3.009×10^{-2}	0.1736	1.285×10^{-2}	-9.801×10^{-5}
14.14	Uniform	0.1300	3.529×10^{-2}	0.1938	1.278×10^{-2}	-2.803×10^{-4}
28.28	SE-optimal	0.1111	3.221×10^{-2}	0.2040	2.827×10^{-2}	-3.391×10^{-4}
28.28	D-optimal	0.0705	1.710×10^{-2}	0.1974	7.444×10^{-3}	-6.045×10^{-5}
28.28	E-optimal	0.0843	1.664×10^{-2}	0.1953	1.381×10^{-2}	4.378×10^{-5}
28.28	Uniform	0.0854	1.792×10^{-2}	0.2122	7.326×10^{-3}	-6.219×10^{-5}

Table 5.8: Estimates and standard errors from the bootstrap method (4.9) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (C, K) = (0.1, 0.2)$, $M = 1000$ bootstraps and $N = 15$, optimization with constraint implementation $(C1)$.

T	Method	$\widehat{C}_{boot}$	$\widehat{SE}(\widehat{C}_{boot})$	$\widehat{K}_{boot}$	$\widehat{SE}(\widehat{K}_{boot})$	$\widehat{\mathrm{Cov}}(\widehat{C}_{boot}, \widehat{K}_{boot})$
14.14	SE-optimal	0.0871	1.460×10^{-2}	0.1988	1.092×10^{-2}	8.013×10^{-5}
14.14	D-optimal	0.1035	1.565×10^{-2}	0.2025	8.225×10^{-3}	-1.731×10^{-5}
14.14	E-optimal	0.0603	2.861×10^{-2}	0.1743	1.297×10^{-2}	6.383×10^{-5}
14.14	Uniform	0.1170	2.469×10^{-2}	0.1989	1.009×10^{-2}	-4.978×10^{-5}
28.28	SE-optimal	0.0827	2.486×10^{-2}	0.1991	1.624×10^{-2}	1.567×10^{-4}
28.28	D-optimal	0.0705	1.428×10^{-2}	0.1972	7.177×10^{-3}	-5.355×10^{-6}
28.28	E-optimal	0.0843	2.138×10^{-2}	0.2035	2.598×10^{-2}	3.099×10^{-6}
28.28	Uniform	0.0837	1.475×10^{-2}	0.2122	6.350×10^{-3}	-4.436×10^{-6}

For asymptotic estimates:

Comparing optimal design methods based on which has parameter estimates closest to the true values, there is no method that is always the best. For constraint implementation $(C1)$, with $T = 14.14$ (Table 5.7) the closest parameter estimates result from either SE-optimal or D-optimal, for $T = 28.28$ no method is consistently closest. For constraint implementation



(a) (b)

Figure 5.10: Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (C, K)$, $N = 15$, with $T = 14.14$ (one period) (panel (a)) and $T = 28.28$ (two periods) (panel (b)). Optimization with constraint implementation $(C2)$.

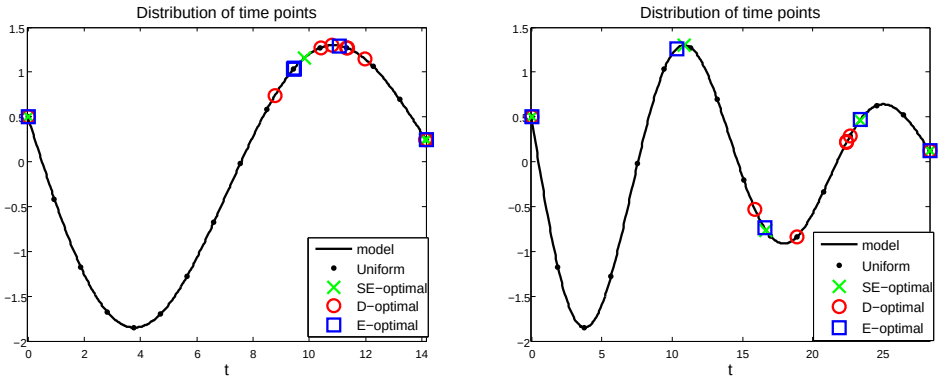
Table 5.9: Estimates and standard errors from the asymptotic theory (4.6) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (C, K) = (0.1, 0.2)$ and $N = 15$, optimization with constraint implementation $(C2)$.

T	Method	$\widehat{C}_{asy}$	$\widehat{SE}(\widehat{C}_{asy})$	$\widehat{K}_{asy}$	$\widehat{SE}(\widehat{K}_{asy})$	$\widehat{\text{Cov}}(\widehat{C}_{asy}, \widehat{K}_{asy})$
14.14	SE-optimal	0.0841	2.852×10^{-2}	0.2314	1.996×10^{-2}	-3.540×10^{-4}
14.14	D-optimal	0.0934	2.635×10^{-2}	0.2054	9.968×10^{-3}	-1.414×10^{-4}
14.14	E-optimal	0.1076	2.220×10^{-2}	0.1952	1.060×10^{-2}	-9.008×10^{-5}
14.14	Uniform	0.1300	3.529×10^{-2}	0.1938	1.278×10^{-2}	-2.803×10^{-4}
28.28	SE-optimal	0.0649	1.440×10^{-2}	0.1842	7.006×10^{-3}	2.883×10^{-6}
28.28	D-optimal	0.1088	1.888×10^{-2}	0.2086	8.425×10^{-3}	-6.880×10^{-5}
28.28	E-optimal	0.1115	2.397×10^{-2}	0.2046	2.073×10^{-2}	-1.256×10^{-4}
28.28	Uniform	0.0854	1.792×10^{-2}	0.2122	7.326×10^{-3}	-6.219×10^{-5}

$(C2)$ (Table 5.9), either D-optimal or E-optimal had the closest parameter estimates to the true values when looking at results from $T = 14.14$ and $T = 28.28$. For constraint implementation $(C3)$ (Table 5.11), either D-optimal or E-optimal had the closest parameter estimate for C, and either SE-optimal or D-optimal has the closest estimate for K. For constraint implementation

Table 5.10: Estimates and standard errors from the bootstrap method (4.9) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (C, K) = (0.1, 0.2)$, $M = 1000$ bootstraps and $N = 15$, optimization with constraint implementation $(C2)$.

T	Method	$\widehat{C}_{boot}$	$\widehat{SE}(\widehat{C}_{boot})$	$\widehat{K}_{boot}$	$\widehat{SE}(\widehat{K}_{boot})$	$\widehat{\text{Cov}}(\widehat{C}_{boot}, \widehat{K}_{boot})$
14.14	SE-optimal	0.0783	2.076×10^{-2}	0.2320	1.628×10^{-2}	4.751×10^{-5}
14.14	D-optimal	0.0976	2.243×10^{-2}	0.2070	9.921×10^{-3}	-4.040×10^{-5}
14.14	E-optimal	0.1031	1.930×10^{-2}	0.1956	9.636×10^{-3}	3.043×10^{-5}
14.14	Uniform	0.1170	2.469×10^{-2}	0.1989	1.009×10^{-2}	-4.978×10^{-5}
28.28	SE-optimal	0.0576	1.479×10^{-2}	0.1842	6.057×10^{-3}	3.937×10^{-5}
28.28	D-optimal	0.1194	1.694×10^{-2}	0.2105	8.317×10^{-3}	4.750×10^{-6}
28.28	E-optimal	0.0947	2.161×10^{-2}	0.2045	1.927×10^{-2}	1.499×10^{-4}
28.28	Uniform	0.0837	1.475×10^{-2}	0.2122	6.350×10^{-3}	-4.436×10^{-6}



(a) (b)

Figure 5.11: Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (C, K)$, $N = 15$, with $T = 14.14$ (one period) (panel (a)) and $T = 28.28$ (two periods) (panel (b)). Optimization with constraint implementation $(C3)$.

$(C4)$ (Table 5.13), SE-optimal and D-optimal had the closest estimates for $T = 14.14$, and E-optimal had the closest estimates for $T = 28.28$.

Comparing the optimal design methods based on the estimated standard errors and covariance between parameters, we find that no method is always the best. For constraint

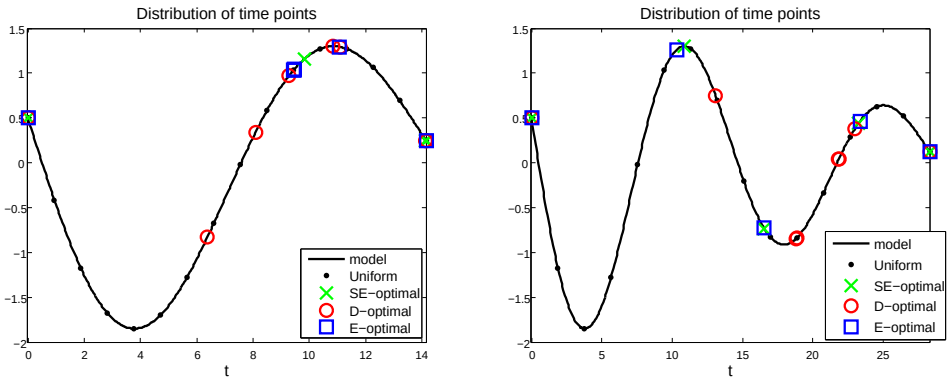
Table 5.11: Estimates and standard errors from the asymptotic theory (4.6) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (C, K) = (0.1, 0.2)$ and $N = 15$, optimization with constraint implementation $(C3)$.

T	Method	$\widehat{C}_{asy}$	$\widehat{SE}(\widehat{C}_{asy})$	$\widehat{K}_{asy}$	$\widehat{SE}(\widehat{K}_{asy})$	$\widehat{\text{Cov}}(\widehat{C}_{asy}, \widehat{K}_{asy})$
14.14	SE-optimal	0.1238	2.515×10^{-2}	0.2028	2.302×10^{-2}	-1.640×10^{-4}
14.14	D-optimal	0.0970	2.061×10^{-2}	0.1997	7.973×10^{-3}	-9.382×10^{-5}
14.14	E-optimal	0.1156	2.204×10^{-2}	0.1953	2.055×10^{-2}	6.635×10^{-5}
14.14	Uniform	0.1300	3.529×10^{-2}	0.1938	1.278×10^{-2}	-2.803×10^{-4}
28.28	SE-optimal	0.0899	1.617×10^{-2}	0.2015	1.368×10^{-2}	-5.288×10^{-5}
28.28	D-optimal	0.0966	1.540×10^{-2}	0.2084	6.787×10^{-3}	-4.422×10^{-5}
28.28	E-optimal	0.1029	1.705×10^{-2}	0.2098	2.111×10^{-2}	-1.575×10^{-4}
28.28	Uniform	0.0854	1.792×10^{-2}	0.2122	7.326×10^{-3}	-6.219×10^{-5}

Table 5.12: Estimates and standard errors from the bootstrap method (4.9) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (C, K) = (0.1, 0.2)$, $M = 1000$ bootstraps and $N = 15$, optimization with constraint implementation $(C3)$.

T	Method	$\widehat{C}_{boot}$	$\widehat{SE}(\widehat{C}_{boot})$	$\widehat{K}_{boot}$	$\widehat{SE}(\widehat{K}_{boot})$	$\widehat{\text{Cov}}(\widehat{C}_{boot}, \widehat{K}_{boot})$
14.14	SE-optimal	0.1204	2.652×10^{-2}	0.2047	2.186×10^{-2}	3.199×10^{-4}
14.14	D-optimal	0.0919	1.574×10^{-2}	0.1978	7.301×10^{-3}	-2.363×10^{-5}
14.14	E-optimal	0.1069	2.756×10^{-2}	0.1978	1.967×10^{-2}	3.763×10^{-4}
14.14	Uniform	0.1170	2.469×10^{-2}	0.1989	1.009×10^{-2}	-4.978×10^{-5}
28.28	SE-optimal	0.0870	1.753×10^{-2}	0.2028	1.542×10^{-2}	8.280×10^{-5}
28.28	D-optimal	0.0906	1.133×10^{-2}	0.2030	5.540×10^{-3}	5.491×10^{-6}
28.28	E-optimal	0.0926	1.783×10^{-2}	0.2113	2.202×10^{-2}	4.422×10^{-5}
28.28	Uniform	0.0837	1.475×10^{-2}	0.2122	6.350×10^{-3}	-4.436×10^{-6}

implementation $(C1)$ (Table 5.7), when $T = 14.14$ SE-optimal had the smallest standard errors and covariance, when $T = 28.28$ either E-optimal or D-optimal had the smallest standard errors and covariances. For constraint implementation $(C2)$ (Table 5.9), the smallest standard errors and covariances came from E-optimal when $T = 14.14$ and SE-optimal when $T = 28.28$, and followed by D-optimal in both cases. For constraint implementation $(C3)$ (Table 5.11), the



(a) (b)

Figure 5.12: Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (C, K)$, $N = 15$, with $T = 14.14$ (one period) (panel (a)) and $T = 28.28$ (two periods) (panel (b)). Optimization with constraint implementation $(C4)$.

Table 5.13: Estimates and standard errors from the asymptotic theory (4.6) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (C, K) = (0.1, 0.2)$ and $N = 15$, optimization with constraint implementation $(C4)$.

T	Method	$\widehat{C}_{asy}$	$\widehat{SE}(\widehat{C}_{asy})$	$\widehat{K}_{asy}$	$\widehat{SE}(\widehat{K}_{asy})$	$\widehat{\mathrm{Cov}}(\widehat{C}_{asy}, \widehat{K}_{asy})$
14.14	SE-optimal	0.0837	2.312×10^{-2}	0.2136	2.087×10^{-2}	-2.385×10^{-4}
14.14	D-optimal	0.0771	2.206×10^{-2}	0.1827	8.461×10^{-3}	-1.074×10^{-4}
14.14	E-optimal	0.0343	1.387×10^{-2}	0.1719	7.166×10^{-3}	2.825×10^{-5}
14.14	Uniform	0.1300	3.529×10^{-2}	0.1938	1.278×10^{-2}	-2.803×10^{-4}
28.28	SE-optimal	0.0908	1.473×10^{-2}	0.2206	1.330×10^{-2}	-1.480×10^{-4}
28.28	D-optimal	0.1160	2.358×10^{-2}	0.1875	9.149×10^{-3}	-9.691×10^{-5}
28.28	E-optimal	0.0964	1.218×10^{-2}	0.2070	1.445×10^{-2}	-4.984×10^{-5}
28.28	Uniform	0.0854	1.792×10^{-2}	0.2122	7.326×10^{-3}	-6.219×10^{-5}

smallest standard errors and covariances came from D-optimal or E-optimal when $T = 14.14$, and from D-optimal followed by SE-optimal when $T = 28.28$. For constraint implementation $(C4)$ when $T = 14.14$, E-optimal had the smallest standard errors and covariances followed by D-optimal.

Table 5.14: Estimates and standard errors from the bootstrap method (4.9) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (C, K) = (0.1, 0.2)$, $M = 1000$ bootstraps and $N = 15$, optimization with constraint implementation $(C4)$.

T	Method	$\widehat{C}_{boot}$	$\widehat{SE}(\widehat{C}_{boot})$	$\widehat{K}_{boot}$	$\widehat{SE}(\widehat{K}_{boot})$	$\widehat{\text{Cov}}(\widehat{C}_{boot}, \widehat{K}_{boot})$
14.14	SE-optimal	0.0856	2.185×10^{-2}	0.2184	2.027×10^{-2}	1.904×10^{-4}
14.14	D-optimal	0.0658	1.611×10^{-2}	0.1822	7.297×10^{-3}	-2.611×10^{-5}
14.14	E-optimal	0.0334	1.769×10^{-2}	0.1729	6.841×10^{-3}	7.838×10^{-5}
14.14	Uniform	0.1170	2.469×10^{-2}	0.1989	1.009×10^{-2}	-4.978×10^{-5}
28.28	SE-optimal	0.0835	8.868×10^{-3}	0.2111	7.826×10^{-3}	2.677×10^{-5}
28.28	D-optimal	0.1265	1.872×10^{-2}	0.1986	9.266×10^{-3}	-1.156×10^{-5}
28.28	E-optimal	0.0963	1.594×10^{-2}	0.2195	2.443×10^{-2}	-1.188×10^{-4}
28.28	Uniform	0.0837	1.475×10^{-2}	0.2122	6.350×10^{-3}	-4.436×10^{-6}

For bootstrap estimates:

Comparing optimal design methods based on which has bootstrapping parameter estimates closest to the true value, again no method is always the best. For constraint implementations $(C1)$ and $(C4)$ (Tables 5.8 and 5.14), when $T = 14.14$ either SE-optimal or D-optimal have the closest estimates. For constraint implementation $(C2)$ (Table 5.10), either D-optimal or E-optimal had parameter estimates closest to the true values. For $T = 14.14$ (Table 5.10), the parameter estimate for K was in fact closest from the uniform mesh, followed by D-optimal. For constraint implementation $(C3)$ (Table 5.12), when $T = 14.14$ either D-optimal or E-optimal had the closest estimates. For cases that were not reported, there was no method that was consistently better in terms of closeness of parameter estimates to the true values.

Comparing optimal design methods based on which method produces the smallest bootstrapping estimated standard errors and parameter estimates, no method is consistently favorable. For constraint implementation $(C1)$ (Table 5.8), D-optimal has the smallest standard errors and covariances. For constraint implementation $(C2)$ (Table 5.10), when $T = 14.14$ the smallest standard errors and covariances come from E-optimal, when $T = 28.28$ either SE-optimal or the uniform grim had the smallest standard errors and covariances, followed by D-optimal. For constraint implementation $(C3)$ (Table 5.12), the smallest standard errors and covariances are from D-optimal, followed by SE-optimal. For constraint implementation $(C4)$ (Table 5.14), when $T = 14.14$ the smallest standard errors and covariances come from D-optimal followed by E-optimal, when $T = 28.28$ either SE-optimal or D-optimal were the smallest.

In conclusion, all of the optimal design methods are favorable under specific conditions. In many of the cases the parameter estimates, standard errors, and covariances are on the same order of magnitude resulting from different optimal design criteria.

5.3 A Simple Glucose Regulation Model

Next we will consider a well-known model for the intervenous glucose tolerance test (IVGTT). This model is referred to as the *minimal model* in the literature [11, 17, 32]. Prior to the IVGTT the patient is asked to fast. When the patient comes in for the IVGTT, measurements of their baseline glucose and insulin concentrations, G_b and I_b, respectively, are first taken. The IVGTT procedure consists of injecting a bolus resulting in an initial concentration p_0 of glucose into the blood, and measuring the glucose and insulin concentrations in the blood at various time points after the injection.

The body carefully regulates the glucose concentration in the blood within a narrow range. Extremely high blood glucose concentration is referred to as hyperglycemia, whereas hypoglycemia results when the blood glucose concentration is too low. The IVGTT initially brings the blood glucose concentration to hyperglycemic levels. In normal healthy patients, the high level of glucose in the blood signals the beta cells of the pancreas to secrete insulin. Insulin helps the fat and muscle cells to uptake glucose from the blood, either for fuel or for storage as glycogen. When the blood glucose concentration is too low, the pancreas secretes glucagon which releases glucose stored in the liver into the blood. Glucagon is another dynamic variable [4] during the IVGTT. Though glucagon is not included in this model, it is acknowledged that the liver can regulate glucose independently from insulin through glucagon.

5.3.1 Model

The minimal model is given by the following system of ordinary differential equations (see [11, 17, 32] for details):

$$\dot{G}(t) = -p_1\big(G(t) - G_b\big) - X(t)G(t), \quad G(0) = p_0, \tag{5.1}$$

$$\dot{X}(t) = -p_2 X(t) + p_3\big(I(t) - I_b\big), \quad X(0) = 0, \tag{5.2}$$

$$\dot{I}(t) = p_4 t \max\big(0, G(t) - p_5\big) - p_6\big(I(t) - I_b\big), \quad I(0) = p_7 + I_b, \tag{5.3}$$

where $G(t)$ is the glucose concentration (in $\mathsf{mg/dl}$) in plasma at time t, $I(t)$ is the insulin concentration (in $\mu\mathsf{U/ml}$) in plasma at time t and $X(t)$ represents insulin-dependent glucose uptake activity (proportional to a remote insulin compartment) in units $\mathsf{1/min}$.

We use the following approximate max function in equation (5.3) since it is continuously differentiable:

$$\text{maxfunc}_1(v) = \begin{cases} v & \text{for } v > \epsilon_0, \\ 0 & \text{for } v < -\epsilon_0, \\ \dfrac{1}{4\epsilon_0}\big(v + \epsilon_0\big)^2 & \text{for } v \in [-\epsilon_0, \epsilon_0], \end{cases}$$

where $\epsilon_0 > 0$ is chosen sufficiently small (for example, $\epsilon_0 = 10^{-5}$).

An interpretation of the parameters is given in Table 5.15.

Table 5.15: Description of model parameters and typical values.

θ	Description	value
G_b	basal pre-injection level of glucose	83.7 mg/dl
I_b	basal pre-injection level of insulin	11 μU/ml
p_0	the theoretical glucose concentration in plasma at time $t = 0$	279 mg/dl
p_1	the rate constant of insulin-independent glucose uptake in muscles, and adipose tissue	2.6×10^{-2} min^{-1}
p_2	the rate constant for decrease in tissue glucose uptake ability	0.025 min^{-1}
p_3	the rate constant for the insulin-dependent increase in glucose uptake ability in tissue per unit of insulin concentration above I_b	1.25×10^{-5} min^{-2}(μU/ml)$^{-1}$
p_4	the rate constant for insulin secretion by the pancreatic β-cells after the glucose injection and with glucose concentration above p_5	4.1×10^{-3} (μU/ml) min^{-2}(mg/dl)$^{-1}$
p_5	the threshold value of glucose in plasma above which the pancreatic β-cells secrete insulin	83.7 mg/dl
p_6	the first order decay rate for insulin in plasma	0.27 min^{-1}
p_7	$p_7 + I_b$ is the theoretical insulin concentration in plasma at time $t = 0$	352.7 μU/ml

In the following we will describe the model and its underlying assumptions.

Equation (5.1) (Glucose concentration in plasma)

At $t = 0$ a bolus of glucose is injected such that the initial glucose concentration in the blood is p_0. The first term represents hepatic glucose balance, which occurs independent of insulin level. The second term is the loss of glucose due to insulin-dependent uptake by peripheral tissues.

Equation (5.2) (Insulin-dependent glucose uptake activity)

At $t = 0$ there is no glucose uptake activity. Spontaneously, tissue loses the ability to uptake glucose, even in the presence of insulin. Glucose uptake activity increases proportionally to the amount by which insulin concentration is greater than baseline insulin concentration.

Equation (5.3) (Insulin concentration in the plasma)

At $t = 0$ the initial insulin concentration is at some level over baseline, given by $p_7 + I_b$. The increase in insulin concentration is proportional to the amount by which glucose concentration exceeds some threshold, p_5, and the amount of time that has elapsed since the glucose injection. There is a loss of insulin to degradation in the plasma. The pancreas secretes low levels of insulin, even in hypoglycemic conditions, to maintain insulin concentration at or above baseline I_b.

The analysis of this model found in [11, 32] gives a metabolic portrait for the first phase sensitivity to glucose (ϕ_1) (corresponding to initial secretion of insulin), the second phase glucose sensitivity (S_G) (corresponding to a secondary phase of insulin secretion), and the insulin sensitivity index (S_I). The metabolic portrait is given by

$$S_I = \frac{p_3}{p_2}, \quad S_G = p_1, \quad \phi_1 = \frac{I_{\max} - I_b}{p_6\big(p_0 - G_b\big)}, \tag{5.4}$$

where $I_{\max}$ is the maximal value of insulin concentration in plasma.

Bergman et al., [10] suggest the use of this model in the clinical IVGTT setting. Parameters from the model are estimated using patient-specific data. The parameter estimates are then used in the metabolic portrait for diabetes diagnosis purpose for that patient. This process was made readily available to clinicians in the computer software MINMOD [25]. Since the estimation of these parameters plays such a crucial role in the diagnosis, it appears that optimal design methods would be of great assistance. Data sampled at the optimal time points would result in a more accurate metabolic portrait produced by this mathematical model.

Next we will describe the corresponding statistical model for this system involving vector observations. We obtain numerical solutions using MATLAB's *ode45* since there does not exist an analytical solution to this system of differential equations. Let $\vec{z}(t, \theta_0) = (G(t, \theta_0), X(t, \theta_0), I(t, \theta_0))^T$

represent our model solution. Since we can observe realizations of $G(t,\theta_0)$ and $I(t,\theta_0)$, but not $X(t,\theta_0)$, our observation process is given by

$$\vec{y}(t) = \vec{f}(t,\theta_0) = (G(t,\theta_0), I(t,\theta_0))^T.$$

Our statistical model is given by the stochastic process

$$\vec{Y}(t) = \vec{f}(t,\theta_0) + \vec{\mathcal{E}}(t),$$

where $\vec{\mathcal{E}}(t)$ is a noisy vector random process. We assume the following about the vector random variable $\vec{\mathcal{E}}(t)$:

$$\begin{aligned}
\mathrm{E}(\vec{\mathcal{E}}(t)) &= 0, \quad t \in [0,T], \\
\mathrm{Var}\vec{\mathcal{E}}(t) &= \mathrm{diag}(\sigma^2_{0,G}, \sigma^2_{0,I}), \quad t \in [0,T], \\
\mathrm{Cov}(\mathcal{E}_1(t)\mathcal{E}_1(s)) &= \sigma^2_{0,G}\delta(t-s), \quad t,s \in [0,T], \\
\mathrm{Cov}(\mathcal{E}_2(t)\mathcal{E}_2(s)) &= \sigma^2_{0,I}\delta(t-s), \quad t,s \in [0,T], \\
\mathrm{Cov}(\mathcal{E}_1(t)\mathcal{E}_2(s)) &= 0, \quad t,s \in [0,T].
\end{aligned}$$

We assume constant variance, $\sigma^2_{0,G} = 25$ and $\sigma^2_{0,I} = 4$. A realization of the observation process is given by

$$\vec{y}(t) = \vec{f}(t,\theta_0) + \vec{\varepsilon}(t), \ t \in [0,T],$$

where the measurement error $\vec{\varepsilon}(t)$ is a realization of $\vec{\mathcal{E}}(t)$.

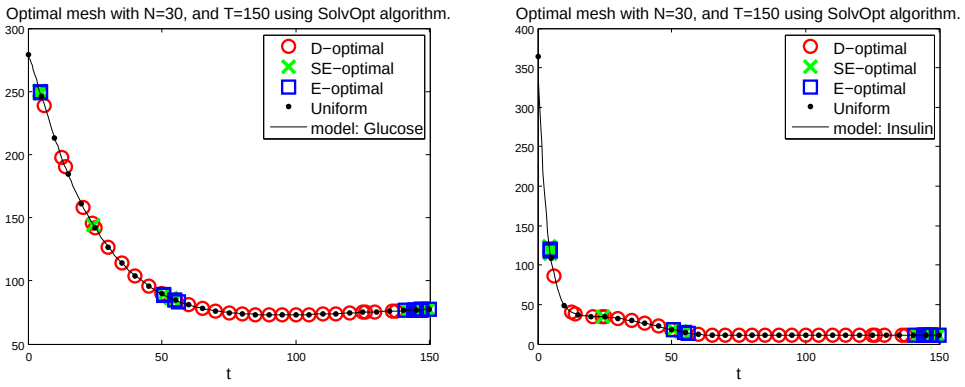
We compute the optimal time mesh using SE-optimality, D-optimality, and E-optimality for a subset of the parameters $\theta = (p_1, p_2, p_3, p_4)$, and a fixed number of time points ($N = 30$) and a final time of $T = 150$ minutes. We remark that a subset of parameters was chosen to avoid an ill-conditioned FIM. The subset of parameters was chosen based on the traditional sensitivity functions. The glucose and insulin model solutions were most sensitive to $\theta = (p_1, p_2, p_3, p_4)$. The approximate asymptotic standard errors (4.5) for $\theta = (p_1, p_2, p_3, p_4)$ were computed on the optimal mesh corresponding to an optimal design method (see Section 4.2.1).

The optimal design methods were implemented using the constrained minimization algorithm SolvOpt. The variations on the constraint employed were the same as in the previous section ($(C1) - (C4)$). We compare SE-optimal, D-optimal and E-optimal design methods based on these approximate asymptotic standard errors (4.5).

5.3.2 Results for the Glucose Regulation Model

The optimal time points (found using the SolvOpt algorithm) for each of the three optimal design methods are plotted with the model for $T = 150$ minutes and $N = 30$ under the first constraint implementation $(C1)$ in Fig. 5.13, the second constraint implementation $(C2)$ in Fig. 5.14, the third constraint implementation $(C3)$ in Fig. 5.15, and the last constraint implementation $(C4)$ in Fig. 5.16. The standard errors (4.5) from the asymptotic theory corresponding to these optimal meshes are given in Table 5.16-5.19, respectively for the four different constraint implementations.

Note that for constraint implementations $(C2)$ and $(C4)$ initializing SolvOpt with the uniform mesh resulted in a terminal error for D-optimal, stating that the gradient at the starting point was zero. In these cases other initial mesh points were chosen such that D-optimal's initial gradient was non-zero, and optimization could be achieved. To be consistent, all three methods were initialized by the same non-uniform mesh. For $(C2)$ the initial mesh was $\tau^0 = \{0, \ldots, 0, 10, 37, 150, \ldots, 150\}$, and for $(C4)$ it was $\tau^0 = \{5, 15, 19, 21, 24, 26, 42, 59, 63, 73, 82, 95, 98, 98, 102, 111, 114, 119, 120, 122, 127, 136, 137, 137, 140, 144, 144, 144, 145, 146\}$. Optimal design methods are guaranteed to converge in a local sense.

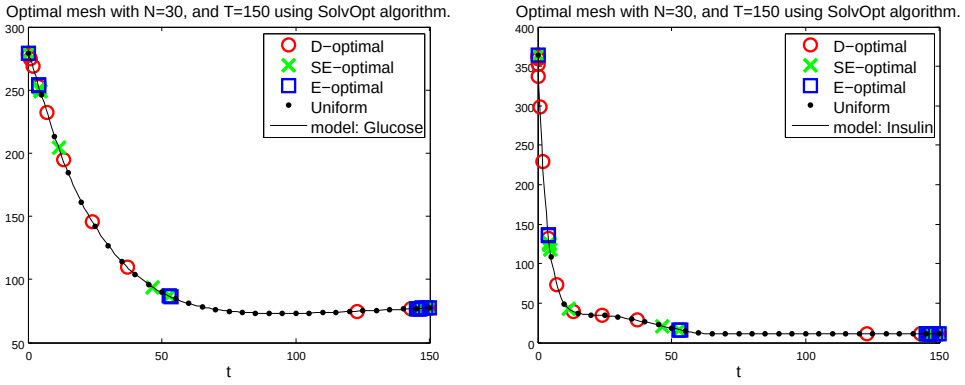


(a) (b)

Figure 5.13: Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (p_1, p_2, p_3, p_4)$, with $T = 150$ for $N = 30$. Optimal time points with the Glucose model in panel (a) and with the Insulin model in panel (b). Optimization, using SolvOpt, with constraint implementation $(C1)$.

Table 5.16: Approximate asymptotic standard errors from the asymptotic theory (4.5) resulting from different optimal design methods for $\theta_0 = (p_1, p_2, p_3, p_4)$, optimization, using SolvOpt, with constraint implementation $(C1)$.

Method	$SE(p_1)$	$SE(p_2)$	$SE(p_3)$	$SE(p_4)$
SE-optimal	4.173×10^{-3}	6.501×10^{-3}	3.100×10^{-6}	2.959×10^{-4}
D-optimal	8.411×10^{-3}	1.236×10^{-2}	6.133×10^{-6}	1.714×10^{-4}
E-optimal	4.381×10^{-3}	6.520×10^{-3}	3.182×10^{-6}	4.941×10^{-4}



(a) (b)

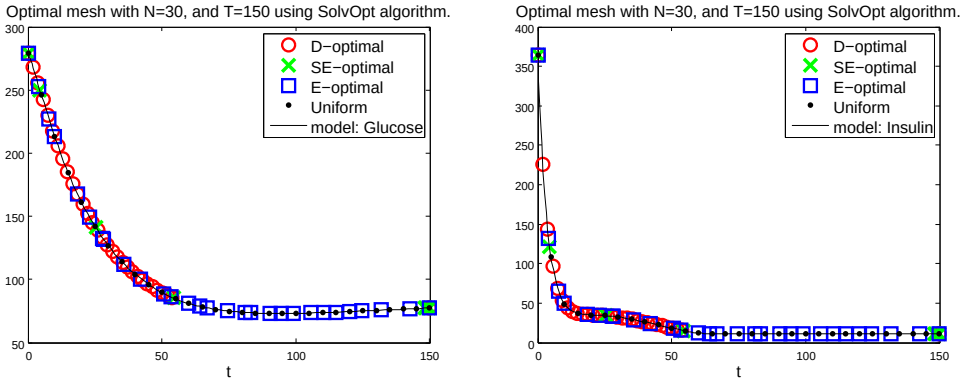
Figure 5.14: Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (p_1, p_2, p_3, p_4)$, with $T = 150$ for $N = 30$. Optimal time points with the Glucose model in panel (a) and with the Insulin model in panel (b). Optimization, using SolvOpt, with constraint implementation $(C2)$.

5.3.3 Discussion for the Glucose Regulation Model

Comparing the optimal design methods using approximate asymptotic standard errors, we find that the optimal design methods that are best for (p_1, p_2, p_3) are different than the ones best for the standard error of p_4. For constraint implementation $(C1)$ (Table 5.16), SE-optimal followed by E-optimal had the smallest standard errors for (p_1, p_2, p_3), and D-optimal followed by SE-optimal had the smallest standard errors for p_4. For constraint implementation $(C2)$ (Table 5.17), the smallest standard errors were from E-optimal followed by SE-optimal for (p_1, p_2, p_3), and for p_4 it was D-optimal followed by SE-optimal. For constraint implementations

Table 5.17: Approximate asymptotic standard errors from the asymptotic theory (4.5) resulting from different optimal design methods for $\theta_0 = (p_1, p_2, p_3, p_4)$, optimization, using SolvOpt, with constraint implementation $(C2)$.

Method	$SE(p_1)$	$SE(p_2)$	$SE(p_3)$	$SE(p_4)$
SE-optimal	4.019×10^{-3}	6.451×10^{-3}	3.088×10^{-6}	3.452×10^{-4}
D-optimal	8.322×10^{-3}	1.103×10^{-2}	6.230×10^{-6}	2.748×10^{-4}
E-optimal	3.882×10^{-3}	6.284×10^{-3}	3.063×10^{-6}	5.390×10^{-4}



(a) (b)

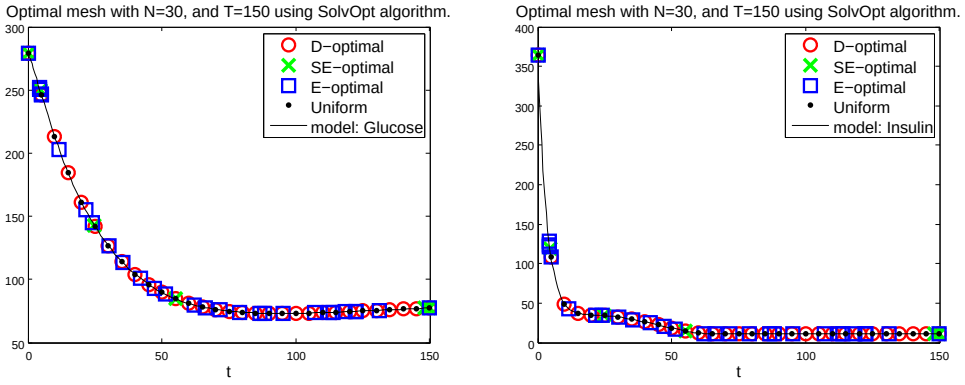
Figure 5.15: Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (p_1, p_2, p_3, p_4)$, with $T = 150$ for $N = 30$. Optimal time points with the Glucose model in panel (a) and with the Insulin model in panel (b). Optimization, using SolvOpt, with constraint implementation $(C3)$.

$(C3)$ and $(C4)$ (Tables 5.18 and 5.19), SE-optimal followed by E-optimal had the smallest standard errors for (p_1, p_2, p_3), and D-optimal followed by E-optimal had the smallest standard errors for p_4.

In conclusion, D-optimal tended to have the smallest standard errors for p_4, whereas SE-optimal or E-optimal had the smallest standard errors for (p_1, p_2, p_3). In the next section we compute the estimated standard errors from simulated data using asymptotic theory and bootstrapping as a different method of comparing the optimal design methods.

Table 5.18: Approximate asymptotic standard errors from the asymptotic theory (4.5) resulting from different optimal design methods for $\theta_0 = (p_1, p_2, p_3, p_4)$, optimization, using SolvOpt, with constraint implementation $(C3)$.

Method	$SE(p_1)$	$SE(p_2)$	$SE(p_3)$	$SE(p_4)$
SE-optimal	4.205×10^{-3}	6.535×10^{-3}	3.151×10^{-6}	3.041×10^{-4}
D-optimal	7.434×10^{-3}	1.517×10^{-2}	6.171×10^{-6}	1.181×10^{-4}
E-optimal	7.528×10^{-3}	1.123×10^{-2}	5.509×10^{-6}	1.833×10^{-4}



(a) (b)

Figure 5.16: Plot of model with optimal time points resulting from different optimal design methods for $\theta_0 = (p_1, p_2, p_3, p_4)$, with $T = 150$ for $N = 30$. Optimal time points with the Glucose model in panel (a) and with the Insulin model in panel (b). Optimization, using SolvOpt, with constraint implementation $(C4)$.

5.3.4 Result for the Glucose Regulation Model with the Inverse Problem

As in the harmonic oscillator example, we use the inverse problem with the OLS formulation to obtain parameter estimates and standard errors from both asymptotic theory (4.6) and the bootstrapping method (4.9) (see Sections 4.2.1 and 4.2.2). We create simulated noisy data corresponding to the optimal time meshes (presented in the previous section) in agreement with our statistical model (absolute error, with independent error processes for G and I) assuming true values θ_0 to be the parameter values found in Table 5.15 and *iid* noise with $\vec{\mathcal{E}}_j \sim \mathcal{N}(0, \vec{\sigma}_0^2)$. We assume the true variances: $\sigma_{0,G}^2 = 25$ and $\sigma_{0,I}^2 = 4$. In this section we only estimate a

Table 5.19: Approximate asymptotic standard errors from the asymptotic theory (4.5) resulting from different optimal design methods for $\theta_0 = (p_1, p_2, p_3, p_4)$, optimization, using SolvOpt, with constraint implementation $(C4)$.

Method	$SE(p_1)$	$SE(p_2)$	$SE(p_3)$	$SE(p_4)$
SE-optimal	4.921×10^{-3}	6.995×10^{-3}	3.633×10^{-6}	4.796×10^{-4}
D-optimal	8.767×10^{-3}	1.249×10^{-2}	6.405×10^{-6}	1.965×10^{-4}
E-optimal	7.154×10^{-3}	1.020×10^{-2}	5.253×10^{-6}	2.302×10^{-4}

subset of the parameters $\theta = (p_1, p_2, p_3, p_4)$. In addition to the estimates and standard errors, we also report the estimated covariance between estimated parameters according to asymptotic theory (4.6) and bootstrapping (4.9). For comparison purposes we also present these results for a uniform grid using the same $T = 150$ and $N = 30$.

The optimal time points for each of the three optimal design methods are the same as computed in the previous results section, and are plotted with the model in Figs. 5.13-5.16 for the four different constraints. The parameter estimates, standard errors and covariances are estimated from the asymptotic theory (4.6) corresponding to these optimal meshes are given in Tables 5.20, 5.22, 5.24, and 5.26, respectively for the four different constraints. The parameter estimates, standard errors, and covariance between parameters are estimated from the bootstrapping method (4.9) corresponding to these optimal meshes are given in Tables 5.21, 5.23, 5.25, and 5.27, respectively for the four different constraints. In each of the tables are also results on the uniform grid of time points.

5.3.5 Discussion for the Glucose Regulation Model with the Inverse Problem

Comparing the resulting parameter estimates from simulated data on the different optimal meshes to the true parameter values, $\theta_0 = (p_1, p_2, p_3, p_4) = (2.6 \times 10^{-2}, 2.5 \times 10^{-2}, 1.25 \times 10^{-5}, 4.1 \times 10^{-3})$, we find there is no optimal design method that is always favorable. Using either asymptotic theory or bootstrapping to compute parameter estimates for different optimal design methods and different constraints, we examine how close the parameter estimates are to the true values. Often (but not always) these parameter estimates from the different optimal meshes are the same order of magnitude as the true values.

The results for the uniform mesh are given for comparison. In most cases, the optimal design methods produce closer parameter estimates with smaller standard errors and covariances (as estimated by asymptotic theory and bootstrapping) than the uniform mesh.

Table 5.20: Estimates, standard errors, and covariances between parameters from the asymptotic theory (4.6) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (p_1, p_2, p_3, p_4) = (2.6 \times 10^{-2}, 2.5 \times 10^{-2}, 1.25 \times 10^{-5}, 4.1 \times 10^{-3})$ and $N = 30$, optimization, using fmincon, with constraint implementation $(C1)$.

	SE-optimal	D-optimal	E-optimal	Uniform
$\widehat{p}_1$	3.036×10^{-2}	2.303×10^{-2}	2.267×10^{-2}	2.045×10^{-2}
$\widehat{SE}(\widehat{p}_1)$	4.977×10^{-3}	1.138×10^{-2}	4.634×10^{-3}	1.056×10^{-2}
$\widehat{p}_2$	2.180×10^{-2}	2.723×10^{-2}	2.818×10^{-2}	3.607×10^{-2}
$\widehat{SE}(\widehat{p}_2)$	7.657×10^{-3}	1.660×10^{-2}	6.818×10^{-3}	1.536×10^{-2}
$\widehat{p}_3$	9.213×10^{-6}	1.414×10^{-5}	1.565×10^{-5}	1.766×10^{-5}
$\widehat{SE}(\widehat{p}_3)$	3.946×10^{-6}	8.421×10^{-6}	3.732×10^{-6}	7.787×10^{-6}
$\widehat{p}_4$	3.544×10^{-3}	4.174×10^{-3}	4.238×10^{-3}	4.027×10^{-3}
$\widehat{SE}(\widehat{p}_4)$	7.822×10^{-4}	4.510×10^{-4}	1.140×10^{-3}	4.817×10^{-4}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_2)$	-3.377×10^{-5}	-1.846×10^{-4}	-2.779×10^{-5}	-1.579×10^{-4}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_3)$	-1.873×10^{-8}	-9.520×10^{-8}	-1.630×10^{-8}	-8.160×10^{-8}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_4)$	5.458×10^{-7}	8.794×10^{-7}	1.117×10^{-6}	8.615×10^{-7}
$\widehat{\mathrm{Cov}}(\widehat{p}_2, \widehat{p}_3)$	2.815×10^{-8}	1.383×10^{-7}	2.289×10^{-8}	1.181×10^{-7}
$\widehat{\mathrm{Cov}}(\widehat{p}_2, \widehat{p}_4)$	2.851×10^{-7}	-6.379×10^{-7}	2.679×10^{-7}	-5.341×10^{-7}
$\widehat{\mathrm{Cov}}(\widehat{p}_3, \widehat{p}_4)$	-5.0551×10^{-10}	-5.785×10^{-10}	-1.308×10^{-9}	-5.605×10^{-10}

Asymptotic theory: parameter estimates.

For the constraint implementation $(C1)$ using asymptotic theory (Table 5.20), the estimates for p_1, p_2, p_3, and p_4 are closest to the true values for D-optimal followed by E-optimal. In other constraint implementations, which optimal sampling distribution produced estimates closest to the true values was different depending on the parameter.

For constraint implementation $(C2)$ (Table 5.22), parameters estimates of (p_1, p_2, p_3) were closest to the true values for the optimal sampling distributions from D-optimal followed by SE-optimal. For p_4 the closest parameter estimates were from the uniform mesh, followed by D-optimal.

For constraint implementation $(C3)$ (Table 5.24), the closest parameter estimate for p_1 came from D-optimal followed by SE-optimal. For (p_2, p_3, p_4) the closest estimates came from E-optimal followed by SE-optimal (for p_2, p_3) and D-optimal (for p_4).

For the last constraint implementation $(C4)$ (Table 5.26), D-optimal followed by SE-optimal

Table 5.21: Estimates, standard errors, and covariances between parameters from the bootstrap method (4.9) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (p_1, p_2, p_3, p_4) = (2.6 \times 10^{-2}, 2.5 \times 10^{-2}, 1.25 \times 10^{-5}, 4.1 \times 10^{-3})$, $M = 1000$ bootstraps and $N = 30$, optimization, using fmincon, with constraint implementation $(C1)$.

	SE-optimal	D-optimal	E-optimal	Uniform
$\widehat{p}_1$	2.908×10^{-2}	2.220×10^{-2}	2.073×10^{-2}	1.973×10^{-2}
$\widehat{SE}(\widehat{p}_1)$	6.215×10^{-3}	8.052×10^{-3}	5.708×10^{-3}	8.563×10^{-3}
$\widehat{p}_2$	2.486×10^{-2}	2.855×10^{-2}	3.126×10^{-2}	3.730×10^{-2}
$\widehat{SE}(\widehat{p}_2)$	1.179×10^{-2}	1.169×10^{-2}	9.810×10^{-3}	1.279×10^{-2}
$\widehat{p}_3$	1.075×10^{-5}	1.552×10^{-5}	1.864×10^{-5}	1.916×10^{-5}
$\widehat{SE}(\widehat{p}_3)$	6.541×10^{-6}	6.570×10^{-6}	6.302×10^{-6}	8.017×10^{-6}
$\widehat{p}_4$	3.688×10^{-3}	4.215×10^{-3}	3.809×10^{-3}	3.984×10^{-3}
$\widehat{SE}(\widehat{p}_4)$	3.743×10^{-4}	1.855×10^{-4}	6.223×10^{-4}	2.098×10^{-4}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_2)$	-6.799×10^{-5}	-9.116×10^{-5}	-5.323×10^{-5}	-1.053×10^{-4}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_3)$	-3.868×10^{-8}	-5.198×10^{-8}	-3.479×10^{-8}	-6.722×10^{-8}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_4)$	-3.337×10^{-7}	2.268×10^{-7}	1.075×10^{-7}	5.716×10^{-8}
$\widehat{\mathrm{Cov}}(\widehat{p}_2, \widehat{p}_3)$	7.452×10^{-8}	7.529×10^{-8}	6.005×10^{-8}	9.990×10^{-8}
$\widehat{\mathrm{Cov}}(\widehat{p}_2, \widehat{p}_4)$	1.050×10^{-6}	-1.262×10^{-7}	7.158×10^{-7}	2.310×10^{-7}
$\widehat{\mathrm{Cov}}(\widehat{p}_3, \widehat{p}_4)$	5.432×10^{-10}	-8.735×10^{-11}	-1.465×10^{-10}	8.308×10^{-11}

had parameter estimates closest to the true values for parameters (p_1, p_2, p_3). E-optimal followed by SE-optimal had the closest estimate of p_4.

Asymptotic theory: standard errors.

Here we compare the optimal design methods based on which has the smallest standard error estimates. Again, the results are dependent on the parameter and the constraint implementation.

For the first constraint implementation $(C1)$ (Table 5.20), the smallest standard errors estimated using asymptotic theory are from D-optimal followed by SE-optimal for parameters (p_1, p_2, p_3), and followed by E-optimal for p_4.

For the constraint implementation $(C2)$ (Table 5.22), the smallest standard error for parameters (p_1, p_2, p_3) come from SE-optimal followed by E-optimal. For p_4, the smallest standard error estimates are from the uniform mesh followed by D-optimal.

For the constraint implementation $(C3)$ (Table 5.24), the standard error estimates for

Table 5.22: Estimates, standard errors, and covariances between parameters from the asymptotic theory (4.6) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (p_1, p_2, p_3, p_4) = (2.6 \times 10^{-2}, 2.5 \times 10^{-2}, 1.25 \times 10^{-5}, 4.1 \times 10^{-3})$ and $N = 30$, optimization, using fmincon, with constraint implementation $(C2)$.

	SE-optimal	D-optimal	E-optimal	Uniform
$\widehat{p}_1$	2.118×10^{-2}	2.232×10^{-2}	2.116×10^{-2}	2.045×10^{-2}
$\widehat{SE}(\widehat{p}_1)$	5.063×10^{-3}	8.596×10^{-3}	5.298×10^{-3}	1.056×10^{-2}
$\widehat{p}_2$	3.509×10^{-2}	3.337×10^{-2}	4.356×10^{-2}	3.607×10^{-2}
$\widehat{SE}(\widehat{p}_2)$	8.020×10^{-3}	1.139×10^{-2}	8.465×10^{-3}	1.536×10^{-2}
$\widehat{p}_3$	1.772×10^{-5}	1.628×10^{-5}	1.958×10^{-5}	1.766×10^{-5}
$\widehat{SE}(\widehat{p}_3)$	4.247×10^{-6}	6.573×10^{-6}	4.874×10^{-6}	7.787×10^{-6}
$\widehat{p}_4$	4.486×10^{-3}	3.993×10^{-3}	4.249×10^{-3}	4.027×10^{-3}
$\widehat{SE}(\widehat{p}_4)$	9.537×10^{-4}	5.919×10^{-4}	1.607×10^{-3}	4.817×10^{-4}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_2)$	-3.569×10^{-5}	-9.416×10^{-5}	-3.811×10^{-5}	-1.579×10^{-4}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_3)$	-2.036×10^{-8}	-5.566×10^{-8}	-2.376×10^{-8}	-8.160×10^{-8}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_4)$	6.620×10^{-7}	1.227×10^{-7}	1.774×10^{-6}	8.615×10^{-7}
$\widehat{\mathrm{Cov}}(\widehat{p}_2, \widehat{p}_3)$	3.131×10^{-8}	7.280×10^{-8}	3.585×10^{-8}	1.181×10^{-7}
$\widehat{\mathrm{Cov}}(\widehat{p}_2, \widehat{p}_4)$	4.626×10^{-7}	4.238×10^{-7}	9.670×10^{-7}	-5.341×10^{-7}
$\widehat{\mathrm{Cov}}(\widehat{p}_3, \widehat{p}_4)$	-6.824×10^{-10}	9.532×10^{-13}	-2.353×10^{-9}	-5.605×10^{-10}

parameters (p_1, p_2, p_3) are smallest using the mesh from SE-optimal, followed by D-optimal (for p_1) and E-optimal (for p_2, p_3). For parameter p_4, the smallest standard error is from D-optimal followed by E-optimal.

For the last constraint implementation $(C4)$ (Table 5.26), E-optimal has the smallest standard errors for parameters (p_1, p_3, p_4), followed by SE-optimal (for p_1, p_3) and D-optimal (for p_4). For p_2, the smallest standard errors are from SE-optimal followed by E-optimal.

Asymptotic theory: covariance estimates.

We also compare the optimal design methods based on which has the smallest covariance estimates in absolute value.

For the first constraint implementation (C_1) (Table 5.20), the smallest in absolute value covariance estimates come from either SE-optimal or E-optimal for different pairs of parameters.

For constraint implementation (C_2) (Table 5.22), SE-optimal or D-optimal have the smallest

Table 5.23: Estimates, standard errors, and covariances between parameters from the bootstrap method (4.9) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (p_1, p_2, p_3, p_4) = (2.6 \times 10^{-2}, 2.5 \times 10^{-2}, 1.25 \times 10^{-5}, 4.1 \times 10^{-3})$, $M = 1000$ bootstraps and $N = 30$, optimization, using fmincon, with constraint implementation $(C2)$.

	SE-optimal	D-optimal	E-optimal	Uniform
$\widehat{p}_1$	1.874×10^{-2}	1.883×10^{-2}	2.104×10^{-2}	1.973×10^{-2}
$\widehat{SE}(\widehat{p}_1)$	6.619×10^{-3}	8.291×10^{-3}	6.397×10^{-3}	8.563×10^{-3}
$\widehat{p}_2$	4.034×10^{-2}	4.249×10^{-2}	4.337×10^{-2}	3.730×10^{-2}
$\widehat{SE}(\widehat{p}_2)$	1.305×10^{-2}	1.458×10^{-2}	1.409×10^{-2}	1.279×10^{-2}
$\widehat{p}_3$	2.124×10^{-5}	2.069×10^{-5}	2.075×10^{-5}	1.916×10^{-5}
$\widehat{SE}(\widehat{p}_3)$	8.241×10^{-6}	8.799×10^{-6}	7.733×10^{-6}	8.017×10^{-6}
$\widehat{p}_4$	4.341×10^{-3}	3.920×10^{-3}	3.988×10^{-3}	3.984×10^{-3}
$\widehat{SE}(\widehat{p}_4)$	4.228×10^{-4}	3.107×10^{-4}	6.192×10^{-4}	2.098×10^{-4}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_2)$	-8.157×10^{-5}	-1.149×10^{-4}	-7.952×10^{-5}	-1.053×10^{-4}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_3)$	-5.272×10^{-8}	-7.128×10^{-8}	-4.687×10^{-8}	-6.722×10^{-8}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_4)$	1.240×10^{-8}	1.275×10^{-7}	-5.657×10^{-7}	5.716×10^{-8}
$\widehat{\mathrm{Cov}}(\widehat{p}_2, \widehat{p}_3)$	1.048×10^{-7}	1.249×10^{-7}	1.042×10^{-7}	9.990×10^{-8}
$\widehat{\mathrm{Cov}}(\widehat{p}_2, \widehat{p}_4)$	4.220×10^{-7}	8.311×10^{-8}	2.226×10^{-6}	2.310×10^{-7}
$\widehat{\mathrm{Cov}}(\widehat{p}_3, \widehat{p}_4)$	6.133×10^{-11}	-2.764×10^{-11}	9.390×10^{-10}	8.308×10^{-11}

in absolute value covariance estimates.

For constraint implementation (C_3) (Table 5.24), SE-optimal or D-optimal have the smallest in absolute value covariance estimates, except for $\hat{\mathrm{C}}\mathrm{ov}(\widehat{p}_2, \widehat{p}_4)$ where E-optimal is the smallest.

For constraint implementation (C_4) (Table 5.26), E-optimal has the smallest in absolute value covariance estimates, except for $\hat{\mathrm{C}}\mathrm{ov}(\widehat{p}_1, \widehat{p}_2)$ where SE-optimal is the smallest.

Bootstrapping: parameter estimates.

Here we compare the optimal design methods based on which had bootstrapping parameter estimates closest to the true values. Often these results are different for the different parameters, as well as the constraint implementation.

For the first constraint implementation (C_1) (Table 5.21), bootstrapping parameter estimates for (p_1, p_2, p_3) were closest to the true values for SE-optimal followed by D-optimal. For p_4, the closest parameter estimates came from D-optimal and then E-optimal.

Table 5.24: Estimates, standard errors, and covariances between parameters from the asymptotic theory (4.6) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (p_1, p_2, p_3, p_4) = (2.6 \times 10^{-2}, 2.5 \times 10^{-2}, 1.25 \times 10^{-5}, 4.1 \times 10^{-3})$ and $N = 30$, optimization, using fmincon, with constraint implementation $(C3)$.

	SE-optimal	D-optimal	E-optimal	Uniform
$\widehat{p}_1$	2.960×10^{-2}	2.851×10^{-2}	2.970×10^{-2}	2.045×10^{-2}
$\widehat{SE}(\widehat{p}_1)$	5.210×10^{-3}	8.937×10^{-3}	9.018×10^{-3}	1.056×10^{-2}
$\widehat{p}_2$	1.894×10^{-2}	1.303×10^{-2}	1.951×10^{-2}	3.607×10^{-2}
$\widehat{SE}(\widehat{p}_2)$	7.987×10^{-3}	1.853×10^{-2}	1.338×10^{-2}	1.536×10^{-2}
$\widehat{p}_3$	9.558×10^{-6}	8.981×10^{-6}	1.018×10^{-5}	1.766×10^{-5}
$\widehat{SE}(\widehat{p}_3)$	4.201×10^{-6}	7.459×10^{-6}	6.701×10^{-6}	7.787×10^{-6}
$\widehat{p}_4$	3.915×10^{-3}	3.945×10^{-3}	4.166×10^{-3}	4.027×10^{-3}
$\widehat{SE}(\widehat{p}_4)$	8.373×10^{-4}	2.882×10^{-4}	4.366×10^{-4}	4.817×10^{-4}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_2)$	-3.688×10^{-5}	-1.563×10^{-4}	-1.168×10^{-4}	-1.579×10^{-4}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_3)$	-2.081×10^{-8}	-6.595×10^{-8}	-5.988×10^{-8}	-8.160×10^{-8}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_4)$	5.971×10^{-7}	1.045×10^{-8}	5.570×10^{-7}	8.615×10^{-7}
$\widehat{\mathrm{Cov}}(\widehat{p}_2, \widehat{p}_3)$	3.113×10^{-8}	1.353×10^{-7}	8.834×10^{-8}	1.181×10^{-7}
$\widehat{\mathrm{Cov}}(\widehat{p}_2, \widehat{p}_4)$	3.520×10^{-7}	4.275×10^{-7}	-2.100×10^{-7}	-5.341×10^{-7}
$\widehat{\mathrm{Cov}}(\widehat{p}_3, \widehat{p}_4)$	-5.579×10^{-10}	5.546×10^{-11}	-3.461×10^{-10}	-5.605×10^{-10}

For constraint implementation $(C2)$ (Table 5.23), parameter estimates for p_1 the closest parameter estimates came from E-optimal followed by D-optimal. For p_2, the closest parameter estimates came from the uniform mesh followed by E-optimal. For p_3, the uniform mesh then SE-optimal had the closest parameter estimates to the true value. For p_4 the closest estimate came from D-optimal followed by E-optimal.

For constraint implementation $(C3)$ (Table 5.25), the closest estimates for p_1 came from D-optimal followed by E-optimal. For (p_2, p_3, p_4) the closest estimates came from E-optimal followed by SE-optimal (for p_2, p_3) and D-optimal (for p_4).

For the last constraint $(C4)$ (Table 5.27), the closest estimates for (p_1, p_2, p_3) came from D-optimal and then SE-optimal. For p_4 the closest estimate to the true value came from SE-optimal followed by E-optimal.

None of the optimal design methods are consistent with parameter estimates that are the closest to the true values for all cases.

Table 5.25: Estimates, standard errors, and covariances between parameters from the bootstrap method (4.9) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (p_1, p_2, p_3, p_4) = (2.6 \times 10^{-2}, 2.5 \times 10^{-2}, 1.25 \times 10^{-5}, 4.1 \times 10^{-3})$, $M = 1000$ bootstraps and $N = 30$, optimization, using fmincon, with constraint implementation $(C3)$.

	SE-optimal	D-optimal	E-optimal	Uniform
$\widehat{p}_1$	2.812×10^{-2}	2.738×10^{-2}	2.749×10^{-2}	1.973×10^{-2}
$\widehat{SE}(\widehat{p}_1)$	5.495×10^{-3}	6.344×10^{-3}	7.838×10^{-3}	8.563×10^{-3}
$\widehat{p}_2$	2.160×10^{-2}	1.513×10^{-2}	2.300×10^{-2}	3.730×10^{-2}
$\widehat{SE}(\widehat{p}_2)$	9.465×10^{-3}	1.166×10^{-2}	1.183×10^{-2}	1.279×10^{-2}
$\widehat{p}_3$	1.122×10^{-5}	1.030×10^{-5}	1.240×10^{-5}	1.916×10^{-5}
$\widehat{SE}(\widehat{p}_3)$	5.360×10^{-6}	5.291×10^{-6}	6.429×10^{-6}	8.017×10^{-6}
$\widehat{p}_4$	3.695×10^{-3}	4.011×10^{-3}	4.188×10^{-3}	3.984×10^{-3}
$\widehat{SE}(\widehat{p}_4)$	3.244×10^{-4}	1.311×10^{-4}	1.946×10^{-4}	2.098×10^{-4}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_2)$	-4.787×10^{-5}	-6.858×10^{-5}	-8.971×10^{-5}	-1.053×10^{-4}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_3)$	-2.768×10^{-8}	-3.288×10^{-8}	-4.933×10^{-8}	-6.722×10^{-8}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_4)$	-1.523×10^{-7}	2.531×10^{-8}	9.876×10^{-8}	5.716×10^{-8}
$\widehat{\mathrm{Cov}}(\widehat{p}_2, \widehat{p}_3)$	4.913×10^{-8}	5.882×10^{-8}	7.443×10^{-8}	9.990×10^{-8}
$\widehat{\mathrm{Cov}}(\widehat{p}_2, \widehat{p}_4)$	6.607×10^{-7}	9.576×10^{-8}	9.573×10^{-8}	2.310×10^{-7}
$\widehat{\mathrm{Cov}}(\widehat{p}_3, \widehat{p}_4)$	3.331×10^{-10}	3.083×10^{-11}	4.247×10^{-11}	8.308×10^{-11}

Bootstrapping: standard errors.

We compare the optimal design methods based on how small their standard errors are as estimated by the bootstrap method.

Comparing the standard error estimates from the first constraint implementation $(C1)$ (Table 5.21) we find that for parameters (p_1, p_2, p_3) the optimal mesh from E-optimal has the smallest standard errors followed by SE-optimal. For p_4, the smallest standard errors come from D-optimal followed by E-optimal.

For the second constraint implementation $(C2)$ (Table 5.23), the smallest standard errors for parameters (p_1, p_3) are from E-optimal followed by SE-optimal. For p_2, the uniform mesh has the smallest standard errors, followed by SE-optimal. For p_4, the uniform mesh has the smallest standard errors followed by the optimal mesh from D-optimal.

For the constraint implementation $(C3)$ (Table 5.25), the smallest standard errors come

Table 5.26: Estimates, standard errors, and covariances between parameters from the asymptotic theory (4.6) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (p_1, p_2, p_3, p_4) = (2.6 \times 10^{-2}, 2.5 \times 10^{-2}, 1.25 \times 10^{-5}, 4.1 \times 10^{-3})$ and $N = 30$, optimization, using fmincon, with constraint implementation $(C4)$.

	SE-optimal	D-optimal	E-optimal	Uniform
$\widehat{p}_1$	3.285×10^{-2}	2.540×10^{-2}	3.401×10^{-2}	2.045×10^{-2}
$\widehat{SE}(\widehat{p}_1)$	6.396×10^{-3}	1.102×10^{-2}	6.353×10^{-3}	1.056×10^{-2}
$\widehat{p}_2$	1.783×10^{-2}	2.452×10^{-2}	1.079×10^{-2}	3.607×10^{-2}
$\widehat{SE}(\widehat{p}_2)$	8.983×10^{-3}	1.562×10^{-2}	9.052×10^{-3}	1.536×10^{-2}
$\widehat{p}_3$	9.094×10^{-6}	1.226×10^{-5}	6.025×10^{-6}	1.766×10^{-5}
$\widehat{SE}(\widehat{p}_3)$	5.375×10^{-6}	8.277×10^{-6}	4.850×10^{-6}	7.787×10^{-6}
$\widehat{p}_4$	3.980×10^{-3}	3.958×10^{-3}	4.040×10^{-3}	4.027×10^{-3}
$\widehat{SE}(\widehat{p}_4)$	1.362×10^{-3}	4.896×10^{-4}	4.257×10^{-4}	4.817×10^{-4}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_2)$	-5.251×10^{-5}	-1.692×10^{-4}	-5.587×10^{-5}	-1.579×10^{-4}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_3)$	-3.260×10^{-8}	-9.060×10^{-8}	-3.043×10^{-8}	-8.160×10^{-8}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_4)$	2.000×10^{-6}	1.1481×10^{-6}	4.352×10^{-7}	8.615×10^{-7}
$\widehat{\mathrm{Cov}}(\widehat{p}_2, \widehat{p}_3)$	4.436×10^{-8}	1.281×10^{-7}	4.316×10^{-8}	1.181×10^{-7}
$\widehat{\mathrm{Cov}}(\widehat{p}_2, \widehat{p}_4)$	-2.356×10^{-7}	-1.016×10^{-6}	-2.022×10^{-7}	-5.341×10^{-7}
$\widehat{\mathrm{Cov}}(\widehat{p}_3, \widehat{p}_4)$	-2.203×10^{-9}	-8.152×10^{-10}	-3.102×10^{-10}	-5.605×10^{-10}

from SE-optimal for parameters (p_1, p_2) followed by D-optimal. For parameters (p_3, p_4) the optimal mesh from D-optimal has the smallest standard errors, followed by SE-optimal (for p_3) and E-optimal (for p_4).

For the last constraint implementation $(C4)$ (Table 5.27), the smallest standard errors for parameters (p_1, p_2, p_3) are from E-optimal followed by SE-optimal. For parameter p_4, the smallest standard errors are from E-optimal followed by D-optimal.

Bootstrapping: covariance estimates.

For the first constraint implementation $(C1)$ (Table 5.21), the smallest in absolute value covariance estimates as estimated by the bootstrapping method came from the optimal meshes of D-optimal (for $\hat{\mathrm{Cov}}(\widehat{p}_2, \widehat{p}_4)$) or E-optimal (for $\hat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_2)$, $\hat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_3)$, and $\hat{\mathrm{Cov}}(\widehat{p}_2, \widehat{p}_3)$) or the uniform mesh (for $\hat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_4)$ and $\hat{\mathrm{Cov}}(\widehat{p}_3, \widehat{p}_4)$).

For constraint implementation $(C2)$ (Table 5.23), D-optimal or E-optimal have the smallest

Table 5.27: Estimates, standard errors, and covariances between parameters from the bootstrap method (4.9) resulting from different optimal design methods (as well as for the uniform mesh) for $\theta_0 = (p_1, p_2, p_3, p_4) = (2.6 \times 10^{-2}, 2.5 \times 10^{-2}, 1.25 \times 10^{-5}, 4.1 \times 10^{-3})$, $M = 1000$ bootstraps and $N = 30$, optimization, using fmincon, with constraint implementation $(C4)$.

	SE-optimal	D-optimal	E-optimal	Uniform
$\widehat{p}_1$	3.243×10^{-2}	2.409×10^{-2}	3.304×10^{-2}	1.973×10^{-2}
$\widehat{SE}(\widehat{p}_1)$	4.851×10^{-3}	7.282×10^{-3}	3.410×10^{-3}	8.563×10^{-3}
$\widehat{p}_2$	1.899×10^{-2}	2.659×10^{-2}	1.262×10^{-2}	3.730×10^{-2}
$\widehat{SE}(\widehat{p}_2)$	6.864×10^{-3}	1.056×10^{-2}	5.304×10^{-3}	1.279×10^{-2}
$\widehat{p}_3$	9.688×10^{-6}	1.385×10^{-5}	6.571×10^{-6}	1.916×10^{-5}
$\widehat{SE}(\widehat{p}_3)$	3.843×10^{-6}	5.780×10^{-6}	2.220×10^{-6}	8.017×10^{-6}
$\widehat{p}_4$	4.117×10^{-3}	3.967×10^{-3}	4.025×10^{-3}	3.984×10^{-3}
$\widehat{SE}(\widehat{p}_4)$	7.050×10^{-4}	2.147×10^{-4}	1.892×10^{-4}	2.098×10^{-4}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_2)$	-2.970×10^{-5}	-7.462×10^{-5}	-1.700×10^{-5}	-1.053×10^{-4}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_3)$	-1.680×10^{-8}	-4.139×10^{-8}	-7.230×10^{-9}	-6.722×10^{-8}
$\widehat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_4)$	2.879×10^{-8}	6.162×10^{-8}	9.816×10^{-8}	5.716×10^{-8}
$\widehat{\mathrm{Cov}}(\widehat{p}_2, \widehat{p}_3)$	2.477×10^{-8}	5.986×10^{-8}	1.133×10^{-8}	9.990×10^{-8}
$\widehat{\mathrm{Cov}}(\widehat{p}_2, \widehat{p}_4)$	6.321×10^{-7}	1.286×10^{-7}	2.627×10^{-9}	2.310×10^{-7}
$\widehat{\mathrm{Cov}}(\widehat{p}_3, \widehat{p}_4)$	4.890×10^{-11}	2.363×10^{-11}	-1.770×10^{-11}	8.308×10^{-11}

in absolute value covariance estimates, except for $\hat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_4)$ where SE-optimal is the smallest and $\hat{\mathrm{Cov}}(\widehat{p}_2, \widehat{p}_3)$ where the uniform mesh is the smallest.

For the third constraint implementation $(C3)$ (Table 5.25), D-optimal or SE-optimal have the smallest in absolute value covariance estimates, except for $\hat{\mathrm{Cov}}(\widehat{p}_2, \widehat{p}_4)$ where E-optimal is the smallest.

For the last constraint implementation $(C4)$ (Table 5.27), the smallest in absolute value covariance estimates are from E-optimal, except for $\hat{\mathrm{Cov}}(\widehat{p}_1, \widehat{p}_4)$ where SE-optimal is the smallest.

Comparing the optimal design methods based on the bootstrapping covariance estimates, we find there is not one method that is always favorable.

5.4 Conclusions

We compared D-optimal, E-optimal and SE-optimal design methods for a simple differential equation model: the logistic population model, a second order differential equation: the harmonic oscillator model, and a vector system for glucose regulation. D-optimal and E-optimal design methods are more established in the literature. Our comparisons test the performance of SE-optimal design, which is a relatively newer method.

For the logistic example, the optimal design methods were compared using the Monte Carlo method for asymptotic standard errors. Comparing the average and median parameter estimates to their true values, we find that SE-optimal has closest parameter estimates for $N = 10$ time points. For $N = 15$, no method had estimates that were always closest to the true values. In all cases each optimal design methods produced estimates close to the true values. The average and median standard errors for K were smallest from the optimal mesh from E-optimal. For parameters r and x_0, SE-optimal had the smallest average and median standard errors. Overall, no optimal design method is consistently favorable for this logistic example.

For the harmonic example, comparing the approximate asymptotic standard errors, we found that different optimal design methods were favorable for different parameters. D-optimal often had the smallest standard errors for K and x_1. SE-optimal often had the smallest standard errors for C. For x_2, either SE-optimal or E-optimal had the smallest standard errors. We also compared methods using the inverse problem with simulated data and asymptotic theory and bootstrapping. Comparing methods based on who's parameter estimates were closest to the true values, and who had the smallest standard errors or covariances, there was no method that was preferred over the others. In each comparison, the best optimal design method often depended on the constraint implementation, the choice of $T = 14.14$ or $T = 28.28$, and the parameter.

For the glucose regulation model, comparing the approximate asymptotic standard errors, we found that for parameters (p_1, p_2, p_3) either SE-optimal or E-optimal had the smallest standard errors. D-optimal tended to have the smallest standard errors for p_4. We also compared the optimal design methods for the inverse problem using asymptotic theory and bootstrapping. Comparing the parameter estimates to their true values, none of the optimal design methods were consistently closer. Comparing the optimal design methods based on who had the smallest standard errors and covariances we found that no method was preferable over the others. However, the optimal design methods often had smaller standard errors and covariances than the uniform mesh. The constraint implementation, parameter, and choice of asymptotic theory or bootstrapping influenced which optimal design method would be favorable for this example.

The best choice of optimal design method depends on the complexity of the model, the type of constraint one is using, the subset of parameters you are estimating, and even the choice of N and T. The examples in this comparison provide evidence that SE-optimal design is competitive

with D-optimal and E-optimal design, and in some cases SE-optimal design is a more favorable method.

Chapter 6

Monte Carlo Analysis

The results in the previous chapter (Chapter 5) may not be representative of typical behavior since often the results were given for a single realization. In this section we present a Monte Carlo analysis of the performance of the Optimal Design methods for the logistic and harmonic oscillator examples.

6.1 Monte Carlo Methodology

For a single Monte Carlo trial, we generate simulated data on the optimal mesh $\{t_j\}_{j=1}^N$ based on the true parameter values θ_0 corresponding to our statistical model

$$y_i = f(t_j, \theta_0) + \epsilon_j, j = 1, \ldots, N,$$

where the ϵ_j are realizations of $\mathcal{E}_j \sim \mathcal{N}(0, \sigma_0^2)$ for $j = 1, \ldots, N$. Parameters are estimated using OLS taking the initial parameter guess to be $\theta^0 = 1.4\theta_0$. The parameter estimates are stored, and the process is repeated with new simulated data corresponding to the optimal mesh for $M = 1000$ Monte Carlo trials. The choice of $M = 1000$ Monte Carlo trials was made since fewer trials would not reliably depict the distribution of parameter estimates. More trials increase the resolution though the results remain unchanged.

In addition, for each simulated data set we computed the estimated FIM, $\widehat{F}$, to determine if the 95% confidence ellipsoid contains the true parameter values θ_0. The 95% confidence ellipsoid corresponding to normally distributed residuals is defined by [30]

$$(\theta - \hat{\theta})^T \widehat{F} (\theta - \hat{\theta}) = p\mathrm{F}_{p,n-p}^{\alpha},$$

where $\mathrm{F}_{p,n-p}^{\alpha}$ is the F distribution with p numerator degrees of freedom, $n - p$ denominator degrees of freedom, and critical level $\alpha = 0.05$. The confidence ellipsoid will contain the true

parameter values if the following inequality holds:

$$(\theta_0 - \hat{\theta})^T \widehat{F}(\theta_0 - \hat{\theta}) \leq p\mathrm{F}^{\alpha}_{p,n-p}. \tag{6.1}$$

The Monte Carlo analysis was repeated for the three different optimal design methods as well as for the uniform mesh, each time with $M = 1000$ new sets of simulated data. For each Monte Carlo analysis, the average parameter estimate is reported along with its standard deviation. We will compare the standard deviation of these Monte Carlo trials to the standard errors as computed using asymptotic theory reported in a previous section.

In addition, we report the proportion of confidence ellipsoids containing the true value and the corresponding standard error,

$$\hat{P} = \frac{\#\text{ of 95\% Confidence Ellipsoids containing } \theta_0}{M}, \;\; SE(\hat{P}) = \sqrt{\frac{\hat{P}(1-\hat{P})}{M}}.$$

Note that $\hat{P}$ is the estimate of P_0 the true proportion. Figure 6.1 gives an example of 95% confidence ellipsoids corresponding to parameter estimates from simulated data. The true parameter values are plotted as a point in $\mathbb{R}^p = \mathbb{R}^2$. Ellipsoids that contain the true parameter are gray and those that do not are black, as determined by the inequality given in (6.1). We are interested to see if $\hat{p}$ is close to 0.95 (95% confidence ellipsoid) for the different optimal design methods and the uniform mesh.

6.2 Logistic Results

Monte Carlo analysis was performed using the same optimal design points as reported earlier in Section 5.1. As stated previously, these optimal design points (see Figs. 5.1 and 5.3) were found using constraint implementation ($C2$) and SolvOpt with $N = 10$ or $N = 15$, $\theta_0 = (K, r, x_0) = (17.5, 0.7, 0.1)$, and $T = 25$. The Monte Carlo analysis was conducted three times for each optimal design mesh taking different levels of noise in the simulated data, $\sigma_0^2 = (0.01, 0.05, 0.10)$. Tables 6.1-6.3 and 6.5-6.7 contain the average estimates and standard deviations from the Monte Carlo trials for $N = 10$ and $N = 15$ respectively and for the three different noise levels. In addition, the tables contain the asymptotic standard errors (4.5), using the true parameter values, for comparison with the standard deviations of the Monte Carlo parameter estimates. The percentage of 95% confidence ellipsoids which contain the true parameter values are given in Tables 6.4 and 6.8. Histograms of the parameter estimates are plotted in Figures 6.2-6.4 for $N = 10$ and Figures 6.5-6.7 for $N = 15$. Each figure corresponds to a specific value of σ_0^2. Subfigures represent histograms for different parameter estimates,

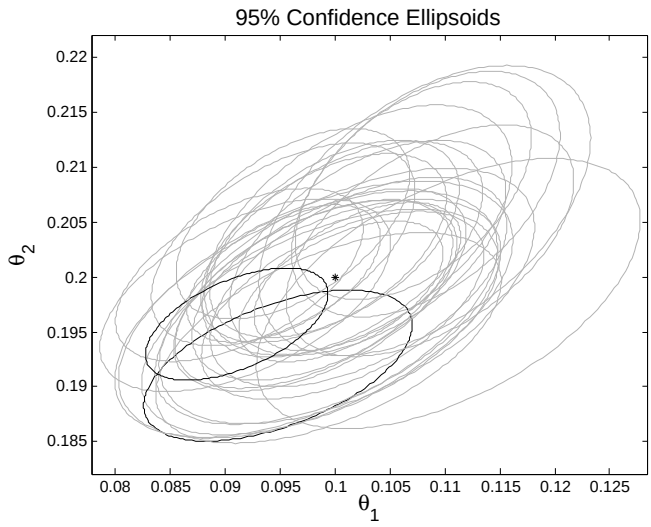


Figure 6.1: A generic example of 95% Confidence Ellipsoids (from $M = 30$ simulated data sets, estimating $p = 2$ parameters), where the true value is given by the star, confidence ellipsoids containing the true value are gray, and confidence ellipsoids not containing the true value are black.

$\hat{\theta} = (\widehat{K}, \hat{r}, \hat{x}_0)$. Within each subfigure are histograms for a specific parameter originating from an optimal design mesh or the uniform mesh.

Table 6.1: Average estimates (θ_{Avg}) with their standard deviations (θ_{SD}) from $M = 1000$ Monte Carlo trials as well as asymptotic standard errors ($SE(\theta)$). Data was simulated using true values $\theta_0 = (17.5, 0.7, 0.1)$, $N = 10$, and 1% noise ($\sigma_0^2 = 0.01$).

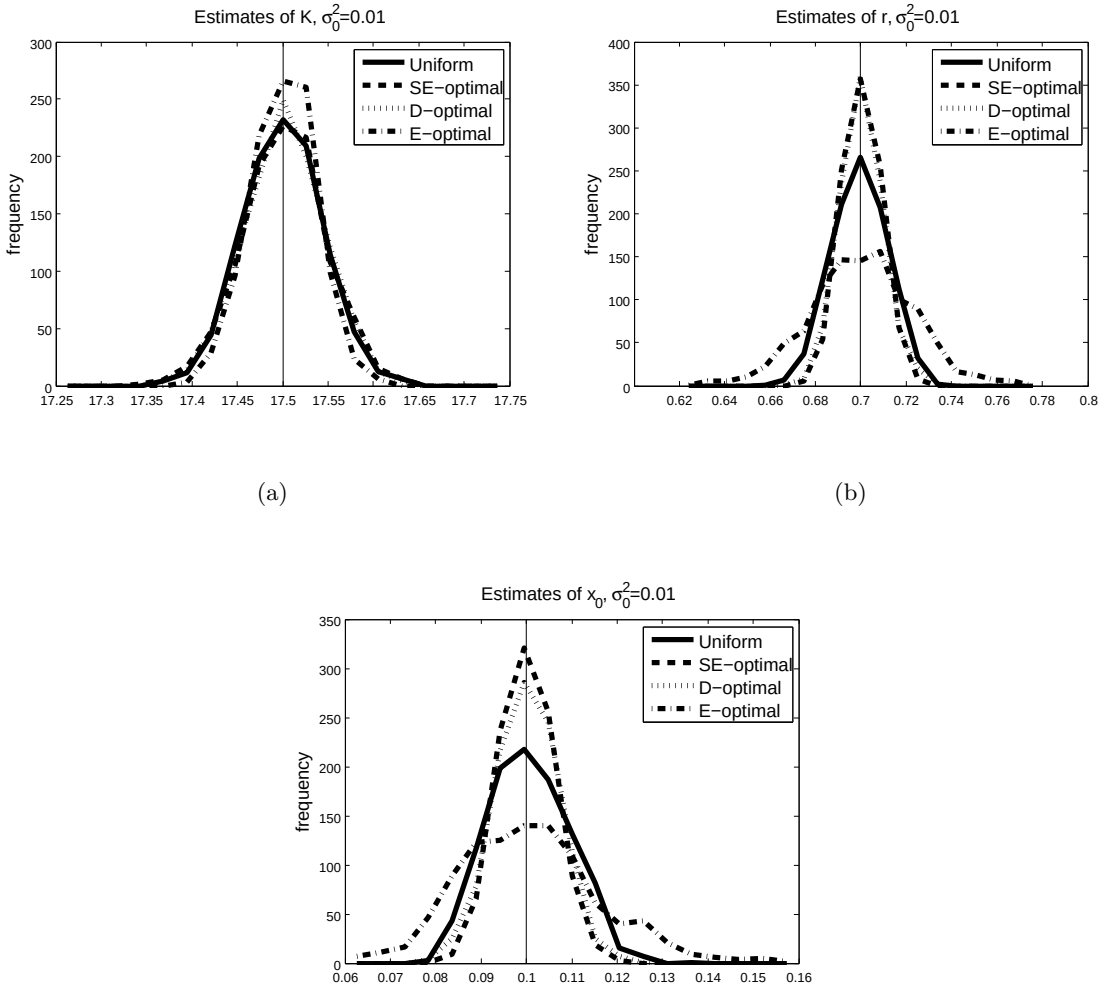
	SE-opt	D-opt	E-opt	Uniform
$\widehat{K}_{Avg}$	17.5019	17.5015	17.5013	17.5003
$\widehat{K}_{SD}$	4.527×10^{-2}	4.378×10^{-2}	3.495×10^{-2}	4.354×10^{-2}
$SE(K)$	4.466×10^{-2}	4.534×10^{-2}	3.535×10^{-2}	4.266×10^{-2}
$\hat{r}_{Avg}$	0.7003	0.7004	0.7004	0.6995
$\hat{r}_{SD}$	8.560×10^{-3}	9.341×10^{-3}	2.247×10^{-2}	1.198×10^{-2}
$SE(r)$	7.167×10^{-3}	7.673×10^{-3}	1.928×10^{-2}	1.005×10^{-2}
$\hat{x}_{0,Avg}$	0.1000	0.1000	0.1009	0.1007
$\hat{x}_{0,SD}$	6.087×10^{-3}	7.058×10^{-3}	1.577×10^{-2}	8.944×10^{-3}
$SE(x_0)$	4.297×10^{-3}	4.965×10^{-3}	9.996×10^{-3}	6.194×10^{-3}

Table 6.2: Average estimates (θ_{Avg}) with their standard deviations (θ_{SD}) from $M = 1000$ Monte Carlo trials as well as asymptotic standard errors ($SE(\theta)$). Data was simulated using true values $\theta_0 = (17.5, 0.7, 0.1)$, $N = 10$, and 5% noise ($\sigma_0^2 = 0.05$).

	SE-opt	D-opt	E-opt	Uniform
$\widehat{K}_{Avg}$	17.5009	17.4979	17.5021	17.5059
$\widehat{K}_{SD}$	1.022×10^{-1}	1.027×10^{-1}	8.133×10^{-2}	9.990×10^{-2}
$SE(K)$	9.987×10^{-2}	1.014×10^{-1}	7.906×10^{-2}	9.540×10^{-2}
$\hat{r}_{Avg}$	0.7002	0.7005	0.7028	0.7007
$\hat{r}_{SD}$	1.895×10^{-2}	2.091×10^{-2}	5.290×10^{-2}	2.748×10^{-2}
$SE(r)$	1.603×10^{-2}	1.716×10^{-2}	4.312×10^{-2}	2.248×10^{-2}
$\hat{x}_{0,Avg}$	0.1007	0.1009	0.1044	0.1017
$\hat{x}_{0,SD}$	1.334×10^{-2}	1.607×10^{-2}	3.612×10^{-2}	2.025×10^{-2}
$SE(x_0)$	9.608×10^{-3}	1.110×10^{-2}	2.235×10^{-2}	1.385×10^{-2}

Table 6.3: Average estimates (θ_{Avg}) with their standard deviations (θ_{SD}) from $M = 1000$ Monte Carlo trials as well as asymptotic standard errors ($SE(\theta)$). Data was simulated using true values $\theta_0 = (17.5, 0.7, 0.1)$, $N = 10$, and 10% noise ($\sigma_0^2 = 0.10$).

	SE-opt	D-opt	E-opt	Uniform
$\widehat{K}_{Avg}$	17.5037	17.4942	17.4912	17.5067
$\widehat{K}_{SD}$	1.390×10^{-1}	1.449×10^{-1}	1.069×10^{-1}	1.369×10^{-1}
$SE(K)$	1.412×10^{-1}	1.4339×10^{-1}	1.118×10^{-1}	1.349×10^{-1}
$\hat{r}_{Avg}$	0.7004	0.7008	0.7079	0.7015
$\hat{r}_{SD}$	2.756×10^{-2}	3.140×10^{-2}	7.641×10^{-2}	3.882×10^{-2}
$SE(r)$	2.266×10^{-2}	2.426×10^{-2}	6.098×10^{-2}	3.179×10^{-2}
$\hat{x}_{0,Avg}$	0.1015	0.1023	0.1072	0.1028
$\hat{x}_{0,SD}$	1.926×10^{-2}	2.441×10^{-2}	5.158×10^{-2}	2.832×10^{-2}
$SE(x_0)$	1.359×10^{-2}	1.570×10^{-2}	3.161×10^{-2}	1.959×10^{-2}



(a)

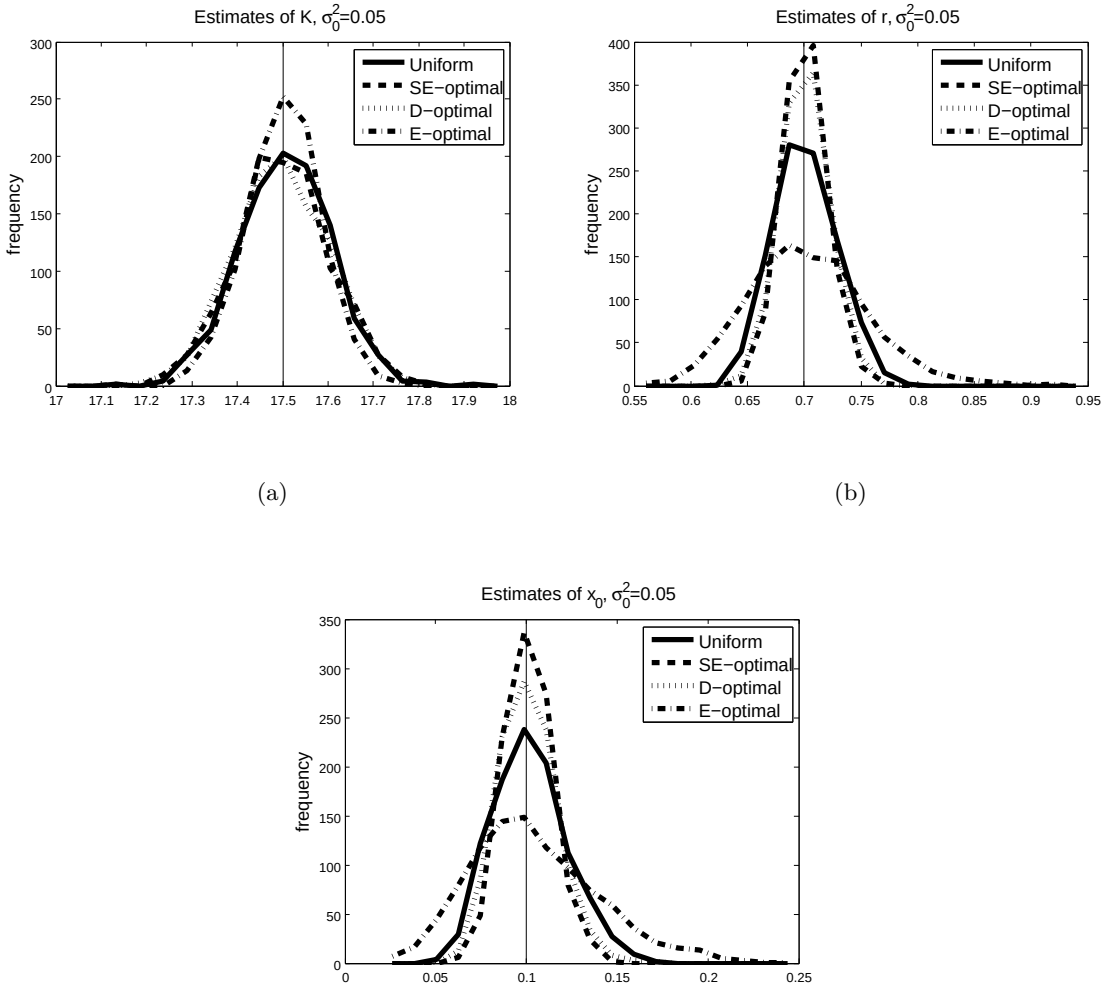
(b)

(c)

Figure 6.2: Histograms of parameter estimates ($\widehat{K}$ in panel (a), $\hat{r}$ in panel (b), $\hat{x}_0$ in panel (c)) resulting from Monte Carlo simulation with $M = 1000$. Different histograms within each subfigure represent results from different optimal design methods as well as from the uniform mesh. Simulated data was generated with $N = 10$ and 1% noise ($\sigma_0^2 = 0.01$).

6.2.1 Discussion of Logistic Monte Carlo Results

A general observation from these results is that as the noise level decreases, variance in the parameters ($\hat{\theta}_{SD}$) decreases, and parameter estimates approach the true parameter values. Comparing the standard deviation of the parameter estimates ($\hat{\theta}_{SD}$) to the asymptotic standard



(a)

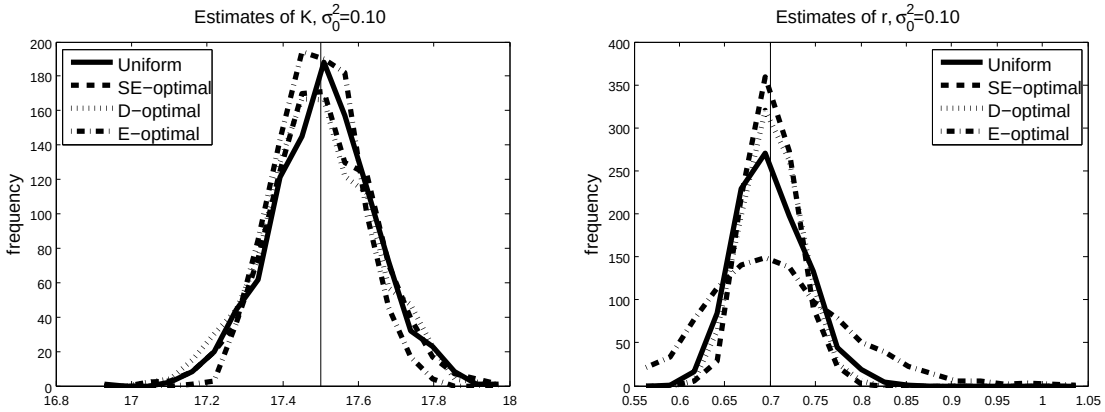
(b)

(c)

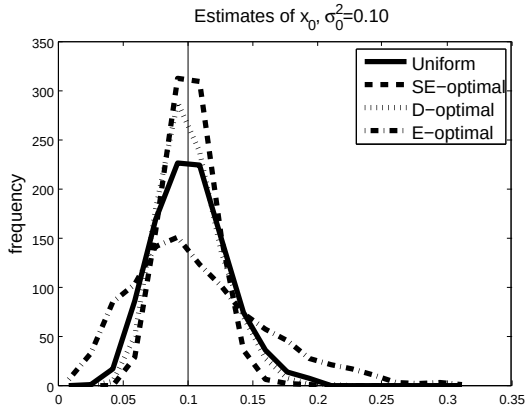
Figure 6.3: Histograms of parameter estimates ($\widehat{K}$ in panel (a), $\hat{r}$ in panel (b), $\hat{x}_0$ in panel (c)) resulting from Monte Carlo simulation with $M = 1000$. Different histograms within each subfigure represent results from different optimal design methods as well as from the uniform mesh. Simulated data was generated with $N = 10$ and 5% noise ($\sigma_0^2 = 0.05$).

errors ($SE(\theta)$) we find that they are always on the same order of magnitude, and often the asymptotic standard errors are slightly smaller.

In all cases the average parameter estimates were very close to the true values. Comparing the optimal design methods based on the standard deviations of the parameter estimates, we find different results based on which parameters are considered. For parameter K, E-optimal



(a) (b)



(c)

Figure 6.4: Histograms of parameter estimates ($\widehat{K}$ in panel (a), $\hat{r}$ in panel (b), $\hat{x}_0$ in panel (c)) resulting from Monte Carlo simulation with $M = 1000$. Different histograms within each subfigure represent results from different optimal design methods as well as from the uniform mesh. Simulated data was generated with $N = 10$ and 10% noise ($\sigma_0^2 = 0.10$).

had the smallest standard deviation in every case, followed by the uniform distribution. For parameters r and x_0, SE-optimal always had the smallest standard deviation, often followed by D-optimal. Visually, E-optimal appears right-skewed for parameters r and x_0 for $\sigma_0^2 = 0.10$, and $N = 10$ and $N = 15$ (see Figs. 6.4 and 6.7). Overall, these results seem to be in agreement with the performance of the optimal design methods in Section 5.1.

Table 6.4: Percent of Confidence Ellipsoids which contain the true parameter values (N=10).

	$\sigma_0^2 = 0.01$	$\sigma_0^2 = 0.05$	$\sigma_0^2 = 0.10$
SE-opt	94.8% ($SE = 0.70\%$)	94.5% ($SE = 0.72\%$)	92.2% ($SE = 0.85\%$)
D-opt	94.7% ($SE = 0.71\%$)	93.3% ($SE = 0.79\%$)	88.3% ($SE = 1.02\%$)
E-opt	91.2% ($SE = 0.90\%$)	82.7% ($SE = 1.20\%$)	76.7% ($SE = 1.34\%$)
Uniform	94.4% ($SE = 0.73\%$)	89.7% ($SE = 0.96\%$)	89.0% ($SE = 0.99\%$)

Table 6.5: Average estimates (θ_{Avg}) with their standard deviations (θ_{SD}) from $M = 1000$ Monte Carlo trials as well as asymptotic standard errors ($SE(\theta)$). Data was simulated using true values $\theta_0 = (17.5, 0.7, 0.1)$, $N = 15$, and 1% noise ($\sigma_0^2 = 0.01$).

	SE-opt	D-opt	E-opt	Uniform
$\widehat{K}_{Avg}$	17.5001	17.4990	17.4980	17.5000
$\widehat{K}_{SD}$	3.747×10^{-2}	3.544×10^{-2}	2.914×10^{-2}	3.534×10^{-2}
$SE(K)$	3.774×10^{-2}	3.612×10^{-2}	3.015×10^{-2}	3.537×10^{-2}
$\hat{r}_{Avg}$	0.7002	0.7000	0.7010	0.7001
$\hat{r}_{SD}$	6.740×10^{-3}	8.543×10^{-3}	1.946×10^{-2}	9.622×10^{-3}
$SE(r)$	5.547×10^{-3}	6.862×10^{-3}	1.713×10^{-2}	8.088×10^{-3}
$\hat{x}_{0,Avg}$	0.0999	0.1002	0.1001	0.1002
$\hat{x}_{0,SD}$	4.895×10^{-3}	6.309×10^{-3}	1.197×10^{-2}	6.993×10^{-3}
$SE(x_0)$	3.524×10^{-3}	4.185×10^{-3}	7.560×10^{-3}	4.989×10^{-3}

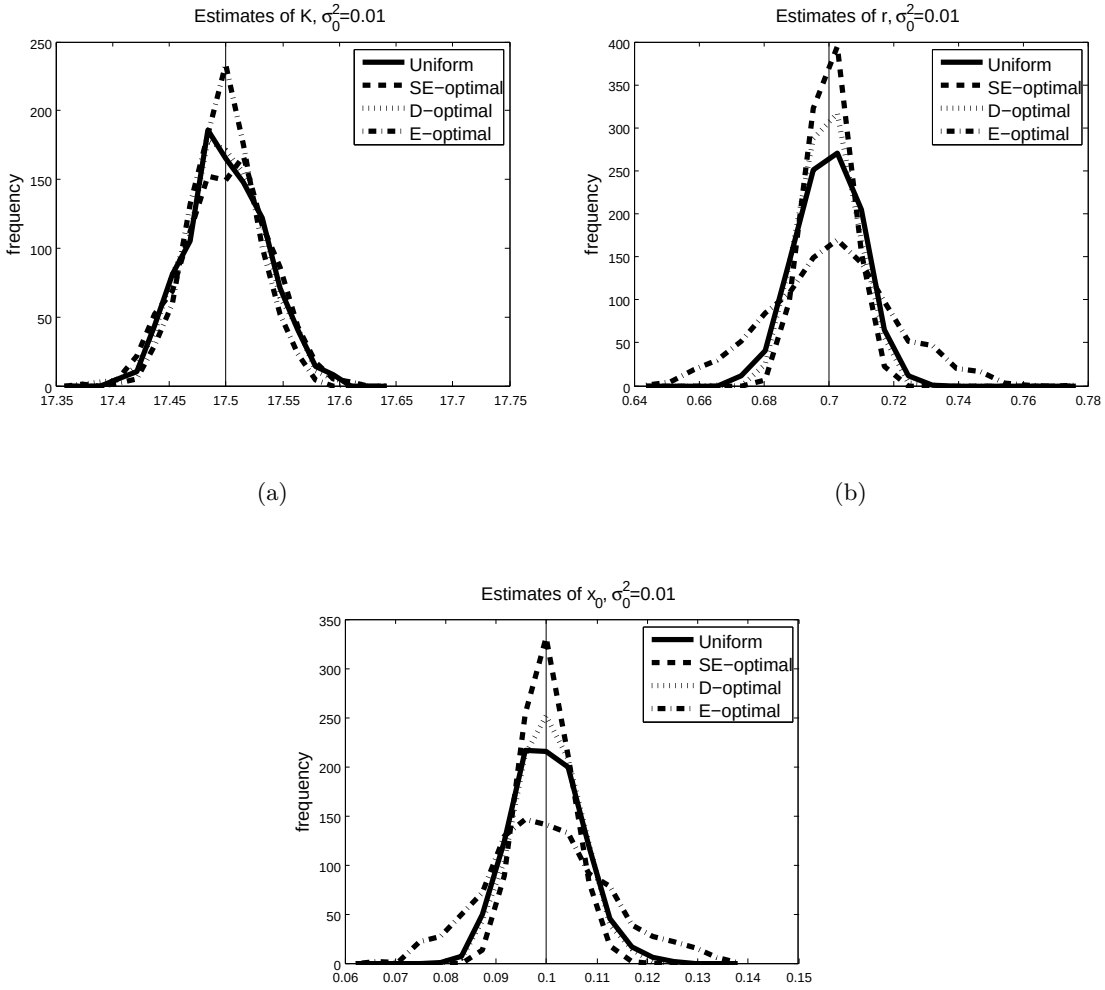
Examining the proportion of 95% confidence ellipsoids that contain the true parameter value (Tables 6.4 and 6.8) we find that often the proportion falls short of the expected 0.95, especially for increased values of σ_0^2. Comparing the optimal design methods based on which has the best coverage of the true parameter value, we find that SE-optimal is the best, followed by D-optimal or the uniform mesh, with E-optimal being the worst for this logistic example.

Table 6.6: Average estimates (θ_{Avg}) with their standard deviations (θ_{SD}) from $M = 1000$ Monte Carlo trials as well as asymptotic standard errors ($SE(\theta)$). Data was simulated using true values $\theta_0 = (17.5, 0.7, 0.1)$, $N = 15$, and 5% noise ($\sigma_0^2 = 0.05$).

	SE-opt	D-opt	E-opt	Uniform
$\widehat{K}_{Avg}$	17.4980	17.5029	17.4994	17.5002
$\widehat{K}_{SD}$	8.612×10^{-2}	7.832×10^{-2}	6.850×10^{-2}	7.883×10^{-2}
$SE(K)$	8.438×10^{-2}	8.077×10^{-2}	6.742×10^{-2}	7.910×10^{-2}
$\hat{r}_{Avg}$	0.7009	0.7005	0.7017	0.7008
$\hat{r}_{SD}$	1.529×10^{-2}	1.966×10^{-2}	4.524×10^{-2}	2.250×10^{-2}
$SE(r)$	1.240×10^{-2}	1.534×10^{-2}	3.831×10^{-2}	1.809×10^{-2}
$\hat{x}_{0,Avg}$	0.1000	0.1006	0.1028	0.1007
$\hat{x}_{0,SD}$	1.120×10^{-2}	1.453×10^{-2}	2.757×10^{-2}	1.647×10^{-2}
$SE(x_0)$	7.880×10^{-3}	9.357×10^{-3}	1.690×10^{-2}	1.116×10^{-2}

Table 6.7: Average estimates (θ_{Avg}) with their standard deviations (θ_{SD}) from $M = 1000$ Monte Carlo trials as well as asymptotic standard errors ($SE(\theta)$). Data was simulated using true values $\theta_0 = (17.5, 0.7, 0.1)$, $N = 15$, and 10% noise ($\sigma_0^2 = 0.10$).

	SE-opt	D-opt	E-opt	Uniform
$\widehat{K}_{Avg}$	17.4987	17.5014	17.5014	17.5071
$\widehat{K}_{SD}$	1.192×10^{-1}	1.135×10^{-1}	9.545×10^{-2}	1.151×10^{-1}
$SE(K)$	1.193×10^{-1}	1.142×10^{-1}	9.535×10^{-2}	1.119×10^{-1}
$\hat{r}_{Avg}$	0.7011	0.7017	0.7046	0.7010
$\hat{r}_{SD}$	2.070×10^{-2}	2.703×10^{-2}	6.565×10^{-2}	3.132×10^{-2}
$SE(r)$	1.754×10^{-2}	2.170×10^{-2}	5.417×10^{-2}	2.558×10^{-2}
$\hat{x}_{0,Avg}$	0.1007	0.1009	0.1043	0.1018
$\hat{x}_{0,SD}$	1.530×10^{-2}	2.026×10^{-2}	4.032×10^{-2}	2.317×10^{-2}
$SE(x_0)$	1.114×10^{-2}	1.323×10^{-2}	2.391×10^{-2}	1.578×10^{-2}

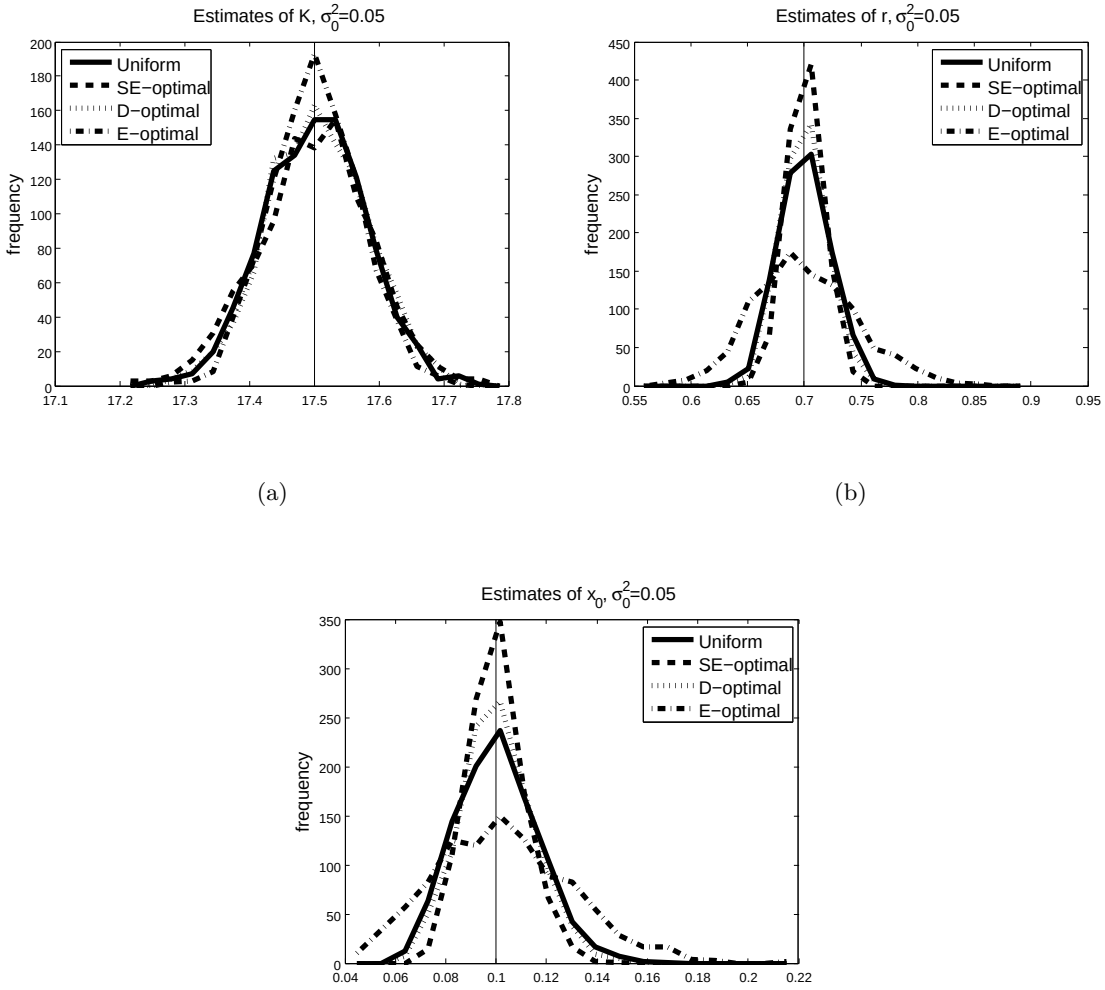


(a)

(b)

(c)

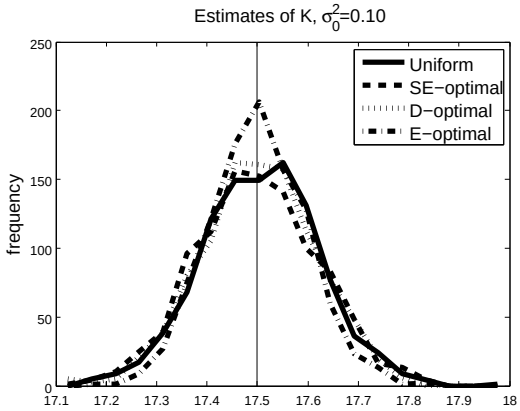
Figure 6.5: Histograms of parameter estimates ($\widehat{K}$ in panel (a), $\hat{r}$ in panel (b), $\hat{x}_0$ in panel (c)) resulting from Monte Carlo simulation with $M = 1000$. Different histograms within each subfigure represent results from different optimal design methods as well as from the uniform mesh. Simulated data was generated with $N = 15$ and 1% noise ($\sigma_0^2 = 0.01$).



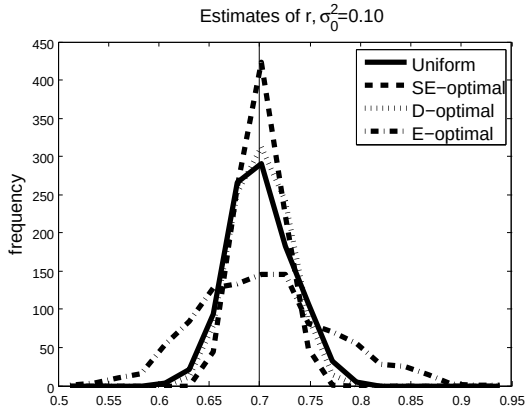
(a) (b)

(c)

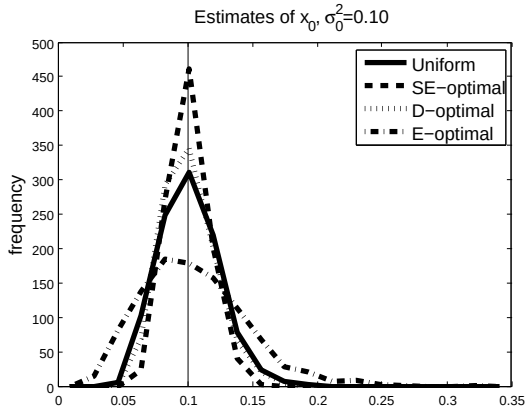
Figure 6.6: Histograms of parameter estimates ($\widehat{K}$ in panel (a), $\hat{r}$ in panel (b), $\hat{x}_0$ in panel (c)) resulting from Monte Carlo simulation with $M = 1000$. Different histograms within each subfigure represent results from different optimal design methods as well as from the uniform mesh. Simulated data was generated with $N = 15$ and 5% noise ($\sigma_0^2 = 0.05$).



(a)



(b)



(c)

Figure 6.7: Histograms of parameter estimates ($\widehat{K}$ in panel (a), $\hat{r}$ in panel (b), $\hat{x}_0$ in panel (c)) resulting from Monte Carlo simulation with $M = 1000$. Different histograms within each subfigure represent results from different optimal design methods as well as from the uniform mesh. Simulated data was generated with $N = 15$ and 10% noise ($\sigma_0^2 = 0.10$).

Table 6.8: Percent of Confidence Ellipsoids which contain the true parameter values (N=15).

	$\sigma_0^2 = 0.01$	$\sigma_0^2 = 0.05$	$\sigma_0^2 = 0.10$
SE-opt	95.7% ($SE = 0.64\%$)	94.0% ($SE = 0.75\%$)	93.5% ($SE = 0.78\%$)
D-opt	95.1% ($SE = 0.68\%$)	92.3% ($SE = 0.84\%$)	89.2% ($SE = 0.98\%$)
E-opt	93.9% ($SE = 0.76\%$)	87.8% ($SE = 1.04\%$)	82.5% ($SE = 1.20\%$)
Uniform	94.0% ($SE = 0.75\%$)	91.2%,($SE = 0.90\%$)	89.7% ($SE = 0.96\%$)

6.3 Harmonic Oscillator Results

Monte Carlo analysis for the Harmonic Oscillator model was conducted using the optimal design points found in the previous section (Section 5.2.3) with $N = 15$, $T = 14.14$ or $T = 28.28$, and four different implementations of the constraint $(C1) - (C4)$ (See Figs. 5.9-5.12). For each optimal mesh or uniform mesh, Monte Carlo analysis was repeated three times for different noise levels $\sigma_0^2 = (0.01, 0.05, 0.10)$. Figures 6.8-6.15 contain histogram plots of the parameter estimates $(\widehat{C}, \widehat{K})$. Each figure represents results from a specific constraint implementation and a value of T. Each subfigure represents one of the parameters and one value of σ_0^2. Within a subfigure are histograms corresponding to each optimal design method (SE-optimal, D-optimal, and E-optimal) and the uniform mesh.

Tables 6.9-6.11, 6.14-6.16, 6.19-6.21, and 6.24-6.26 contain the average estimates and standard deviations from the Monte Carlo trials for the four constraints respectively, and both $T = 14.14$ and $T = 28.28$, as well as for the three different noise levels. In addition, the tables contain the asymptotic standard errors (4.5), using the true parameter values, for comparison with the standard deviations of the Monte Carlo parameter estimates. The percentage of 95% confidence ellipsoids which contain the true parameter values are given in Tables 6.12, 6.13, 6.17, 6.18, 6.22, 6.23, 6.27, and 6.28.

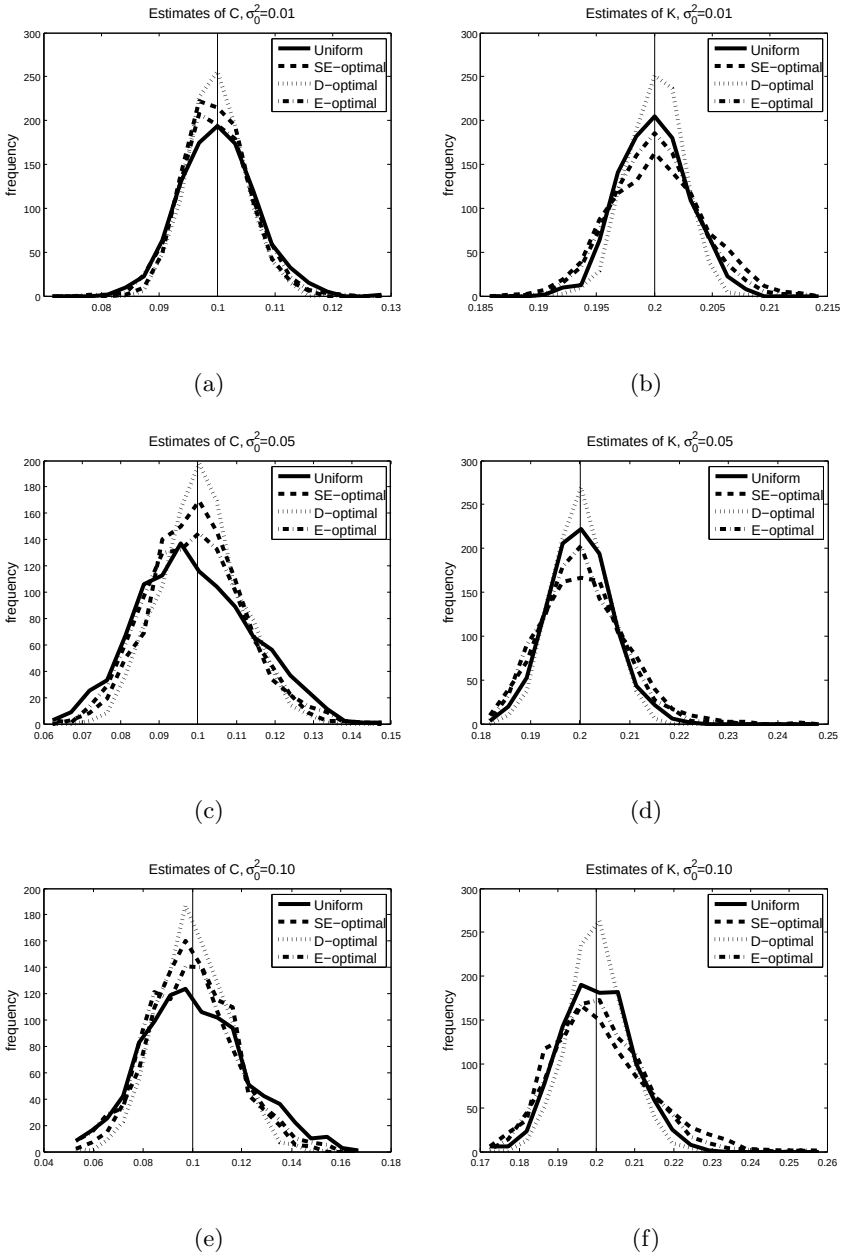


Figure 6.8: Histograms of parameter estimates ($\widehat{C}$ in panels (a),(c),(e), and $\widehat{K}$ in panel (b),(d),(f)) resulting from Monte Carlo simulation. Simulated data was generated with $\sigma_0^2 = 0.01$ (top row), $\sigma_0^2 = 0.05$ (middle row), and $\sigma_0^2 = 0.10$ (bottom row) ($T = 14.14$ and constraint implementation $(C1)$).

Table 6.9: Average estimates (θ_{Avg}) with their standard deviations (θ_{SD}) from $M = 1000$ Monte Carlo trials as well as asymptotic standard errors ($SE(\theta)$). Data was simulated using true values $\theta_0 = (C, K) = (0.1, 0.2)$, fixed parameter values $(x_1, x_2) = (-1, 0.5)$, and noise level $\sigma_0^2 = 0.01$. Simulated data corresponds to optimal design points using $N = 15$, $T = 14.14$ and $T = 28.28$, and constraint implementation ($C1$).

T = 14.14	SE-opt	D-opt	E-opt	Uniform
$\widehat{C}_{Avg}$	0.0998	0.0999	0.0999	0.1002
$\widehat{C}_{SD}$	5.197×10^{-3}	4.904×10^{-3}	5.928×10^{-3}	6.495×10^{-3}
$SE(C)$	5.004×10^{-3}	4.746×10^{-3}	5.818×10^{-3}	6.577×10^{-3}
$\widehat{K}_{Avg}$	0.2002	0.2001	0.2000	0.2000
$\widehat{K}_{SD}$	4.049×10^{-3}	2.360×10^{-3}	3.482×10^{-3}	2.977×10^{-3}
$SE(K)$	3.678×10^{-3}	2.378×10^{-3}	3.416×10^{-3}	2.895×10^{-3}
T = 28.28				
$\widehat{C}_{Avg}$	0.1000	0.0999	0.0999	0.1000
$\widehat{C}_{SD}$	4.493×10^{-3}	4.319×10^{-3}	5.707×10^{-3}	5.123×10^{-3}
$SE(C)$	4.342×10^{-3}	4.284×10^{-3}	5.214×10^{-3}	5.127×10^{-3}
$\widehat{K}_{Avg}$	0.2002	0.2000	0.2002	0.2000
$\widehat{K}_{SD}$	3.238×10^{-3}	1.854×10^{-3}	3.981×10^{-3}	2.413×10^{-3}
$SE(K)$	3.021×10^{-3}	1.994×10^{-3}	3.412×10^{-3}	2.307×10^{-3}

Table 6.10: Average estimates (θ_{Avg}) with their standard deviations (θ_{SD}) from $M = 1000$ Monte Carlo trials as well as asymptotic standard errors ($SE(\theta)$). Data was simulated using true values $\theta_0 = (C, K) = (0.1, 0.2)$, fixed parameter values $(x_1, x_2) = (-1, 0.5)$, and noise level $\sigma_0^2 = 0.05$. Simulated data corresponds to optimal design points using $N = 15$, $T = 14.14$ and $T = 28.28$, and constraint implementation ($C1$).

$\mathbf{T = 14.14}$	SE-opt	D-opt	E-opt	Uniform
$\widehat{C}_{Avg}$	0.0999	0.1008	0.0994	0.0994
$\widehat{C}_{SD}$	1.145×10^{-2}	1.019×10^{-2}	1.286×10^{-2}	1.457×10^{-2}
$SE(C)$	1.119×10^{-2}	1.061×10^{-2}	1.301×10^{-2}	1.471×10^{-2}
$\widehat{K}_{Avg}$	0.2008	0.2000	0.2001	0.2001
$\widehat{K}_{SD}$	8.782×10^{-3}	5.430×10^{-3}	7.761×10^{-3}	6.469×10^{-3}
$SE(K)$	8.224×10^{-3}	5.318×10^{-3}	7.640×10^{-3}	6.473×10^{-3}
$\mathbf{T = 28.28}$				
$\widehat{C}_{Avg}$	0.0996	0.0997	0.0988	0.1001
$\widehat{C}_{SD}$	9.659×10^{-3}	9.588×10^{-3}	1.224×10^{-2}	1.183×10^{-2}
$SE(C)$	9.708×10^{-3}	9.579×10^{-3}	1.166×10^{-2}	1.146×10^{-2}
$\widehat{K}_{Avg}$	0.2004	0.2000	0.2013	0.2001
$\widehat{K}_{SD}$	7.795×10^{-3}	4.295×10^{-3}	1.129×10^{-2}	5.114×10^{-3}
$SE(K)$	6.755×10^{-3}	4.459×10^{-3}	7.629×10^{-3}	5.159×10^{-3}

Table 6.11: Average estimates (θ_{Avg}) with their standard deviations (θ_{SD}) from $M = 1000$ Monte Carlo trials as well as asymptotic standard errors ($SE(\theta)$). Data was simulated using true values $\theta_0 = (C, K) = (0.1, 0.2)$, fixed parameter values $(x_1, x_2) = (-1, 0.5)$, and noise level $\sigma_0^2 = 0.10$. Simulated data corresponds to optimal design points using $N = 15$, $T = 14.14$ and $T = 28.28$, and constraint implementation ($C1$).

T = 14.14	SE-opt	D-opt	E-opt	Uniform
$\widehat{C}_{Avg}$	0.0999	0.1003	0.0991	0.1008
$\widehat{C}_{SD}$	1.621×10^{-2}	1.445×10^{-2}	1.842×10^{-2}	2.107×10^{-2}
$SE(C)$	1.582×10^{-2}	1.501×10^{-2}	1.840×10^{-2}	2.080×10^{-2}
$\widehat{K}_{Avg}$	0.2008	0.2002	0.2007	0.2000
$\widehat{K}_{SD}$	1.286×10^{-2}	7.447×10^{-3}	1.131×10^{-2}	9.192×10^{-3}
$SE(K)$	1.163×10^{-2}	7.521×10^{-3}	1.080×10^{-2}	9.154×10^{-3}
T = 28.28				
$\widehat{C}_{Avg}$	0.0991	0.0998	0.0978	0.0992
$\widehat{C}_{SD}$	1.386×10^{-2}	1.353×10^{-2}	1.714×10^{-2}	1.628×10^{-2}
$SE(C)$	1.373×10^{-2}	1.355×10^{-2}	1.649×10^{-2}	1.621×10^{-2}
$\widehat{K}_{Avg}$	0.2015	0.2003	0.2038	0.2001
$\widehat{K}_{SD}$	1.106×10^{-2}	6.602×10^{-3}	1.807×10^{-2}	7.350×10^{-3}
$SE(K)$	9.553×10^{-3}	6.305×10^{-3}	1.079×10^{-2}	7.295×10^{-3}

Table 6.12: Percent of Confidence Ellipsoids which contain the true parameter values ($T = 14.14$, constraint implementation ($C1$)).

	$\sigma_0^2 = 0.01$	$\sigma_0^2 = 0.05$	$\sigma_0^2 = 0.10$
SE-opt	95.9% ($SE = 0.63\%$)	94.7% ($SE = 0.71\%$)	94.3% ($SE = 0.73\%$)
D-opt	95.2% ($SE = 0.68\%$)	96.7% ($SE = 0.56\%$)	95.9% ($SE = 0.63\%$)
E-opt	95.5% ($SE = 0.66\%$)	94.4% ($SE = 0.73\%$)	94.6% ($SE = 0.71\%$)
Uniform	94.6% ($SE = 0.71\%$)	95.2% ($SE = 0.68\%$)	95.4% ($SE = 0.66\%$)

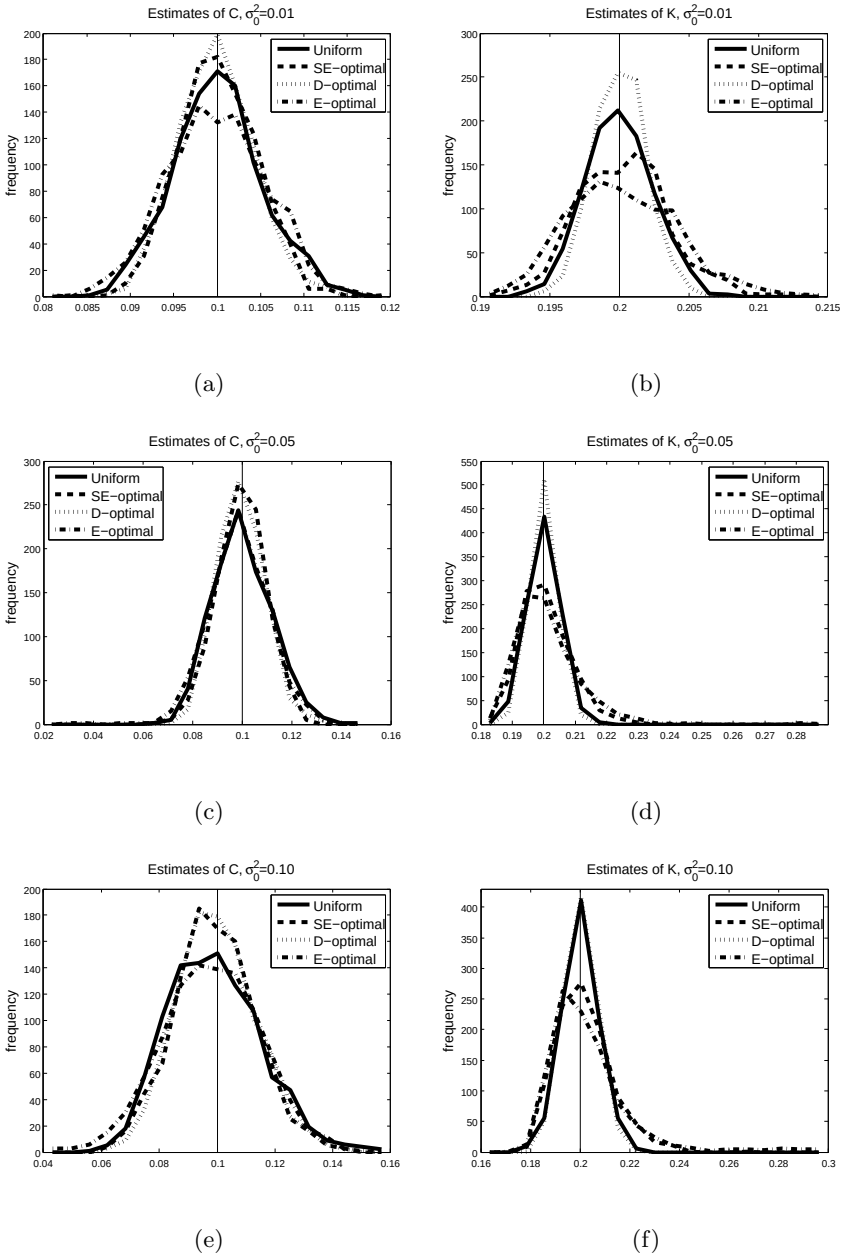


Figure 6.9: Histograms of parameter estimates ($\widehat{C}$ in panels (a),(c),(e), and $\widehat{K}$ in panel (b),(d),(f)) resulting from Monte Carlo simulation. Simulated data was generated with $\sigma_0^2 = 0.01$ (top row), $\sigma_0^2 = 0.05$ (middle row), and $\sigma_0^2 = 0.10$ (bottom row) ($T = 28.28$ and constraint implementation $(C1)$).

Table 6.13: Percent of Confidence Ellipsoids which contain the true parameter values ($T = 28.28$, constraint implementation ($C1$)).

	$\sigma_0^2 = 0.01$	$\sigma_0^2 = 0.05$	$\sigma_0^2 = 0.10$
SE-opt	95.5% ($SE = 0.66\%$)	92.6% ($SE = 0.83\%$)	93.2% ($SE = 0.80\%$)
D-opt	95.7% ($SE = 0.64\%$)	95.1% ($SE = 0.68\%$)	93.1% ($SE = 0.80\%$)
E-opt	94.9% ($SE = 0.70\%$)	92.8% ($SE = 0.82\%$)	90.5% ($SE = 0.93\%$)
Uniform	95.3% ($SE = 0.67\%$)	94.2% ($SE = 0.74\%$)	93.3% ($SE = 0.79\%$)

Table 6.14: Average estimates (θ_{Avg}) with their standard deviations (θ_{SD}) from $M = 1000$ Monte Carlo trials as well as asymptotic standard errors ($SE(\theta)$). Data was simulated using true values $\theta_0 = (C, K) = (0.1, 0.2)$, fixed parameter values $(x_1, x_2) = (-1, 0.5)$, and noise level $\sigma_0^2 = 0.01$. Simulated data corresponds to optimal design points using $N = 15$, $T = 14.14$ and $T = 28.28$ and constraint implementation ($C2$).

$\mathbf{T = 14.14}$	SE-opt	D-opt	E-opt	Uniform
$\widehat{C}_{Avg}$	0.1000	0.1000	0.0998	0.0998
$\widehat{C}_{SD}$	5.264×10^{-3}	5.261×10^{-3}	5.536×10^{-3}	6.431×10^{-3}
$SE(C)$	5.444×10^{-3}	4.987×10^{-3}	5.700×10^{-3}	6.577×10^{-3}
$\widehat{K}_{Avg}$	0.2002	0.2000	0.2002	0.2000
$\widehat{K}_{SD}$	3.424×10^{-3}	2.211×10^{-3}	3.112×10^{-3}	2.867×10^{-3}
$SE(K)$	3.344×10^{-3}	2.181×10^{-3}	3.008×10^{-3}	2.895×10^{-3}
$\mathbf{T = 28.28}$				
$\widehat{C}_{Avg}$	0.0999	0.0998	0.0998	0.1001
$\widehat{C}_{SD}$	4.741×10^{-3}	4.630×10^{-3}	5.454×10^{-3}	5.342×10^{-3}
$SE(C)$	4.537×10^{-3}	4.528×10^{-3}	5.279×10^{-3}	5.127×10^{-3}
$\widehat{K}_{Avg}$	0.2001	0.2000	0.2000	0.2000
$\widehat{K}_{SD}$	3.414×10^{-3}	2.202×10^{-3}	3.509×10^{-3}	2.285×10^{-3}
$SE(K)$	3.144×10^{-3}	2.137×10^{-3}	3.260×10^{-3}	2.307×10^{-3}

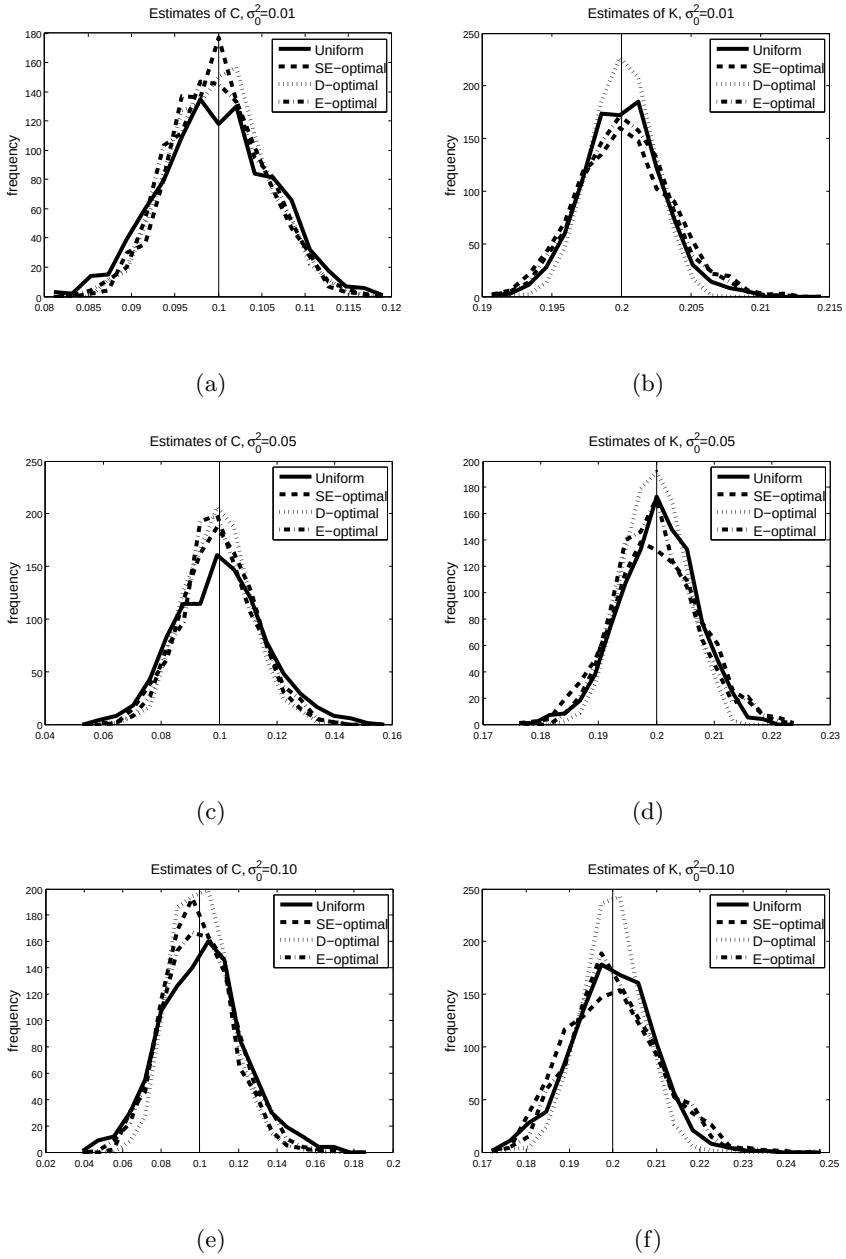


Figure 6.10: Histograms of parameter estimates ($\widehat{C}$ in panels (a),(c),(e), and $\widehat{K}$ in panel (b),(d),(f)) resulting from Monte Carlo simulation. Simulated data was generated with $\sigma_0^2 = 0.01$ (top row), $\sigma_0^2 = 0.05$ (middle row), and $\sigma_0^2 = 0.10$ (bottom row) ($T = 14.14$ and constraint implementation $(C2)$).

Table 6.15: Average estimates (θ_{Avg}) with their standard deviations (θ_{SD}) from $M = 1000$ Monte Carlo trials as well as asymptotic standard errors ($SE(\theta)$). Data was simulated using true values $\theta_0 = (C, K) = (0.1, 0.2)$, fixed parameter values $(x_1, x_2) = (-1, 0.5)$, and noise level $\sigma_0^2 = 0.05$. Simulated data corresponds to optimal design points using $N = 15$, $T = 14.14$ and $T = 28.28$, and constraint implementation ($C2$).

T = 14.14	SE-opt	D-opt	E-opt	Uniform
$\widehat{C}_{Avg}$	0.0996	0.0997	0.0999	0.1005
$\widehat{C}_{SD}$	1.247×10^{-2}	1.125×10^{-2}	1.238×10^{-2}	1.564×10^{-2}
$SE(C)$	1.217×10^{-2}	1.115×10^{-2}	1.275×10^{-2}	1.471×10^{-2}
$\widehat{K}_{Avg}$	0.2002	0.1997	0.1999	0.2002
$\widehat{K}_{SD}$	7.715×10^{-3}	5.107×10^{-3}	6.751×10^{-3}	6.549×10^{-3}
$SE(K)$	7.477×10^{-3}	4.876×10^{-3}	6.727×10^{-3}	6.473×10^{-3}
T = 28.28				
$\widehat{C}_{Avg}$	0.0990	0.0997	0.0998	0.1007
$\widehat{C}_{SD}$	1.008×10^{-2}	9.940×10^{-3}	1.172×10^{-2}	1.153×10^{-2}
$SE(C)$	1.015×10^{-2}	1.012×10^{-2}	1.180×10^{-2}	1.146×10^{-2}
$\widehat{K}_{Avg}$	0.2006	0.2000	0.2009	0.2000
$\widehat{K}_{SD}$	7.797×10^{-3}	4.753×10^{-3}	8.777×10^{-3}	5.443×10^{-3}
$SE(K)$	7.030×10^{-3}	4.778×10^{-3}	7.289×10^{-3}	5.159×10^{-3}

Table 6.16: Average estimates (θ_{Avg}) with their standard deviations (θ_{SD}) from $M = 1000$ Monte Carlo trials as well as asymptotic standard errors ($SE(\theta)$). Data was simulated using true values $\theta_0 = (C, K) = (0.1, 0.2)$, fixed parameter values $(x_1, x_2) = (-1, 0.5)$, and noise level $\sigma_0^2 = 0.10$. Simulated data corresponds to optimal design points using $N = 15$, $T = 14.14$ and $T = 28.28$, and constraint implementation ($C2$).

T = 14.14	SE-opt	D-opt	E-opt	Uniform
$\widehat{C}_{Avg}$	0.0988	0.1002	0.1003	0.1014
$\widehat{C}_{SD}$	1.703×10^{-2}	1.492×10^{-2}	1.863×10^{-2}	2.138×10^{-2}
$SE(C)$	1.721×10^{-2}	1.577×10^{-2}	1.802×10^{-2}	2.080×10^{-2}
$\widehat{K}_{Avg}$	0.1998	0.1997	0.2004	0.2000
$\widehat{K}_{SD}$	1.073×10^{-2}	6.885×10^{-3}	9.921×10^{-3}	9.483×10^{-3}
$SE(K)$	1.057×10^{-2}	6.895×10^{-3}	9.513×10^{-3}	9.154×10^{-3}
T = 28.28				
$\widehat{C}_{Avg}$	0.0990	0.0997	0.0998	0.1007
$\widehat{C}_{SD}$	1.008×10^{-2}	9.940×10^{-3}	1.172×10^{-2}	1.153×10^{-2}
$SE(C)$	1.015×10^{-2}	1.012×10^{-2}	1.180×10^{-2}	1.146×10^{-2}
$\widehat{K}_{Avg}$	0.2006	0.2000	0.2009	0.2000
$\widehat{K}_{SD}$	7.797×10^{-3}	4.753×10^{-3}	8.777×10^{-3}	5.443×10^{-3}
$SE(K)$	7.030×10^{-3}	4.778×10^{-3}	7.289×10^{-3}	5.159×10^{-3}

Table 6.17: Percent of Confidence Ellipsoids which contain the true parameter values ($T = 14.14$, constraint implementation ($C2$)).

	$\sigma_0^2 = 0.01$	$\sigma_0^2 = 0.05$	$\sigma_0^2 = 0.10$
SE-opt	95.5% ($SE = 0.66\%$)	93.9% ($SE = 0.76\%$)	94.3% ($SE = 0.73\%$)
D-opt	95.3% ($SE = 0.67\%$)	93.7% ($SE = 0.77\%$)	95.5% ($SE = 0.66\%$)
E-opt	94.6% ($SE = 0.71\%$)	94.9% ($SE = 0.70\%$)	94% ($SE = 0.75\%$)
Uniform	95.1% ($SE = 0.68\%$)	95.1% ($SE = 0.68\%$)	94.5% ($SE = 0.72\%$)

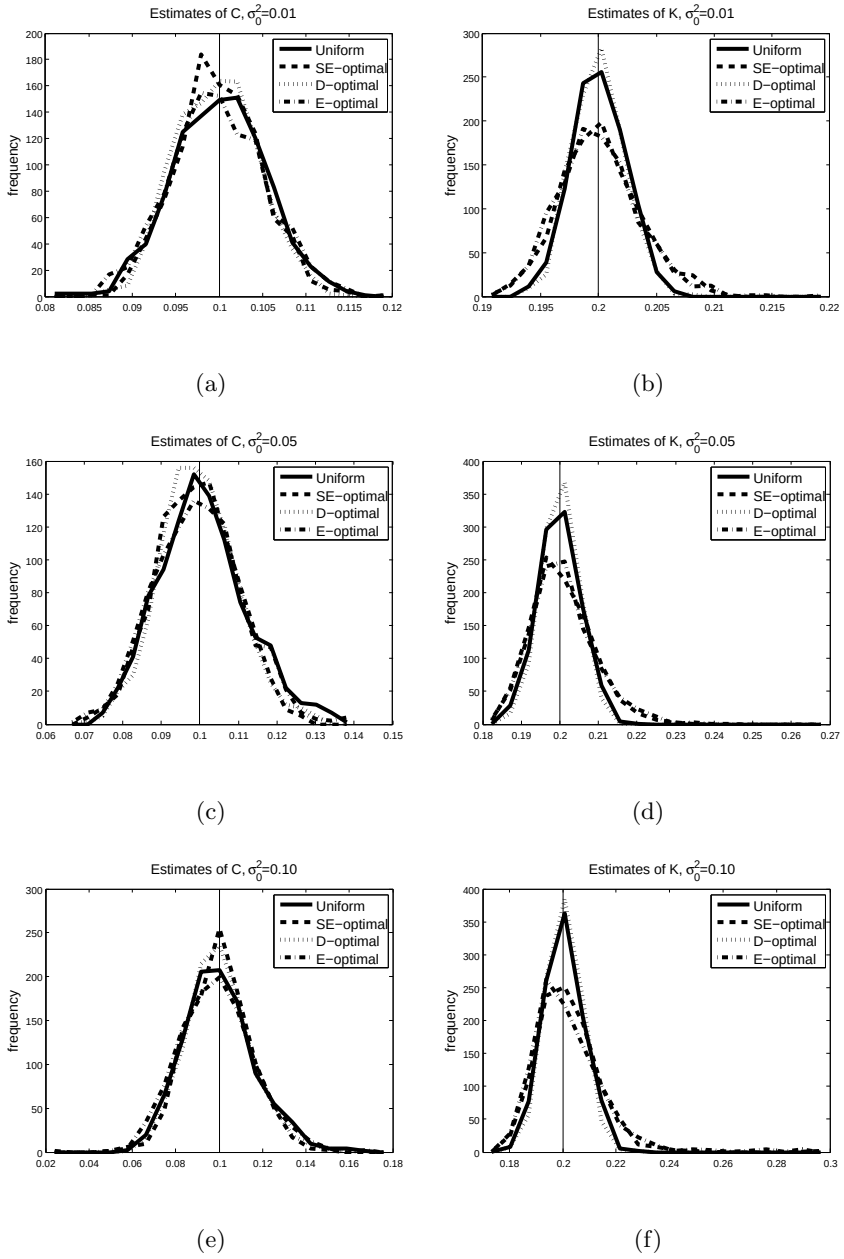


Figure 6.11: Histograms of parameter estimates ($\widehat{C}$ in panels (a),(c),(e), and $\widehat{K}$ in panel (b),(d),(f)) resulting from Monte Carlo simulation. Simulated data was generated with $\sigma_0^2 = 0.01$ (top row), $\sigma_0^2 = 0.05$ (middle row), and $\sigma_0^2 = 0.10$ (bottom row) ($T = 28.28$ and constraint implementation $(C2)$).

Table 6.18: Percent of Confidence Ellipsoids which contain the true parameter values ($T = 28.28$, constraint implementation ($C2$)).

	$\sigma_0^2 = 0.01$	$\sigma_0^2 = 0.05$	$\sigma_0^2 = 0.10$
SE-opt	94.6% ($SE = 0.71\%$)	93.4% ($SE = 0.79\%$)	93.1% ($SE = 0.80\%$)
D-opt	93.7% ($SE = 0.77\%$)	95% ($SE = 0.69\%$)	92.3% ($SE = 0.84\%$)
E-opt	93% ($SE = 0.81\%$)	92.7% ($SE = 0.82\%$)	92.4% ($SE = 0.84\%$)
Uniform	94.6% ($SE = 0.71\%$)	94.9% ($SE = 0.70\%$)	95.2% ($SE = 0.68\%$)

Table 6.19: Average estimates (θ_{Avg}) with their standard deviations (θ_{SD}) from $M = 1000$ Monte Carlo trials as well as asymptotic standard errors ($SE(\theta)$). Data was simulated using true values $\theta_0 = (C, K) = (0.1, 0.2)$, fixed parameter values $(x_1, x_2) = (-1, 0.5)$, and noise levels $\sigma_0^2 = 0.01$. Simulated data corresponds to optimal design points using $N = 15$, $T = 14.14$ and $T = 28.28$, and constraint implementation ($C3$).

T = 14.14	SE-opt	D-opt	E-opt	Uniform
$\widehat{C}_{Avg}$	0.1002	0.1002	0.1003	0.0999
$\widehat{C}_{SD}$	5.396×10^{-3}	4.957×10^{-3}	6.616×10^{-3}	6.664×10^{-3}
$SE(C)$	5.124×10^{-3}	4.938×10^{-3}	6.361×10^{-3}	6.577×10^{-3}
$\widehat{K}_{Avg}$	0.2002	0.2000	0.2002	0.2000
$\widehat{K}_{SD}$	3.888×10^{-3}	2.318×10^{-3}	4.770×10^{-3}	2.888×10^{-3}
$SE(K)$	3.735×10^{-3}	2.248×10^{-3}	4.162×10^{-3}	2.895×10^{-3}
T = 28.28				
$\widehat{C}_{Avg}$	0.0999	0.0999	0.1002	0.1000
$\widehat{C}_{SD}$	4.785×10^{-3}	4.398×10^{-3}	5.593×10^{-3}	5.147×10^{-3}
$SE(C)$	4.492×10^{-3}	4.372×10^{-3}	5.409×10^{-3}	5.127×10^{-3}
$\widehat{K}_{Avg}$	0.2001	0.1999	0.2005	0.1999
$\widehat{K}_{SD}$	3.325×10^{-3}	2.027×10^{-3}	4.067×10^{-3}	2.317×10^{-3}
$SE(K)$	3.088×10^{-3}	2.005×10^{-3}	3.539×10^{-3}	2.307×10^{-3}

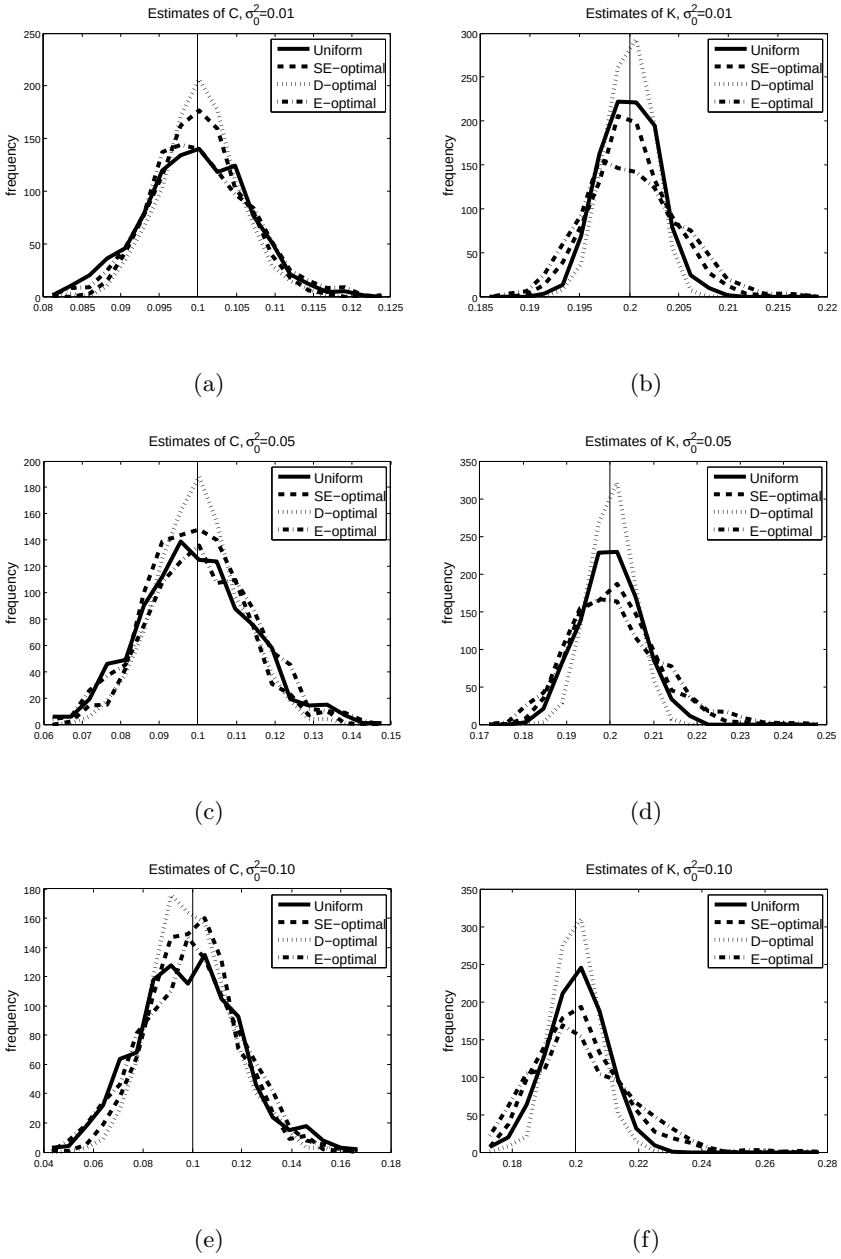


Figure 6.12: Histograms of parameter estimates ($\widehat{C}$ in panels (a),(c),(e), and $\widehat{K}$ in panel (b),(d),(f)) resulting from Monte Carlo simulation. Simulated data was generated with $\sigma_0^2 = 0.01$ (top row), $\sigma_0^2 = 0.05$ (middle row), and $\sigma_0^2 = 0.10$ (bottom row) ($T = 14.14$ and constraint implementation $(C3)$).

Table 6.20: Average estimates (θ_{Avg}) with their standard deviations (θ_{SD}) from $M = 1000$ Monte Carlo trials as well as asymptotic standard errors ($SE(\theta)$). Data was simulated using true values $\theta_0 = (C, K) = (0.1, 0.2)$, fixed parameter values $(x_1, x_2) = (-1, 0.5)$, and noise levels $\sigma_0^2 = 0.05$. Simulated data corresponds to optimal design points using $N = 15$, $T = 14.14$ and $T = 28.28$, and constraint implementation ($C3$).

$\mathbf{T = 14.14}$	SE-opt	D-opt	E-opt	Uniform
$\widehat{C}_{Avg}$	0.0997	0.1001	0.1008	0.0996
$\widehat{C}_{SD}$	1.185×10^{-2}	1.082×10^{-2}	1.451×10^{-2}	1.474×10^{-2}
$SE(C)$	1.146×10^{-2}	1.104×10^{-2}	1.422×10^{-2}	1.471×10^{-2}
$\widehat{K}_{Avg}$	0.2007	0.2002	0.2014	0.2000
$\widehat{K}_{SD}$	9.063×10^{-3}	4.951×10^{-3}	1.089×10^{-2}	6.885×10^{-3}
$SE(K)$	8.351×10^{-3}	5.028×10^{-3}	9.307×10^{-3}	6.473×10^{-3}
$\mathbf{T = 28.28}$				
$\widehat{C}_{Avg}$	0.0993	0.1000	0.0991	0.0992
$\widehat{C}_{SD}$	1.010×10^{-2}	9.973×10^{-3}	1.313×10^{-2}	1.147×10^{-2}
$SE(C)$	1.005×10^{-2}	9.775×10^{-3}	1.209×10^{-2}	1.146×10^{-2}
$\widehat{K}_{Avg}$	0.2006	0.2001	0.2027	0.2003
$\widehat{K}_{SD}$	7.581×10^{-3}	4.465×10^{-3}	1.433×10^{-2}	5.200×10^{-3}
$SE(K)$	6.905×10^{-3}	4.484×10^{-3}	7.913×10^{-3}	5.159×10^{-3}

Table 6.21: Average estimates (θ_{Avg}) with their standard deviations (θ_{SD}) from $M = 1000$ Monte Carlo trials as well as asymptotic standard errors ($SE(\theta)$). Data was simulated using true values $\theta_0 = (C, K) = (0.1, 0.2)$, fixed parameter values $(x_1, x_2) = (-1, 0.5)$, and noise levels $\sigma_0^2 = 0.10$. Simulated data corresponds to optimal design points using $N = 15$, $T = 14.14$ and $T = 28.28$, and constraint implementation ($C3$).

T = 14.14	SE-opt	D-opt	E-opt	Uniform
$\widehat{C}_{Avg}$	0.1002	0.1001	0.0995	0.0988
$\widehat{C}_{SD}$	1.736×10^{-2}	1.603×10^{-2}	2.028×10^{-2}	2.074×10^{-2}
$SE(C)$	1.620×10^{-2}	1.562×10^{-2}	2.012×10^{-2}	2.080×10^{-2}
$\widehat{K}_{Avg}$	0.2012	0.2001	0.2021	0.2004
$\widehat{K}_{SD}$	1.310×10^{-2}	7.205×10^{-3}	1.665×10^{-2}	9.417×10^{-3}
$SE(K)$	1.181×10^{-2}	7.110×10^{-3}	1.316×10^{-2}	9.154×10^{-3}
T = 28.28				
$\widehat{C}_{Avg}$	0.0989	0.1001	0.0968	0.1007
$\widehat{C}_{SD}$	1.448×10^{-2}	1.372×10^{-2}	1.743×10^{-2}	1.730×10^{-2}
$SE(C)$	1.421×10^{-2}	1.382×10^{-2}	1.710×10^{-2}	1.621×10^{-2}
$\widehat{K}_{Avg}$	0.2019	0.2002	0.2056	0.2002
$\widehat{K}_{SD}$	1.218×10^{-2}	6.432×10^{-3}	2.156×10^{-2}	7.437×10^{-3}
$SE(K)$	9.765×10^{-3}	6.341×10^{-3}	1.119×10^{-2}	7.295×10^{-3}

Table 6.22: Percent of Confidence Ellipsoids which contain the true parameter values ($T = 14.14$, constraint implementation ($C3$)).

	$\sigma_0^2 = 0.01$	$\sigma_0^2 = 0.05$	$\sigma_0^2 = 0.10$
SE-opt	94.7% ($SE = 0.71\%$)	94.9% ($SE = 0.70\%$)	93.5% ($SE = 0.78\%$)
D-opt	93.9% ($SE = 0.76\%$)	95.1% ($SE = 0.68\%$)	95.1% ($SE = 0.68\%$)
E-opt	94.9% ($SE = 0.70\%$)	93.8% ($SE = 0.76\%$)	92.1% ($SE = 0.85\%$)
Uniform	94.3% ($SE = 0.73\%$)	93.7% ($SE = 0.77\%$)	95.6% ($SE = 0.65\%$)

Table 6.23: Percent of Confidence Ellipsoids which contain the true parameter values ($T = 28.28$, constraint implementation ($C3$)).

	$\sigma_0^2 = 0.01$	$\sigma_0^2 = 0.05$	$\sigma_0^2 = 0.10$
SE-opt	93.9% ($SE = 0.76\%$)	94.6% ($SE = 0.71\%$)	93.3% ($SE = 0.79\%$)
D-opt	94.4% ($SE = 0.73\%$)	94.7% ($SE = 0.71\%$)	94.6% ($SE = 0.71\%$)
E-opt	95.3% ($SE = 0.67\%$)	91.1% ($SE = 0.90\%$)	86.5% ($SE = 1.08\%$)
Uniform	94.7% ($SE = 0.71\%$)	95.3% ($SE = 0.67\%$)	93.9% ($SE = 0.76\%$)

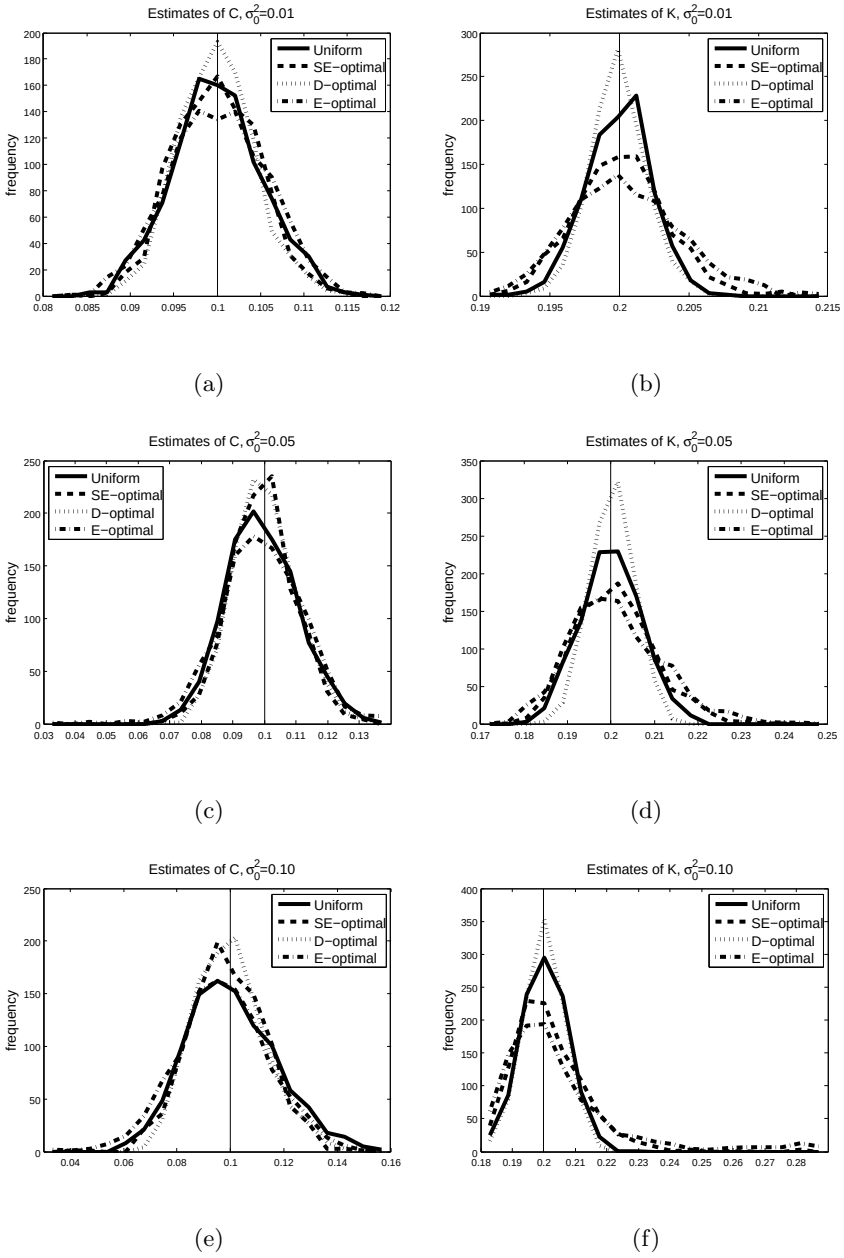


Figure 6.13: Histograms of parameter estimates ($\widehat{C}$ in panels (a),(c),(e), and $\widehat{K}$ in panel (b),(d),(f)) resulting from Monte Carlo simulation. Simulated data was generated with $\sigma_0^2 = 0.01$ (top row), $\sigma_0^2 = 0.05$ (middle row), and $\sigma_0^2 = 0.10$ (bottom row) ($T = 28.28$ and constraint implementation $(C3)$).

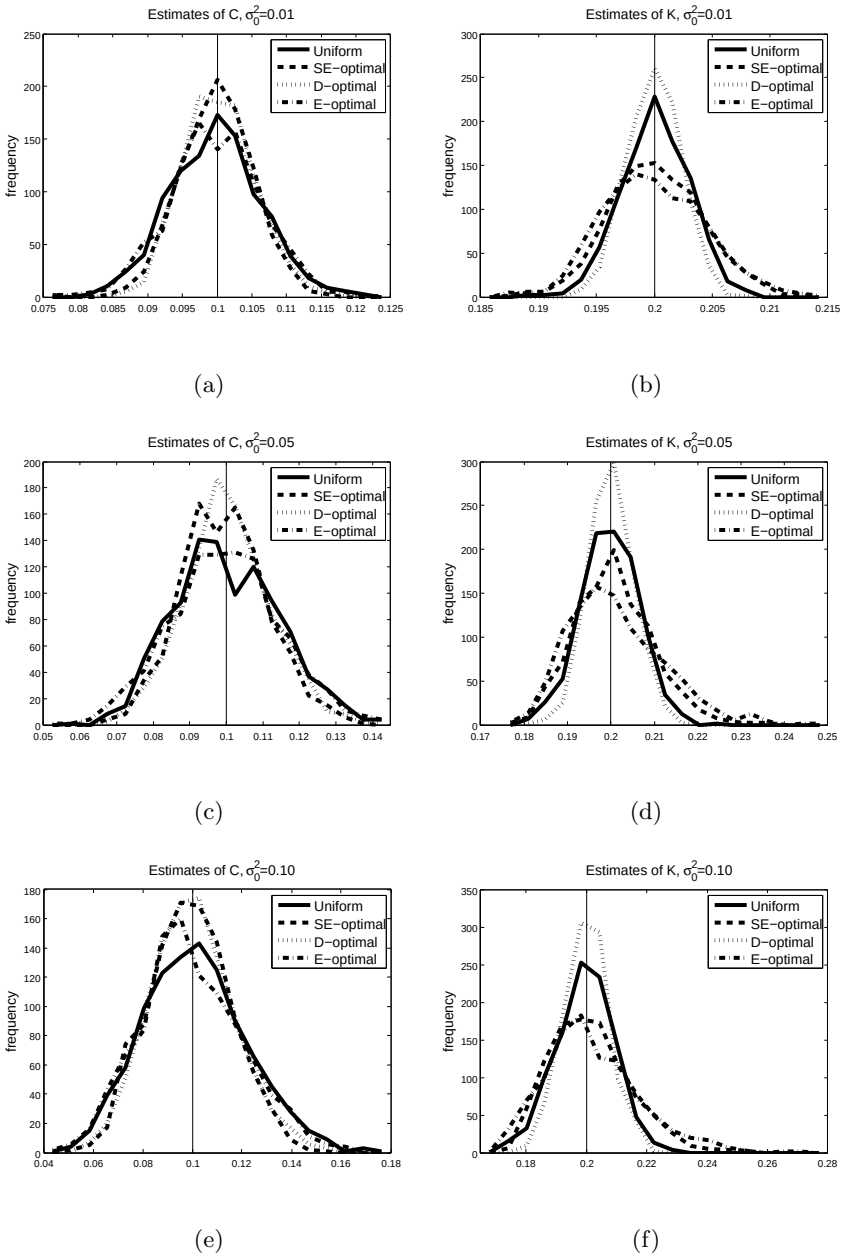


Figure 6.14: Histograms of parameter estimates ($\widehat{C}$ in panels (a),(c),(e), and $\widehat{K}$ in panel (b),(d),(f)) resulting from Monte Carlo simulation. Simulated data was generated with $\sigma_0^2 = 0.01$ (top row), $\sigma_0^2 = 0.05$ (middle row), and $\sigma_0^2 = 0.10$ (bottom row) ($T = 14.14$ and constraint implementation ($C4$)).

Table 6.24: Average estimates (θ_{Avg}) with their standard deviations (θ_{SD}) from $M = 1000$ Monte Carlo trials as well as asymptotic standard errors ($SE(\theta)$). Data was simulated using true values $\theta_0 = (C, K) = (0.1, 0.2)$, fixed parameter values $(x_1, x_2) = (-1, 0.5)$, and noise level $\sigma_0^2 = 0.01$. Simulated data corresponds to optimal design points using $N = 15$, $T = 14.14$ and $T = 28.28$, and constraint implementation ($C4$).

T = 14.14	SE-opt	D-opt	E-opt	Uniform
$\widehat{C}_{Avg}$	0.0999	0.1003	0.0998	0.0998
$\widehat{C}_{SD}$	5.235×10^{-3}	5.136×10^{-3}	6.716×10^{-3}	6.532×10^{-3}
$SE(C)$	5.124×10^{-3}	5.108×10^{-3}	6.361×10^{-3}	6.577×10^{-3}
$\widehat{K}_{Avg}$	0.2001	0.2000	0.2000	0.2001
$\widehat{K}_{SD}$	4.010×10^{-3}	2.289×10^{-3}	4.426×10^{-3}	2.926×10^{-3}
$SE(K)$	3.735×10^{-3}	2.295×10^{-3}	4.162×10^{-3}	2.895×10^{-3}
T = 28.28				
$\widehat{C}_{Avg}$	0.0999	0.1001	0.1001	0.1000
$\widehat{C}_{SD}$	4.434×10^{-3}	4.303×10^{-3}	5.596×10^{-3}	4.906×10^{-3}
$SE(C)$	4.504×10^{-3}	4.391×10^{-3}	5.419×10^{-3}	5.127×10^{-3}
$\widehat{K}_{Avg}$	0.2001	0.1998	0.2002	0.1999
$\widehat{K}_{SD}$	3.354×10^{-3}	1.989×10^{-3}	3.924×10^{-3}	2.371×10^{-3}
$SE(K)$	3.116×10^{-3}	2.003×10^{-3}	3.546×10^{-3}	2.307×10^{-3}

Table 6.25: Average estimates (θ_{Avg}) with their standard deviations (θ_{SD}) from $M = 1000$ Monte Carlo trials as well as asymptotic standard errors ($SE(\theta)$). Data was simulated using true values $\theta_0 = (C, K) = (0.1, 0.2)$, fixed parameter values $(x_1, x_2) = (-1, 0.5)$, and noise level $\sigma_0^2 = 0.05$. Simulated data corresponds to optimal design points using $N = 15$, $T = 14.14$ and $T = 28.28$, and constraint implementation ($C4$).

$\mathbf{T = 14.14}$	SE-opt	D-opt	E-opt	Uniform
$\widehat{C}_{Avg}$	0.0994	0.1004	0.1000	0.1004
$\widehat{C}_{SD}$	1.172×10^{-2}	1.196×10^{-2}	1.498×10^{-2}	1.459×10^{-2}
$SE(C)$	1.146×10^{-2}	1.142×10^{-2}	1.422×10^{-2}	1.471×10^{-2}
$\widehat{K}_{Avg}$	0.2006	0.2000	0.2011	0.1996
$\widehat{K}_{SD}$	9.001×10^{-3}	5.172×10^{-3}	1.100×10^{-2}	6.683×10^{-3}
$SE(K)$	8.351×10^{-3}	5.133×10^{-3}	9.307×10^{-3}	6.473×10^{-3}
$\mathbf{T = 28.28}$				
$\widehat{C}_{Avg}$	0.0997	0.0997	0.0991	0.1000
$\widehat{C}_{SD}$	9.840×10^{-3}	1.019×10^{-2}	1.299×10^{-2}	1.147×10^{-2}
$SE(C)$	1.007×10^{-2}	9.820×10^{-3}	1.212×10^{-2}	1.146×10^{-2}
$\widehat{K}_{Avg}$	0.2007	0.2000	0.2022	0.1999
$\widehat{K}_{SD}$	7.336×10^{-3}	4.351×10^{-3}	1.268×10^{-2}	5.005×10^{-3}
$SE(K)$	6.967×10^{-3}	4.478×10^{-3}	7.928×10^{-3}	5.159×10^{-3}

Table 6.26: Average estimates (θ_{Avg}) with their standard deviations (θ_{SD}) from $M = 1000$ Monte Carlo trials as well as asymptotic standard errors ($SE(\theta)$). Data was simulated using true values $\theta_0 = (C, K) = (0.1, 0.2)$, fixed parameter values $(x_1, x_2) = (-1, 0.5)$, and noise level $\sigma_0^2 = 0.10$. Simulated data corresponds to optimal design points using $N = 15$, $T = 14.14$ and $T = 28.28$, and constraint implementation ($C4$).

T = 14.14	SE-opt	D-opt	E-opt	Uniform
$\widehat{C}_{Avg}$	0.0992	0.1002	0.0996	0.1013
$\widehat{C}_{SD}$	1.635×10^{-2}	1.691×10^{-2}	2.099×10^{-2}	2.089×10^{-2}
$SE(C)$	1.621×10^{-2}	1.615×10^{-2}	2.012×10^{-2}	2.080×10^{-2}
$\widehat{K}_{Avg}$	0.2013	0.2000	0.2020	0.1994
$\widehat{K}_{SD}$	1.326×10^{-2}	7.446×10^{-3}	1.611×10^{-2}	9.633×10^{-3}
$SE(K)$	1.181×10^{-2}	7.259×10^{-3}	1.316×10^{-2}	9.154×10^{-3}
T = 28.28				
$\widehat{C}_{Avg}$	0.0991	0.1004	0.0979	0.0995
$\widehat{C}_{SD}$	1.404×10^{-2}	1.377×10^{-2}	1.752×10^{-2}	1.579×10^{-2}
$SE(C)$	1.424×10^{-2}	1.389×10^{-2}	1.714×10^{-2}	1.621×10^{-2}
$\widehat{K}_{Avg}$	0.2012	0.2002	0.2046	0.2000
$\widehat{K}_{SD}$	1.168×10^{-2}	6.757×10^{-3}	1.960×10^{-2}	7.242×10^{-3}
$SE(K)$	9.853×10^{-3}	6.333×10^{-3}	1.121×10^{-2}	7.295×10^{-3}

Table 6.27: Percent of Confidence Ellipsoids which contain the true parameter values ($T = 14.14$, constraint implementation ($C4$)).

	$\sigma_0^2 = 0.01$	$\sigma_0^2 = 0.05$	$\sigma_0^2 = 0.10$
SE-opt	94.6% ($SE = 0.71\%$)	93.9% ($SE = 0.76\%$)	95.5% ($SE = 0.66\%$)
D-opt	94.8% ($SE = 0.70\%$)	94.4% ($SE = 0.73\%$)	94.3% ($SE = 0.73\%$)
E-opt	94.6% ($SE = 0.71\%$)	94% ($SE = 0.75\%$)	93.1% ($SE = 0.80\%$)
Uniform	95.2% ($SE = 0.68\%$)	96.3% ($SE = 0.60\%$)	94.5% ($SE = 0.72\%$)

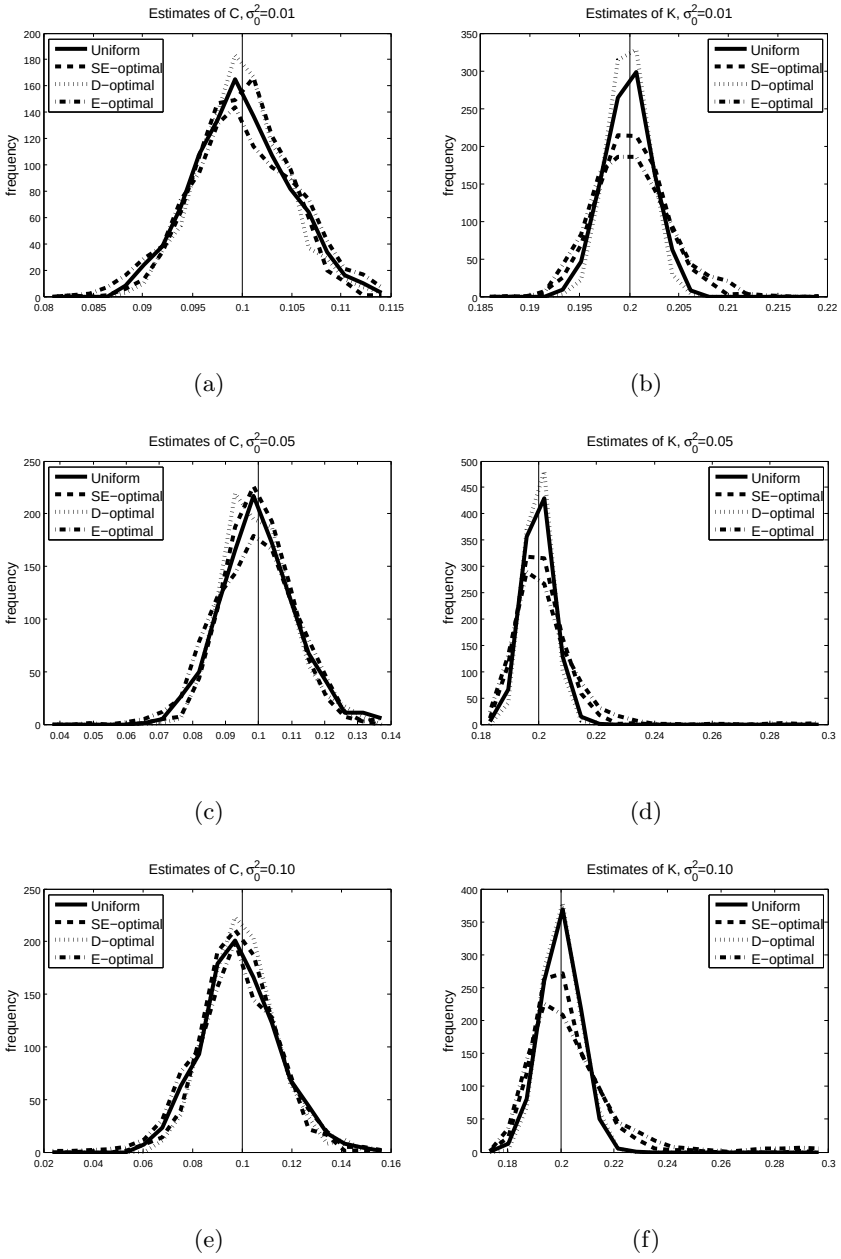


Figure 6.15: Histograms of parameter estimates ($\widehat{C}$ in panels (a),(c),(e), and $\widehat{K}$ in panel (b),(d),(f)) resulting from Monte Carlo simulation. Simulated data was generated with $\sigma_0^2 = 0.01$ (top row), $\sigma_0^2 = 0.05$ (middle row), and $\sigma_0^2 = 0.10$ (bottom row) ($T = 28.28$ and constraint implementation $(C4)$).

Table 6.28: Percent of Confidence Ellipsoids which contain the true parameter values ($T = 28.28$, constraint implementation ($C4$)).

	$\sigma_0^2 = 0.01$	$\sigma_0^2 = 0.05$	$\sigma_0^2 = 0.10$
SE-opt	94.9% ($SE = 0.70\%$)	94.4% ($SE = 0.73\%$)	93.9% ($SE = 0.76\%$)
D-opt	94.2% ($SE = 0.74\%$)	94.4% ($SE = 0.73\%$)	95% ($SE = 0.70\%$)
E-opt	94.7% ($SE = 0.71\%$)	92.5% ($SE = 0.83\%$)	89.9% ($SE = 0.95\%$)
Uniform	96.2% ($SE = 0.60\%$)	94.9% ($SE = 0.70\%$)	94.7% ($SE = 0.71\%$)

6.3.1 Discussion of Harmonic Oscillator Monte Carlo Results

From the harmonic oscillator Monte Carlo results we again observe that as the level of noise in the data decreases so does the variance in the parameter estimates ($\hat{\theta}_{SD}$) and the estimates approach the true values. Examining the average parameter estimates, we find that in all cases the averages are close to the true values $\theta = (C, K) = (0.1, 0.2)$. However in Figures 6.8-6.15 often it appears as though for noise levels $\sigma_0^2 = 0.05$ and 0.10 E-optimal and SE-optimal are right-skewed for the parameter K. Also, often E-optimal appears to be left-skewed for parameter C for the higher levels of noise in the data. This causes the average estimates of K to be slightly larger than the true value for E-optimal and SE-optimal, and the average estimate of C to be slightly smaller than the true value for E-optimal.

Comparing the standard deviations of the parameter estimates ($\hat{\theta}_{SD}$) to the asymptotic standard errors ($SE(\theta)$) we find that they are very close in every case. Comparing the standard deviations of the parameter estimates among the optimal design methods and the uniform mesh often we find that D-optimal had the smallest standard deviations, followed by SE-optimal for parameter C and the uniform mesh for parameter K. In Tables 6.14-6.16 for $T = 28.28$, and 6.24-6.26 D-optimal and SE-optimal had the smallest standard deviations for the parameter C. For parameter K, D-optimal had the smallest standard deviation again followed by the uniform mesh.

For the harmonic oscillator example 95% confidence ellipsoid coverage of the true parameter value is better overall than in the logistic example. This may not be a reasonable comparison since only $p = 2$ parameters are estimated for the harmonic oscillator example, whereas in the logistic example $p = 3$. In this example, no optimal design mesh is consistently better in terms of confidence ellipsoid coverage of the true parameter value. However, in almost every case E-optimal has the worst coverage compared to the other optimal design methods (though still

fairy decent).

Chapter 7

Autocorrelated Data with Applications to Optimal Design Methods

7.1 Introduction

Data taken at time points too close together may be similar and in fact contain little or no new information. The phenomena of serial correlation, or autocorrelation, among repeated measurements is often expected when the time points are taken too close together. Data corresponding to time points sufficiently spaced apart may not be correlated, and thus independence of the measurements is a good assumption in this case.

In this chapter we will discuss how to detect autocorrelation in the data, how to model it, and an algorithm for solving the inverse problem while estimating autocorrelation and model parameters. We will then define the optimal design methods in the context of a statistical model with autocorrelation. We will compare the optimal design methods for the logistic model with autocorrelated data.

7.2 Detection

Autocorrelation, which is summarized in [16], can be detected by examination of the residuals based on the assumption that the errors are independent. In the case of a constant variance statistical model, the assumption of independence is investigated using the OLS weighted residuals:

$$\epsilon_{i,OLS} = \frac{y_i - f(t_i, \hat{\theta}_{OLS})}{\hat{\sigma}_{OLS}}, \text{ for } i = 1, \ldots, n,$$

which are realizations of the random variable

$$\mathcal{E}_i = \frac{Y_i - f(t_i, \theta^0)}{\sigma}, \text{ for } i = 1, \ldots, n,$$

where θ^0 is the true parameter vector, σ is the true standard deviation and Y_i is the observation process.

In the case that you have a non-constant variance model, e.g,

$$\text{Var}(\mathcal{E}_i) = \sigma^2 g^2(\theta, t_i),$$

it can be useful to analyze the GLS weighted residuals

$$\epsilon_{i,GLS} = \frac{y_i - f(t_i, \hat{\theta})}{\hat{\sigma} g(\hat{\theta}, t_i)},$$

which is a realization of the random variable

$$\mathcal{E}_i = \frac{Y_i - f(t_i, \theta^0)}{\sigma g(\theta^0, t_i)}, \text{ for } i = 1, \ldots, n. \tag{7.1}$$

Residual plots are plots of the weighted residuals, for GLS, versus the time points, t_i, for $i = 1, \ldots, n$. Residual plots can be used to test if the variance assumption is accurate. Plotting the weighted residuals should exhibit a random pattern if your variance assumption is correct. To test independence of errors, we examine the lagged weighted residuals. For a fixed l, $1 \leq l \leq n-1$, the lagged weighted residuals of lag l are defined by the random variable

$$\mathcal{E}_{i+l} = \frac{Y_{i+l} - f(t_{i+l}, \theta^0)}{\sigma g(\theta^0, t_{i+l})}.$$

Autocorrelation can be observed visually by plotting ϵ_{i+l} versus ϵ_i, for a fixed lag l. This correlation plot will resemble a symmetric blob centered about zero if the weighted residuals corresponding to different times are uncorrelated. The autocorrelation function is a way to visualize correlation for multiple lags. The autocorrelation function is defined by

$$\rho(l) = \text{Corr}(\mathcal{E}_i, \mathcal{E}_{i+l}) = \frac{\text{Cov}(\mathcal{E}_i, \mathcal{E}_{i+l})}{\sqrt{\text{Var}(\mathcal{E}_i)}\sqrt{\text{Var}(\mathcal{E}_{i+l})}}, \text{ for all } i = 1 \ldots n, \text{ and a fixed } l = 1 \ldots, n-1.$$

The autocorrelation function, $\rho(l)$, is a discrete function of l which can be plotted (for $l = 1 \ldots, n-1$) to examine patterns in autocorrelation. Often we would expect that as the lag increases the autocorrelation would decrease in absolute value to zero.

For a small number of time points, n, it may be difficult to have sufficient supporting evidence

for or against autocorrelation. If data is available for multiple subjects, typically the weighted residuals from all individuals are pooled to detect autocorrelation. This tacitly assumes that the pattern of correlation, if any, would be similar for all individuals. Having more residuals makes a stronger case for the pattern of autocorrelation observed. Since standard errors are not available for the autocorrelation as estimated by the sample correlation between times of lag l, autocorrelation functions based on a small number of residuals should be interpreted with caution.

The autocorrelation function defined above assumes equally spaced points. Something similar can be done for unequally spaced time points. Having more replications at each time point would be useful computing the autocorrelation function with more precision.

7.3 Autocorrelation Models

There are a number of models for autocorrelation. Each model defines a correlation matrix, $\Gamma = \Gamma(\alpha, t)$. A common family of models involves a correlation matrix in the statistical model with the form,

$$\mathrm{Var}(\mathcal{E}) = V = T^{1/2}\Gamma T^{1/2},$$

where T is a diagonal $n \times n$ matrix with either constant or non-constant variance values along the diagonal. A number of specific forms for Γ can be considered [16, 19].

Unstructured Correlation Model

$$\Gamma(\alpha) = \begin{pmatrix} 1 & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1n} \\ \alpha_{21} & 1 & \alpha_{23} & \cdots & \alpha_{2n} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{n,n-1} & 1 \end{pmatrix}$$

Here α is a vector of arbitrary correlation parameters of length $n(n-1)/2$.

Compound Symmetric Model

$$\Gamma(\alpha) = \begin{pmatrix} 1 & \alpha & \cdots & \alpha \\ \alpha & 1 & \cdots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \cdots & \alpha & 1 \end{pmatrix}$$

This model only depends on a single scalar, α.

One-Dependent Model

$$\Gamma(\alpha) = \begin{pmatrix} 1 & \alpha_1 & 0 & \cdots & 0 \\ \alpha_1 & 1 & \alpha_2 & \cdots & 0 \\ 0 & \alpha_2 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \alpha_{n-1} & 1 \end{pmatrix}$$

A special case of this model is to assume that $\alpha_j \equiv \alpha$ for all j:

$$\Gamma(\alpha) = \begin{pmatrix} 1 & \alpha & 0 & \cdots & 0 \\ \alpha & 1 & \alpha & \cdots & 0 \\ 0 & \alpha & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \alpha & 1 \end{pmatrix}$$

Autoregressive Model of Order 1: AR(1)

$$\Gamma(\alpha) = \begin{pmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{n-1} \\ \alpha & 1 & \alpha & \alpha^2 & \cdots \\ \alpha^2 & \alpha & 1 & \alpha & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha^{n-1} & \cdots & \alpha^2 & \alpha & 1 \end{pmatrix},$$

where $0 \leq \alpha \leq 1$. This correlation model describes the amount of autocorrelation decreasing as the amount of time between points increases, for equally spaced time points.

Exponential Correlation Model

The exponential correlation model is represented by the autocorrelation function

$$\rho(l) = \exp(-\alpha l), \alpha > 0,$$

implying

$$\text{Corr}(Y_i, Y_j) = \exp(-\alpha |t_i - t_j|),$$

$$\Gamma(\alpha,t)=\begin{pmatrix} 1 & (\alpha^*)^{|t_1-t_2|} & (\alpha^*)^{|t_1-t_3|} & \cdots & (\alpha^*)^{|t_1-t_{n-1}|} \\ & 1 & (\alpha^*)^{|t_2-t_3|} & \cdots & \vdots \\ & & \ddots & \vdots & \vdots \\ & & & 1 & (\alpha^*)^{|t_{n-1}-t_{n-2}|} \\ & & & & 1 \end{pmatrix},$$

where $\alpha^* = \exp(-\alpha)$, and $\Gamma(\alpha)$ is a symmetric matrix. The exponential correlation model is a generalization of the AR(1) model for unequally-spaced times.

Gaussian Correlation Model

The exponential correlation model is represented by the autocorrelation function

$$\rho(l) = \exp(-\alpha l^2), \alpha > 0,$$

implying

$$\mathrm{Corr}(Y_i, Y_j) = \exp(-\alpha(t_i - t_j)^2).$$

7.4 Estimation of Parameters in Nonlinear Mathematical Models using Statistical Models with Autocorrelation

When one adds a correlation model to the statistical model the number of variance parameters to be estimated increases $\gamma = (\sigma, \alpha)$, or more generally $\gamma = (\vec{\sigma}, \vec{\alpha}) = (\sigma_1, \ldots, \sigma_n, \alpha_1, \ldots, \alpha_{n-1})$. Approaches to this type of estimation procedure are possible using maximum likelihood, or methods similar to generalized least squares (GLS). The latter will be described here. The method will be given for a generic correlation model matrix, Γ, though later we implement it for the exponential correlation model.

The model is given by

$$\begin{aligned} Y_i &= f(t_i, \theta) + \tilde{\mathcal{E}}_i, \\ \mathrm{E}(\tilde{\mathcal{E}}) &= 0, \\ \mathrm{Var}(\tilde{\mathcal{E}}) &= V = T^{1/2}\Gamma T^{1/2}, \end{aligned}$$

where $\tilde{\mathcal{E}}$ is the random variable representing the observation error. Note that our unknown model parameters are given by the vector θ and our unknown variance parameters are given by $\gamma = (\sigma, \alpha)$ where both σ and α may be vectors. Let q represent the total number of variance

parameters to be estimated which are contained in the vector γ. Below we describe an algorithm to solve for model and variance parameters. This algorithm is one in a family of methods solving this type of inverse problem using General Estimating Equations [18, 24, 26, 27, 34].

Three-step Algorithm to solve the Linear General Estimating Equations (GEE-1), [16]:

(i) Estimate θ by $\hat{\theta}^{(0)}$ using the OLS estimator (assuming independence), and set $k = 0$.

(ii) Estimate the variance parameters $\gamma = (\sigma, \alpha)$, while taking $\hat{\theta}^{(k)}$ to be fixed, and store as $\hat{\gamma}^{(k)} = (\hat{\sigma}^{(k)}, \hat{\alpha}^{(k)})$. To do this we start by defining the elements of $V = V(\gamma) = \mathrm{Var}(\tilde{\mathcal{E}})$:

$$\begin{aligned} \mathrm{E}\left[\{Y_i - f(t_i, \theta)\}^2\right] &= V_{ii}(\gamma), \\ \mathrm{E}\left[\{Y_i - f(t_i, \theta)\}\{Y_j - f(t_j, \theta)\}\right] &= V_{ij}(\gamma), \end{aligned}$$

which can be stored in the $(n(n+1)/2)$-vector, v, not repeating symmetric elements. The vector will be of length $n(n-1)/2$ if a constant variance model with known σ_0^2, and correlated errors are assumed. Note that the matrix V follows from the assumptions made about our statistical model. For symmetric matrices A (such as V), let us define the following notation for $\mathrm{vech}(A)$, for example in the 3×3 case,

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}, \quad \mathrm{vech}(A) = \begin{pmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{22} \\ a_{23} \\ a_{33} \end{pmatrix}.$$

This vector can be obtained by going along the rows from the diagonal to the end, starting from the top row and working your way down. Let $v(\gamma) = \mathrm{vech}(V(\gamma))$. Also define a symmetric matrix U with components

$$U_{ij} = \{Y_i - f(t_i, \theta)\}\{Y_j - f(t_j, \theta)\} = \tilde{\mathcal{E}}_i \tilde{\mathcal{E}}_j.$$

Note that $\mathrm{E}(U) = V(\gamma)$. The realizations of U are given by the $(n(n+1)/2)$-vector u, formed as described above by the vech() function. The elements of the vector u are based on the corresponding residuals:

$$u_{ij} = (y_i - f(t_i, \hat{\theta}))(y_j - f(t_i, \hat{\theta})).$$

Define the gradient matrix of the vector of variance components v:

$$H(\gamma) = \partial/\partial\gamma[v(\gamma)].$$

Note that $H(\gamma)$ will have $n(n+1)/2$ rows and q columns, where q is the number of variance or correlation parameters being estimated. Define the $(n(n+1)/2) \times (n(n+1)/2)$ covariance matrix of the vector vech(U):

$$Z(\gamma) \quad = \quad \text{Var}(\text{vech}(U)),$$

with off-diagonal elements of Z defined by

$$\text{Cov}(U_{ij}, U_{lk}) = \text{E}(U_{ij}U_{lk}) - \text{E}(U_{ij})\text{E}(U_{lk}),$$

for example, where $1 \leq i, j, k, l \leq n$. In order to define Z we must make some assumptions. Note that our weighted residuals, defined in equation (7.1), have

$$\text{E}(\mathcal{E}_i) = 0, \ \text{Var}(\mathcal{E}_i) = 1, \ \text{for all } i = 1, \ldots, n.$$

Assuming our correlation assumptions are correct, the weighted residuals are correlated with $\Gamma(\alpha)$ correlation matrix. We can also derive the following quantities,

$$\begin{aligned}
v_{jj} &= \sigma^2 g^2(\theta, \gamma, t_j)\text{E}(\mathcal{E}_j^2) = \sigma^2 g^2(\theta, \gamma, t_j), \\
v_{jk} &= \sigma^2 g(\theta, \gamma, t_j) g(\theta, \gamma, t_k)\text{E}(\mathcal{E}_j\mathcal{E}_k) = \sigma^2 g(\theta, \gamma, t_j) g(\theta, \gamma, t_k)\Gamma_{j,k}, \\
U_{jj} &= \sigma^2 g^2(\theta, \gamma, t_j)\mathcal{E}_j^2, \\
U_{jk} &= \sigma^2 g(\theta, \gamma, t_j) g(\theta, \gamma, t_k)\mathcal{E}_j\mathcal{E}_k, \\
\text{Cov}(U_{ij}, U_{lk}) &= \sigma^4 g(\theta, \gamma, t_i) g(\theta, \gamma, t_j) g(\theta, \gamma, t_k) g(\theta, \gamma, t_l) \\
&\quad \times \{\text{E}(\mathcal{E}_i\mathcal{E}_j\mathcal{E}_k\mathcal{E}_l) - \text{E}(\mathcal{E}_i\mathcal{E}_j)\text{E}(\mathcal{E}_k\mathcal{E}_l)\}, \\
&= \sigma^4 g(\theta, \gamma, t_i) g(\theta, \gamma, t_j) g(\theta, \gamma, t_k) g(\theta, \gamma, t_l) \{\text{E}(\mathcal{E}_i\mathcal{E}_j\mathcal{E}_k\mathcal{E}_l) - \Gamma_{i,j}\Gamma_{k,l}\},
\end{aligned}$$

for $i, j, k, l = 1, \ldots, n$ and $i \neq j \neq k \neq l$. The assumptions we must make come in defining an approximation for $\text{E}(\mathcal{E}_i\mathcal{E}_j\mathcal{E}_k\mathcal{E}_l)$. Otherwise we are unable to fully define the Z matrix. It is common to make the Gaussian working assumption. This assumes that the Y_i are normally distributed with mean and variance corresponding to the mathematical and statistical model, ($\text{E}(Y) = f(t, \theta^0)$ and $\text{Var}(Y) = V$). Equivalently this assumes the errors are also normally distributed with $\text{E}(\tilde{\mathcal{E}}) = 0$ and $\text{Var}(\tilde{\mathcal{E}}) = V$. Under this assumption we

have:

$$\begin{aligned}
\mathrm{Cov}(U_{ij}, U_{lk}) &= \sigma^4 g(\theta,\gamma,t_i)g(\theta,\gamma,t_j)g(\theta,\gamma,t_k)g(\theta,\gamma,t_l) \\
&\quad \times\{\mathrm{E}(\mathcal{E}_i\mathcal{E}_k)\mathrm{E}(\mathcal{E}_j\mathcal{E}_l) + \mathrm{E}(\mathcal{E}_i\mathcal{E}_l)\mathrm{E}(\mathcal{E}_j\mathcal{E}_k)\}, \\
&= \sigma^4 g(\theta,\gamma,t_i)g(\theta,\gamma,t_j)g(\theta,\gamma,t_k)g(\theta,\gamma,t_l)\{\Gamma_{i,k}\Gamma_{j,l} + \Gamma_{i,l}\Gamma_{j,k}\}, \\
\mathrm{Var}(U_{jj}) &= 2\sigma^4 g^4(\theta,\gamma,t_j).
\end{aligned}$$

This assumption allows for the matrix $Z(\gamma) = \mathrm{Var}(\mathrm{vech}(U))$ to be clearly defined. There are other ways to go about this, but this approach is the most common.
Now that we have defined the matrices H and Z and the vectors v and $\mathrm{vech}(U)$, we can give the estimating equation for estimating $\gamma = (\sigma, \alpha)$ while holding $\hat{\theta}^{(k)}$ fixed:

$$H(\gamma)^T Z(\gamma)^{-1}\{\mathrm{vech}(U) - v(\gamma)\} = 0.$$

Note that H is a $(n(n+1)/2) \times q$ matrix, Z is a $(n(n+1)/2) \times (n(n+1)/2)$ matrix and $\mathrm{vech}(U)$ and v are $(n(n+1)/2)$-vectors, so the equation above is equivalent to a system of q equations with q unknowns since q is the number of variance parameters being estimated in γ.
Store the estimate of γ as $\hat{\gamma}^{(k)} = (\hat{\sigma}^{(k)}, \hat{\alpha}^{(k)})$.

(iii) Re-estimate θ by solving for θ in

$$(\nabla_\theta f(\vec{t},\theta)))^{\mathrm{T}} V^{-1}(\hat{\sigma}^{(k)}, \hat{\alpha}^{(k)})(\vec{y} - f(\vec{t},\theta)) = 0,$$

where $\vec{t} = \{t_i\}_{i=1}^n$, $\vec{y} = \{y_i\}_{i=1}^n$, and $\nabla_\theta f(\vec{t},\theta)$ is the $n \times p$ sensitivity matrix. Note that V is an $n \times n$ matrix and $(\vec{y} - f(\vec{t},\theta))$ is a n-vector, making the equation above a system of p equations with p unknowns: θ.
Store the estimate of θ as $\hat{\theta}^{(k)}$. Set $k = k + 1$ and go to (ii).

Iterate between (ii) and (iii) to some convergence criteria.

The estimating equations given in steps (ii) and (iii) are nonlinear functions where solving for the parameters involves finding the function's root. Parameter estimates can be obtained from the estimating equations using the Gauss-Newton algorithm, which is implemented in MATLAB by the function *fsolve*. The implementation of the Gauss-Newton algorithm used by MATLAB's *fsolve* is called the Trust-region Dogleg implementation.

7.5 Application to Optimal Design Methods

An extension of optimal design methods to this case where there is correlation between the time points is possible due to the corresponding definition of the FIM [30, 33],

$$F = \chi^T(\theta)V^{-1}(\gamma)\chi(\theta).$$

Recall that χ is the $n \times p$ sensitivity matrix ($\chi_{j,k} = \partial f(t_j, \theta)/\partial\theta_k$, for $j = 1, \ldots n, k = 1, \ldots p$), and $V(\gamma) = T^{1/2}(\sigma)\Gamma(\alpha)T^{1/2}(\sigma)$ is the $n \times n$ matrix where T is a $n \times n$ diagonal matrix containing the constant or non-constant variances (σ) along the diagonal, and $\Gamma = \Gamma(\alpha)$ is the $n \times n$ matrix from our assumed correlation model.

Optimal design methods can be related to the FIM. It is possible to find the optimal mesh while optimizing on any set of the parameters, θ, and including the variance parameters $\gamma = (\sigma, \alpha)$ as they appear in the V matrix.

7.5.1 Methodology for the Comparison of Optimal Design Methods with Autocorrelation

Optimal design methods are compared based on parameter estimates for (θ, γ), and approximate asymptotic standard errors. The approximate asymptotic standard errors for θ are defined as follows

$$\begin{aligned} \hat{\Sigma} &= [\chi^T(\hat{\theta})V^{-1}(\hat{\gamma})\chi(\hat{\theta})]^{-1}, \\ SE_k(\hat{\theta}) &= \sqrt{(\hat{\Sigma})_{kk}}, \quad k = 1, \ldots, p, \end{aligned} \tag{7.2}$$

where p is the number of model parameters in the vector θ. Standard errors are not available for the variance parameters, $\gamma = (\sigma^2, \alpha)$.

In addition, optimal design methods are compared based on average parameter estimates, for (θ, α), with their standard deviations from $M = 1000$ Monte Carlo trials.

Data was simulated according to our statistical model with realizations

$$y_i = f(t_i, \theta) + \epsilon_i, \quad i = 1, \ldots, n,$$

which correspond to the optimal design mesh points $\tau^* = \{t_i\}$, $i = 1, \ldots, n$. Our autocorrelated errors are generated using a multivariate normal

$$\vec{\mathcal{E}} \sim \mathcal{N}(\vec{0}, V),$$

where $V = V(\gamma) = T^{1/2}(\sigma)\Gamma(\alpha)T^{1/2}(\sigma)$ is an $n \times n$ matrix with constant variance σ^2 along the

diagonal, where $\Gamma(\alpha)$ is modeled using the exponential correlation model which also depends on the optimal time points:

$$\Gamma_{i,j} = \text{Corr}(Y_i, Y_j) = \exp(-\alpha|t_i - t_j|) = (\alpha^*)^{|t_i - t_j|},$$

where $\alpha^* = \exp(-\alpha)$. The resulting $\vec{\epsilon}$ is an n-vector.

The inverse problem was solved using the GEE-1 algorithm described above. Our statistical model implies that the variance function in the GEE-1 algorithm description reduces to $g(\theta, \gamma, t) = 1$. Also, the Gaussian working assumption in the definition of the matrix Z is reasonable since our simulated autocorrelated errors are generated using a multivariate normal.

Most of the time MATLAB's *fsolve* converged to a solution when solving the estimating equations within the GEE-1 algorithm. However, in less than 8% of the Monte Carlo trials the algorithm failed to converge to a solution. This error occurred within the Gauss-Newton algorithm as it attempted to solve an estimating equation within the GEE-1 algorithm at step (ii) or (iii). During our Monte Carlo simulation, data sets were not used if the algorithm failed to converge, and more data was simulated until we obtained $M = 1000$ successful solutions (sets of parameter estimates) from the GEE-1 algorithm.

In practice, one would choose a different initial guess of the parameter values if the GEE-1 algorithm failed to converge rather than throwing out the data set. Since this analysis consists of Monte Carlo analysis of simulated data with $M = 1000$, it was feasible to use the same initial guess for all trials and generate new data if the algorithm failed to converge for a particular data set.

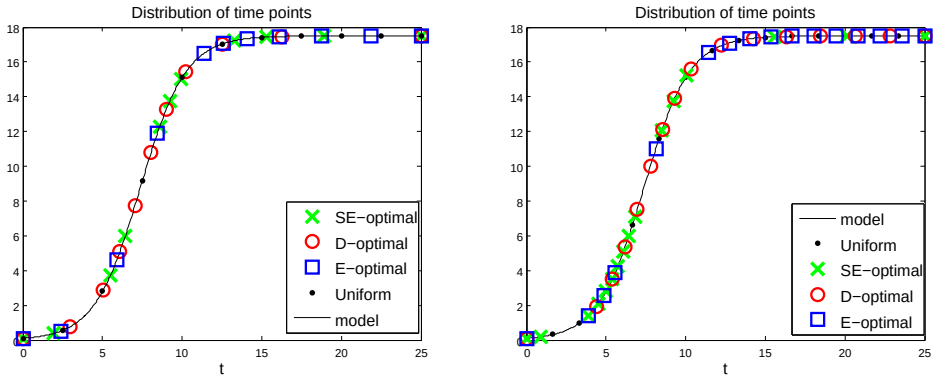
7.5.2 Results of Optimal Design Methods with Autocorrelation

Optimal design points were obtained corresponding to the same statistical model assumed for our simulated data with autocorrelation, and the following assumed parameter values $(\theta, \gamma) = (K, r, x_0, \alpha^*, \sigma^2) = (17.5, 0.7, 0.1, 0.67, 0.10)$. Figure 7.1 contains the optimal time points using constraint implementation $(C3)$ for $N = 10$ (panel (a)) and $N = 15$ (panel (b)).

For all inverse problems using the GEE-1 algorithm our initial parameter guess was $(\theta^0, \gamma^0) = (1.4\theta_0, \gamma_0)$, where $(\theta_0, \gamma_0) = (K, r, x_0, \alpha^*) = (17.5, 0.7, 0.1, 0.67)$ are the true parameter values. Note that the variance parameter σ^2 was held fixed ($\sigma^2 = 0.10$) and was not estimated. Our convergence criteria for the GEE-1 algorithm is given by

$$\left(|\hat{K}^{(k+1)} - \hat{K}^{(k)}|, |\hat{r}^{(k+1)} - \hat{r}^{(k)}|, |\hat{x}_0^{(k+1)} - x_0^{(k)}|, |\hat{\alpha}^{*(k+1)} - \hat{\alpha}^{*(k)}|\right)^T \leq (10^{-3}, 10^{-5}, 10^{-5}, 10^{-5})^T,$$

where in this case the $\leq$ is component-wise in the 4-vector. Since we are estimating a total of four parameters, we chose to compare the method based on the optimal design points with



(a) (b)

Figure 7.1: The distribution of optimal time points and uniform sampling time points plotted on the logistic curve. Optimal times points assuming an exponential autocorrelation model were obtained using SolvOpt, with $N = 10$ (panel (a)) and $N = 15$ (panel (b)).

$N = 15$ points (Fig. 7.1(b)).

We compare the optimal design methods based on the parameter estimates from the GEE-1 algorithm and their approximate asymptotic standard errors (7.2). Table 7.1 contains the parameter estimates and approximate asymptotic standard errors (7.2) from a set of autocorrelated data corresponding to the optimal time points (with $N = 15$ and constraint implementation $(C3)$).

We also compared the optimal design methods based on $M = 1000$ Monte Carlo trials. Histograms of the parameter estimates obtained by the GEE-1 algorithm are given in Fig. 7.2. Each subfigure corresponds to a specific parameter. Within each subfigure are histograms corresponding to the three different optimal design methods (SE-optimal, D-optimal, E-optimal) and the uniform mesh. Table 7.2 contains the average and standard deviation of the $M = 1000$ parameter estimates.

7.5.3 Discussion of Optimal Design Methods with Autocorrelation

Examining the results from the inverse problem using asymptotic theory in Table 7.1, we find that the parameter estimates closest to the true values resulted from D-optimal followed by SE-optimal for K, and the uniform mesh followed by E-optimal for r and x_0. E-optimal had the closest estimate of the autocorrelation parameter α, followed by D-optimal. The uniform mesh and SE-optimal estimated α to be very small. Comparing the optimal design methods based

Table 7.1: Estimates and standard estimates (7.2) from the asymptotic theory resulting from different optimal design methods (as well as the uniform mesh) for data simulated with $\theta_0 = (K, r, x_0) = (17.5, 0.7, 0.1)$ and $\gamma_0 = (\alpha^*, \sigma^2) = (0.67, 0.10)$ and $N = 15$, optimization with constraint implementation $(C3)$. Parameters (K, r, x_0, α^*) were estimated using the GEE-1 algorithm.

	SE-opt	D-opt	E-opt	Uniform
$\hat{K}$	17.3630	17.4104	17.7273	17.3436
$SE(\hat{K})$	5.671×10^{-2}	1.411×10^{-1}	1.918×10^{-1}	1.182×10^{-1}
$\hat{r}$	0.7971	0.6555	0.6880	0.6966
$SE(\hat{r})$	3.316×10^{-3}	3.002×10^{-2}	4.991×10^{-2}	6.269×10^{-3}
$\hat{x}_0$	0.0581	0.1211	0.1119	0.1088
$SE(\hat{x}_0)$	1.511×10^{-3}	2.759×10^{-2}	4.099×10^{-2}	5.584×10^{-3}
$\hat{\alpha}^*$	3.178×10^{-4}	0.7297	0.6921	4.793×10^{-3}

Table 7.2: Average estimates ($\hat{\theta}_{Avg}$) with their standard deviations ($\hat{\theta}_{SD}$) from $M = 1000$ Monte Carlo trials. Autocorrelated data corresponding to the optimal time points was simulated using true values $(K, r, x_0, \alpha^*) = (17.5, 0.7, 0.1, 0.67)$, $N = 15$, and 10% noise ($\sigma_0^2 = 0.10$). Parameters (K, r, x_0, α^*) were estimated using the GEE-1 algorithm.

	SE-opt	D-opt	E-opt	Uniform
$\hat{K}_{avg}$	17.4920	17.4997	17.4978	17.5077
$\hat{K}_{SD}$	2.042×10^{-1}	1.801×10^{-1}	1.746×10^{-1}	1.852×10^{-1}
$\hat{r}_{avg}$	0.7017	0.7021	0.7024	0.6999
$\hat{r}_{SD}$	3.447×10^{-2}	3.463×10^{-2}	3.951×10^{-2}	3.662×10^{-2}
$\hat{x}_{0,avg}$	0.1019	0.1017	0.1021	0.1042
$\hat{x}_{0,SD}$	2.661×10^{-2}	2.621×10^{-2}	3.091×10^{-2}	2.846×10^{-2}
$\hat{\alpha}^*_{avg}$	0.2504	0.3522	0.4216	0.4784
$\hat{\alpha}^*_{SD}$	2.812×10^{-1}	3.424×10^{-1}	2.834×10^{-1}	2.784×10^{-1}

on their approximate asymptotic standard errors we find that SE-optimal had the smallest standard errors for (K, r, x_0), followed by the uniform mesh.

Parameter estimates from the Monte Carlo analysis (M=1000) are visualized in histograms in Figure 7.2. For parameters K (in panel (a)), r (in panel (b)), x_0 (in panel (c)), the histograms appear similar from the three optimal design methods and the uniform mesh and all are centered

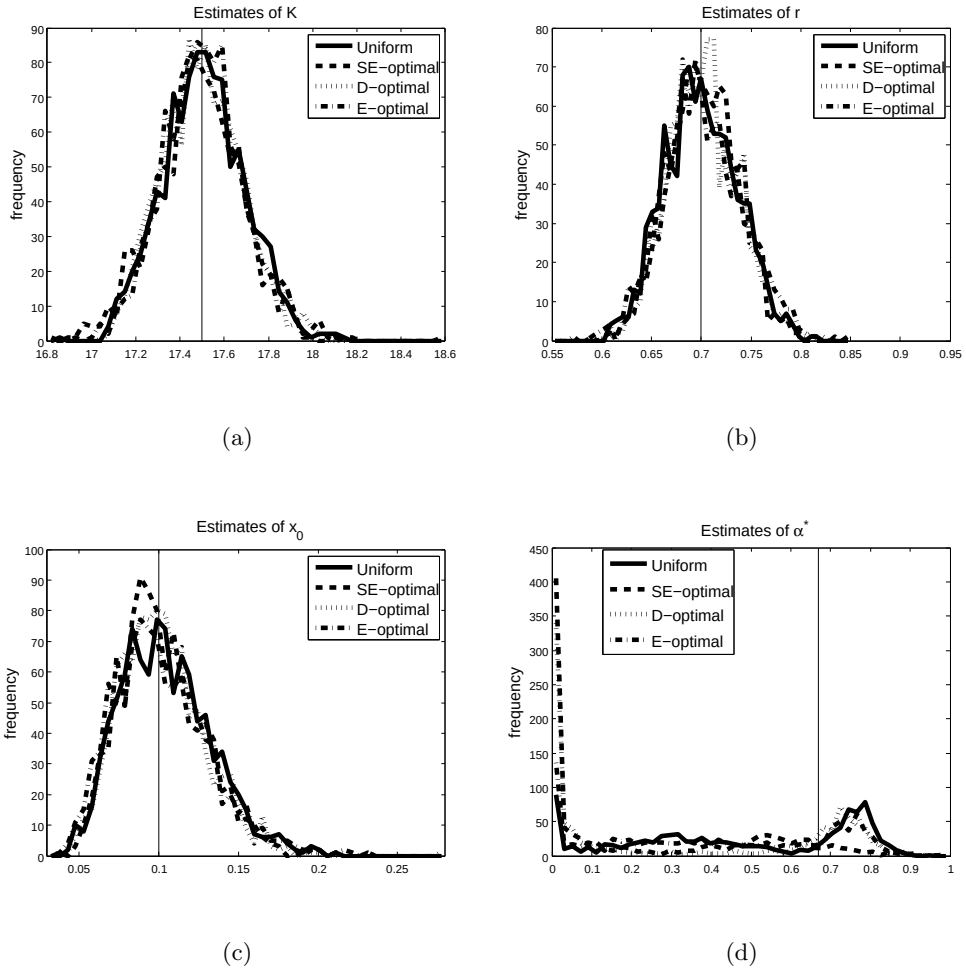


Figure 7.2: Histograms of parameter estimates (K in panel (a), r in panel (b), x_0 in panel (c), α^* in panel (d)) resulting from Monte Carlo simulation with M = 1000. Different histograms within each subgure represent results from different optimal design methods as well as from the uniform mesh. Simulated autocorrelated data was generated with N = 15, $\alpha^* = 0.67$, 10% noise ($\sigma_0^2 = 0.10$), and true parameter values $\theta_0 = (K, r, x_0) = (17.5, 0.7, 0.1)$.

about the true values. The histogram for α^* (in panel (d)) is more complicated. Both D-optimal and SE-optimal have a spike of estimates close to zero, with approximate frequency of 400 out of $M = 1000$. The uniform mesh and E-optimal have a small spike near zero, with approximate frequency of 125. There is another peak that the uniform mesh, D-optimal and E-optimal have around $\alpha^* = 0.75$ (frequency ≈ 150) Recall that the true value for α^* was 0.67. All the optimal design methods and the uniform mesh estimated α to be dispersed between 0 and 0.9, just with lower frequency than the spikes already discussed.

Table 7.2 summarizes the results from the MC trials with the average and standard deviation of the parameter estimates. Comparing the methods based on how close their average parameter estimates are to the true values, we find that for K D-optimal followed by E-optimal was the closest, for r the closest average estimates where from the uniform mesh followed by SE-optimal, for x_0 D-optimal was closest followed by SE-optimal, and for α^* the uniform mesh was closest followed by E-optimal. In general, all parameter estimates were close to the true values for K, r, and x_0 for all the optimal design methods and the uniform mesh.

Comparing the methods based on the standard deviations of the parameter estimates (Table 7.2), for K E-optimal design has the smallest standard deviation followed by D-optimal, for r the smallest standard deviation was from SE-optimal followed by D-optimal, for x_0 D-optimal followed by SE-optimal has the smallest standard deviation, and for α^* the uniform mesh has the smallest standard deviation followed by SE-optimal and E-optimal. For each of the optimal design methods, as well as for the uniform mesh, the standard deviations for α^* were larger than the standard deviation for K even though the true value of α^* is two orders of magnitude smaller than the true value K.

Comparing the approximate asymptotic standard errors (Table 7.1) to the Monte Carlo trial standard deviations (Table 7.2), we find that they are similar for D-optimal and E-optimal. For SE-optimal and the uniform mesh, often the approximate asymptotic standard errors are smaller. This maybe explained by the fact that SE-optimal and the uniform mesh approximated the autocorrelation parameter α^* to be close to zero in Table 7.1. Autocorrelation detected in the data (as is the case for D and E-optimal in Table 7.1) would result in higher standard errors compared to the case where the autocorrelation parameter is estimated to be small.

In conclusion, each optimal design method results in good estimates for the model parameters (K, r, x_0) with no method being favorable over the others. Parameter estimation for the correlation parameter α^* was not as good. The standard deviations in α^* were large for all optimal design methods. SE-optimal and D-optimal tended to estimate α^* close to zero, compared to E-optimal and the uniform mesh. The uniform mesh was the best for estimating α^* in terms of its average estimate and its standard deviation, followed by E-optimal. While the uniform mesh appears favorable for estimating α^*, we find a wide range of estimates for α^* from the $M = 1000$ Monte Carlo from the uniform mesh as is also the case with the optimal design methods.

7.6 Using the Incorrect Assumptions on the Correlation in the Errors

In the previous section we observed that even with $\alpha^* = 0.67$ autocorrelation present in the data, often α^* was estimated to be close to zero (i.e. suggesting independence of the errors).

In this section we will explore the consequences of incorrectly assuming that the data is independent when in fact the data is autocorrelated. We examine this question with autocorrelated data with various levels of autocorrelation: $\alpha^* = 10^{-4}, 0.225, 0.45, 0.675, 0.9$, respectively, true model parameters $\theta_0 = (K, r, x_0) = (17.5, 0.7, 0.1)$ and constant variance $\sigma^2 = 0.10$.

Assuming that the data is independent, we compare the optimal design methods based on their optimal design points assuming the data is i.i.d with constant variance and $N = 15$ and constraint implementation $(C2)$; see Figure 5.3 for the optimal design points. We use Ordinary Least Squares (OLS) to solve the inverse problem. Note that we only estimate the model parameters $\theta = (K, r, x_0)$. We also compute the approximate asymptotic standard errors according to our assumption of independent errors with constant variance:

$$\begin{aligned} \hat{\sigma}^2_{OLS} &= \frac{1}{N-p}\sum_{j=1}^{N}[y_j - f(t_j, \hat{\theta})]^2, \\ \hat{\Sigma} &= \hat{\sigma}^2_{OLS}[\chi^T(\hat{\theta})\chi(\hat{\theta})]^{-1}, \\ SE_k(\hat{\theta}) &= \sqrt{(\hat{\Sigma})_{kk}}, \;\; k = 1, \ldots, p, \end{aligned} \tag{7.3}$$

where p is the number of parameters we are estimating using OLS.

For comparison purposes for a given data set with autocorrelation (corresponding to optimal design points assuming independence), parameters were estimated using both OLS and the GEE-1 algorithm. Using GEE-1, under the same conditions as in the previous section, we estimate parameters $(\theta, \gamma) = (K, r, x_0, \alpha^*)$, and compute approximate standard errors (7.2).

To get an idea of the typical behavior of the inverse problem approaches (OLS and GEE-1) as well as the optimal design methods for each of the five levels of autocorrelation $\alpha^* = 10^{-4}, 0.225, 0.45, 0.675,\ 0.9$, respectively, we repeat the analysis for $M = 250$ Monte Carlo trials. If GEE-1 fails to converge for a given data set, that data set is not used for either GEE-1 or OLS. New autocorrelated data sets are generated until each level of autocorrelation corresponds to $M = 250$ successful sets of OLS and GEE-1 of parameter estimates. Comparisons are based on average parameter estimates $(\hat{\theta}_{avg})$, standard deviation of parameter estimates $(\hat{\theta}_{SD})$, and average approximate standard errors $(SE(\hat{\theta})_{avg})$ (computed differently for OLS and GEE-1).

7.6.1 Results of using the Incorrect Assumptions on the Correlation in the Errors

Results from the inverse problem using OLS (incorrect assumption) are given in Tables 7.3, 7.5, 7.7, 7.9 and 7.11, corresponding to the five different levels of autocorrelation in the data, $\alpha^* = 10^{-4}, 0.225, 0.45, 0.675, 0.9$, respectively. Results from the inverse problem using GEE-1 (correct assumption) are given in Tables 7.4, 7.6, 7.8, 7.10 and 7.12.

7.6.2 Discussion of using the Incorrect Assumption on the Correlation in the Errors

In what follows is a discussion of the tables of results from OLS (incorrect assumption) and GEE-1 (correct assumption) for each level of autocorrelation in the data.

Discussion for $\alpha^* = 10^{-4}$:

We examine the results from OLS in Table 7.3 corresponding to $M = 250$ Monte Carlo trials of data with $\alpha^* = 10^{-4}$ autocorrelation. Comparing the optimal design methods based on which average OLS parameter estimates are closest to the true values, we find that for K and r the uniform mesh is closest followed by SE-optimal (for K) and D-optimal (for r), for x_0 the closest was from D-optimal followed by SE-optimal. In this case, the standard deviations (SD) of the estimates were very close to the average approximate asymptotic standard errors (SE) for all the optimal design methods. The smallest SD's and SE's for parameter K came from E-optimal followed by D-optimal. For r and x_0 the smallest SD's and SE's came from SE-optimal followed by D-optimal.

The results from GEE-1 for the Monte Carlo trials with autocorrelated data with $\alpha^* = 10^{-4}$ are given in Table 7.4. Comparing the optimal design methods based on which has average parameter estimates closest to the true values, for K, r, and α^* the uniform mesh was closest followed by D-optimal. For x_0, SE-optimal was closest to the true value followed by D-optimal. Though the average approximate asymptotic standard errors (SE) are very close to the standard deviations (SD) of the parameter estimates, the results are different in terms of which method had the smallest SD or SE. Comparing the methods based on which had the smallest SD, we find that D-optimal is smallest for (K, r, x_0) followed by the uniform mesh (for K) and SE-optimal (for r and x_0). E-optimal, followed by the uniform mesh, had the smallest SD for α^*. Comparing the methods based on which has the smallest SE, we find for K the uniform mesh followed by D-optimal is the smallest, and for r and x_0 D-optimal is the smallest followed by the uniform mesh.

Comparing the results from OLS (Table 7.3) to the results from GEE-1 (Table 7.4), we find that the parameter estimates and standard deviations for (K, r, x_0) are very similar from the two different inverse problem methods. The average approximate standard errors are on the

Table 7.3: Average estimates obtained from OLS ($\hat{\theta}_{Avg}$) with their standard deviations ($\hat{\theta}_{SD}$) and average asymptotic standard errors (7.3) ($SE(\hat{\theta})_{avg}$) from $M = 250$ Monte Carlo trials. Autocorrelated data corresponding to the optimal time points was simulated using true values $(K, r, x_0) = (17.5, 0.7, 0.1)$, $N = 15$, 10% noise ($\sigma_0^2 = 0.10$), and $\alpha^* = 10^{-4}$ autocorrelation.

	SE-opt	D-opt	E-opt	Uniform
$\hat{K}_{avg}$	17.4986	17.4985	17.5104	17.4990
$\hat{K}_{SD}$	1.199×10^{-1}	1.070×10^{-1}	9.939×10^{-2}	1.151×10^{-1}
$SE(\hat{K})_{avg}$	1.171×10^{-1}	1.112×10^{-1}	9.255×10^{-2}	1.115×10^{-1}
$\hat{r}_{avg}$	0.7015	0.7007	0.7052	0.7003
$\hat{r}_{SD}$	2.135×10^{-2}	2.481×10^{-2}	6.439×10^{-2}	3.111×10^{-2}
$SE(\hat{r})_{avg}$	2.100×10^{-2}	2.560×10^{-2}	6.153×10^{-2}	3.114×10^{-2}
$\hat{x}_{0,avg}$	0.1002	0.1012	0.1037	0.1030
$\hat{x}_{0,SD}$	1.567×10^{-2}	1.864×10^{-2}	3.965×10^{-2}	2.323×10^{-2}
$SE(\hat{x}_0)_{avg}$	1.523×10^{-2}	1.894×10^{-2}	3.710×10^{-2}	2.314×10^{-2}

Table 7.4: Average estimates obtained from GEE-1 ($\hat{\theta}_{Avg}$) with their standard deviations ($\hat{\theta}_{SD}$) and average asymptotic standard errors (7.2) ($SE(\hat{\theta})_{avg}$) from $M = 250$ Monte Carlo trials. Autocorrelated data corresponding to the optimal time points was simulated using true values $(K, r, x_0) = (17.5, 0.7, 0.1)$, $N = 15$, 10% noise ($\sigma_0^2 = 0.10$), and $\alpha^* = 10^{-4}$ autocorrelation.

	SE-opt	D-opt	E-opt	Uniform
$\hat{K}_{avg}$	17.4989	17.4986	17.5060	17.4990
$\hat{K}_{SD}$	1.276×10^{-1}	1.069×10^{-1}	1.242×10^{-1}	1.151×10^{-1}
$SE(\hat{K})_{avg}$	1.910×10^{-1}	1.118×10^{-1}	1.602×10^{-1}	1.115×10^{-1}
$\hat{r}_{avg}$	0.7027	0.7006	0.7053	0.7003
$\hat{r}_{SD}$	2.633×10^{-2}	2.489×10^{-2}	6.602×10^{-2}	3.111×10^{-2}
$SE(\hat{r})_{avg}$	3.915×10^{-2}	2.835×10^{-2}	7.677×10^{-2}	3.114×10^{-2}
$\hat{x}_{0,avg}$	0.0998	0.1014	0.1040	0.1030
$\hat{x}_{0,SD}$	2.047×10^{-2}	1.885×10^{-2}	4.073×10^{-2}	2.323×10^{-2}
$SE(\hat{x}_0)_{avg}$	2.906×10^{-2}	2.128×10^{-2}	4.712×10^{-2}	2.314×10^{-2}
$\hat{\alpha}^*_{avg}$	3.521×10^{-6}	3.245×10^{-3}	1.306×10^{-66}	9.999×10^{-5}
$\hat{\alpha}^*_{SD}$	5.354×10^{-5}	1.817×10^{-2}	2.061×10^{-65}	4.346×10^{-19}

Table 7.5: Average estimates obtained from OLS ($\hat{\theta}_{Avg}$) with their standard deviations ($\hat{\theta}_{SD}$) and average asymptotic standard errors (7.3) ($SE(\hat{\theta})_{avg}$) from $M = 250$ Monte Carlo trials. Autocorrelated data corresponding to the optimal time points was simulated using true values $(K, r, x_0) = (17.5, 0.7, 0.1)$, $N = 15$, 10% noise ($\sigma_0^2 = 0.10$), and $\alpha^* = 0.225$ autocorrelation.

	SE-opt	D-opt	E-opt	Uniform
$\hat{K}_{avg}$	17.5390	17.5104	17.5345	17.5054
$\hat{K}_{SD}$	2.982×10^{-1}	1.240×10^{-1}	3.322×10^{-1}	1.213×10^{-1}
$SE(\hat{K})_{avg}$	3.131×10^{-2}	9.477×10^{-2}	2.828×10^{-2}	1.074×10^{-1}
$\hat{r}_{avg}$	0.7014	0.7068	0.7126	0.6999
$\hat{r}_{SD}$	4.791×10^{-2}	4.337×10^{-2}	1.107×10^{-1}	3.344×10^{-2}
$SE(\hat{r})_{avg}$	5.651×10^{-3}	2.194×10^{-2}	1.995×10^{-2}	2.988×10^{-2}
$\hat{x}_{0,avg}$	0.1042	0.0997	0.1122	0.1028
$\hat{x}_{0,SD}$	3.628×10^{-2}	3.265×10^{-2}	6.448×10^{-2}	2.414×10^{-2}
$SE(\hat{x}_0)_{avg}$	4.165×10^{-3}	1.583×10^{-2}	1.154×10^{-2}	2.237×10^{-2}

Table 7.6: Average estimates obtained from GEE-1 ($\hat{\theta}_{Avg}$) with their standard deviations ($\hat{\theta}_{SD}$) and average asymptotic standard errors (7.2) ($SE(\hat{\theta})_{avg}$) from $M = 250$ Monte Carlo trials. Autocorrelated data corresponding to the optimal time points was simulated using true values $(K, r, x_0) = (17.5, 0.7, 0.1)$, $N = 15$, 10% noise ($\sigma_0^2 = 0.10$), and $\alpha^* = 0.225$ autocorrelation.

	SE-opt	D-opt	E-opt	Uniform
$\hat{K}_{avg}$	17.5405	17.5114	17.5339	17.5053
$\hat{K}_{SD}$	2.989×10^{-1}	1.233×10^{-1}	3.330×10^{-1}	1.213×10^{-1}
$SE(\hat{K})_{avg}$	7.628×10^{-2}	9.887×10^{-2}	8.694×10^{-2}	1.095×10^{-1}
$\hat{r}_{avg}$	0.7011	0.7063	0.7116	0.6999
$\hat{r}_{SD}$	4.795×10^{-2}	4.303×10^{-2}	1.090×10^{-1}	3.342×10^{-2}
$SE(\hat{r})_{avg}$	1.169×10^{-2}	2.782×10^{-2}	3.249×10^{-2}	3.025×10^{-2}
$\hat{x}_{0,avg}$	0.1044	0.1000	0.1126	0.1028
$\hat{x}_{0,SD}$	3.637×10^{-2}	3.245×10^{-2}	6.476×10^{-2}	2.413×10^{-2}
$SE(\hat{x}_0)_{avg}$	8.573×10^{-3}	2.058×10^{-2}	1.965×10^{-2}	2.267×10^{-2}
$\hat{\alpha}^*_{avg}$	0.1402	0.0550	0.0744	0.0572
$\hat{\alpha}^*_{SD}$	1.497×10^{-1}	1.040×10^{-1}	9.364×10^{-2}	9.446×10^{-2}

Table 7.7: Average estimates obtained from OLS ($\hat{\theta}_{Avg}$) with their standard deviations ($\hat{\theta}_{SD}$) and average asymptotic standard errors (7.3) ($SE(\hat{\theta})_{avg}$) from $M = 250$ Monte Carlo trials. Autocorrelated data corresponding to the optimal time points was simulated using true values $(K, r, x_0) = (17.5, 0.7, 0.1)$, $N = 15$, 10% noise ($\sigma_0^2 = 0.10$), and $\alpha^* = 0.45$ autocorrelation.

	SE-opt	D-opt	E-opt	Uniform
$\hat{K}_{avg}$	17.4942	17.4940	17.5236	17.4977
$\hat{K}_{SD}$	2.871×10^{-1}	1.616×10^{-1}	3.030×10^{-1}	1.368×10^{-1}
$SE(\hat{K})_{avg}$	2.957×10^{-2}	9.268×10^{-2}	2.448×10^{-2}	1.055×10^{-1}
$\hat{r}_{avg}$	0.7046	0.7050	0.7188	0.7009
$\hat{r}_{SD}$	4.134×10^{-2}	3.971×10^{-2}	1.098×10^{-1}	3.659×10^{-2}
$SE(\hat{r})_{avg}$	5.388×10^{-3}	2.143×10^{-2}	1.737×10^{-2}	2.945×10^{-2}
$\hat{x}_{0,avg}$	0.1010	0.1012	0.1075	0.1029
$\hat{x}_{0,SD}$	3.035×10^{-2}	3.048×10^{-2}	6.004×10^{-2}	2.802×10^{-2}
$SE(\hat{x}_0)_{avg}$	3.875×10^{-3}	1.582×10^{-2}	9.813×10^{-3}	2.190×10^{-2}

Table 7.8: Average estimates obtained from GEE-1 ($\hat{\theta}_{Avg}$) with their standard deviations ($\hat{\theta}_{SD}$) and average asymptotic standard errors (7.2) ($SE(\hat{\theta})_{avg}$) from $M = 250$ Monte Carlo trials. Autocorrelated data corresponding to the optimal time points was simulated using true values $(K, r, x_0) = (17.5, 0.7, 0.1)$, $N = 15$, 10% noise ($\sigma_0^2 = 0.10$), and $\alpha^* = 0.45$ autocorrelation.

	SE-opt	D-opt	E-opt	Uniform
$\hat{K}_{avg}$	17.4941	17.4948	17.5253	17.4977
$\hat{K}_{SD}$	2.875×10^{-1}	1.609×10^{-1}	3.014×10^{-1}	1.381×10^{-1}
$SE(\hat{K})_{avg}$	7.076×10^{-2}	1.054×10^{-1}	7.370×10^{-2}	1.169×10^{-1}
$\hat{r}_{avg}$	0.7045	0.7048	0.7191	0.7015
$\hat{r}_{SD}$	4.159×10^{-2}	3.896×10^{-2}	1.088×10^{-1}	3.718×10^{-2}
$SE(\hat{r})_{avg}$	1.097×10^{-2}	2.912×10^{-2}	2.832×10^{-2}	3.083×10^{-2}
$\hat{x}_{0,avg}$	0.1011	0.1011	0.1068	0.1026
$\hat{x}_{0,SD}$	3.057×10^{-2}	2.993×10^{-2}	5.854×10^{-2}	2.815×10^{-2}
$SE(\hat{x}_0)_{avg}$	7.830×10^{-3}	2.176×10^{-2}	1.641×10^{-2}	2.286×10^{-2}
$\hat{\alpha}^*_{avg}$	0.2606	0.1536	0.1678	0.1565
$\hat{\alpha}^*_{SD}$	2.318×10^{-1}	2.0446×10^{-1}	1.794×10^{-1}	1.978×10^{-1}

Table 7.9: Average estimates obtained from OLS ($\hat{\theta}_{Avg}$) with their standard deviations ($\hat{\theta}_{SD}$) and average asymptotic standard errors (7.3) ($SE(\hat{\theta})_{avg}$) from $M = 250$ Monte Carlo trials. Autocorrelated data corresponding to the optimal time points was simulated using true values $(K, r, x_0) = (17.5, 0.7, 0.1)$, $N = 15$, 10% noise ($\sigma_0^2 = 0.10$), and $\alpha^* = 0.675$ autocorrelation.

	SE-opt	D-opt	E-opt	Uniform
$\hat{K}_{avg}$	17.5183	17.4833	17.4952	17.5060
$\hat{K}_{SD}$	2.851×10^{-1}	1.829×10^{-1}	3.042×10^{-1}	1.754×10^{-1}
$SE(\hat{K})_{avg}$	2.525×10^{-2}	8.211×10^{-2}	2.223×10^{-2}	9.064×10^{-2}
$\hat{r}_{avg}$	0.6998	0.7009	0.7040	0.7009
$\hat{r}_{SD}$	3.755×10^{-2}	3.894×10^{-2}	1.020×10^{-1}	4.212×10^{-2}
$SE(\hat{r})_{avg}$	4.522×10^{-3}	1.895×10^{-2}	1.513×10^{-2}	2.534×10^{-2}
$\hat{x}_{0,avg}$	0.1024	0.1032	0.1176	0.1047
$\hat{x}_{0,SD}$	2.798×10^{-2}	2.891×10^{-2}	6.622×10^{-2}	3.256×10^{-2}
$SE(\hat{x}_0)_{avg}$	3.365×10^{-3}	1.420×10^{-2}	9.546×10^{-3}	1.890×10^{-2}

Table 7.10: Average estimates obtained from GEE-1 ($\hat{\theta}_{Avg}$) with their standard deviations ($\hat{\theta}_{SD}$) and average asymptotic standard errors (7.2) ($SE(\hat{\theta})_{avg}$) from $M = 250$ Monte Carlo trials. Autocorrelated data corresponding to the optimal time points was simulated using true values $(K, r, x_0) = (17.5, 0.7, 0.1)$, $N = 15$, 10% noise ($\sigma_0^2 = 0.10$), and $\alpha^* = 0.675$ autocorrelation.

	SE-opt	D-opt	E-opt	Uniform
$\hat{K}_{avg}$	17.5187	17.4834	17.4966	17.5083
$\hat{K}_{SD}$	2.838×10^{-1}	1.874×10^{-1}	3.033×10^{-1}	1.744×10^{-1}
$SE(\hat{K})_{avg}$	6.181×10^{-2}	1.286×10^{-1}	6.232×10^{-2}	1.345×10^{-1}
$\hat{r}_{avg}$	0.6998	0.7019	0.7026	0.6994
$\hat{r}_{SD}$	3.779×10^{-2}	4.007×10^{-2}	1.031×10^{-1}	4.109×10^{-2}
$SE(\hat{r})_{avg}$	9.027×10^{-3}	2.634×10^{-2}	2.377×10^{-2}	2.731×10^{-2}
$\hat{x}_{0,avg}$	0.1024	0.1026	0.1191	0.1058
$\hat{x}_{0,SD}$	2.839×10^{-2}	2.991×10^{-2}	6.726×10^{-2}	3.213×10^{-2}
$SE(\hat{x}_0)_{avg}$	6.664×10^{-3}	2.013×10^{-2}	1.585×10^{-2}	2.105×10^{-2}
$\hat{\alpha}^*_{avg}$	0.3724	0.4945	0.2178	0.4980
$\hat{\alpha}^*_{SD}$	3.049×10^{-1}	3.079×10^{-1}	2.509×10^{-1}	2.806×10^{-1}

Table 7.11: Average estimates obtained from OLS ($\hat{\theta}_{Avg}$) with their standard deviations ($\hat{\theta}_{SD}$) and average asymptotic standard errors (7.3) ($SE(\hat{\theta})_{avg}$) from $M = 250$ Monte Carlo trials. Autocorrelated data corresponding to the optimal time points was simulated using true values $(K, r, x_0) = (17.5, 0.7, 0.1)$, $N = 15$, 10% noise ($\sigma_0^2 = 0.10$), and $\alpha^* = 0.9$ autocorrelation.

	SE-opt	D-opt	E-opt	Uniform
$\hat{K}_{avg}$	17.4851	17.5056	17.4932	17.4922
$\hat{K}_{SD}$	3.061×10^{-1}	2.626×10^{-1}	3.304×10^{-1}	2.626×10^{-1}
$SE(\hat{K})_{avg}$	2.547×10^{-2}	6.181×10^{-2}	1.458×10^{-2}	7.234×10^{-2}
$\hat{r}_{avg}$	0.7036	0.6995	0.7133	7.0199
$\hat{r}_{SD}$	2.906×10^{-2}	2.764×10^{-2}	8.105×10^{-2}	3.698×10^{-2}
$SE(\hat{r})_{avg}$	4.649×10^{-3}	1.423×10^{-2}	1.052×10^{-2}	2.025×10^{-2}
$\hat{x}_{0,avg}$	0.0992	0.1035	0.1050	0.1039
$\hat{x}_{0,SD}$	2.129×10^{-2}	2.426×10^{-2}	5.448×10^{-2}	3.259×10^{-2}
$SE(\hat{x}_0)_{avg}$	3.246×10^{-3}	1.078×10^{-2}	5.682×10^{-3}	1.513×10^{-2}

Table 7.12: Average estimates obtained from GEE-1 ($\hat{\theta}_{Avg}$) with their standard deviations ($\hat{\theta}_{SD}$) and average asymptotic standard errors (7.2) ($SE(\hat{\theta})_{avg}$) from $M = 250$ Monte Carlo trials. Autocorrelated data corresponding to the optimal time points was simulated using true values $(K, r, x_0) = (17.5, 0.7, 0.1)$, $N = 15$, 10% noise ($\sigma_0^2 = 0.10$), and $\alpha^* = 0.9$ autocorrelation.

	SE-opt	D-opt	E-opt	Uniform
$\hat{K}_{avg}$	17.4865	17.5050	17.4931	17.4904
$\hat{K}_{SD}$	3.061×10^{-1}	2.597×10^{-1}	3.301×10^{-1}	2.593×10^{-1}
$SE(\hat{K})_{avg}$	6.105×10^{-2}	1.034×10^{-1}	3.336×10^{-2}	1.243×10^{-1}
$\hat{r}_{avg}$	0.7033	0.6996	0.7133	0.7022
$\hat{r}_{SD}$	2.916×10^{-2}	2.804×10^{-2}	8.438×10^{-2}	3.173×10^{-2}
$SE(\hat{r})_{avg}$	9.184×10^{-3}	1.883×10^{-2}	1.619×10^{-2}	2.190×10^{-2}
$\hat{x}_{0,avg}$	0.0996	0.1033	0.1062	0.1022
$\hat{x}_{0,SD}$	2.143×10^{-2}	2.457×10^{-2}	5.751×10^{-2}	2.761×10^{-2}
$SE(\hat{x}_0)_{avg}$	6.450×10^{-3}	1.465×10^{-2}	9.067×10^{-3}	1.656×10^{-2}
$\hat{\alpha}^*_{avg}$	0.3149	0.5717	0.1143	0.5998
$\hat{\alpha}^*_{SD}$	4.026×10^{-1}	3.030×10^{-1}	2.910×10^{-1}	2.645×10^{-1}

same order of magnitude from the two different tables. Since the autocorrelation is so small in this example, we expect that the two different formulations for approximate standard errors (equations (7.2) and (7.3)) would be very similar.

Discussion for $\alpha^* = 0.225$:

For data simulated with autocorrelation level $\alpha^* = 0.225$, we compare the optimal design methods based on their results using OLS (Table 7.5). Comparing the optimal design methods based on how close their average estimates are to the true values, we find that for parameters K and r the uniform mesh is the closest followed by D-optimal, and for x_0 D-optimal is the closest followed by the uniform mesh. Comparing the optimal design methods based on which has smallest standard deviation we find that for all model parameters (K, r, x_0) the uniform mesh followed by D-optimal has the smallest SD's. Comparing the average approximate asymptotic standard errors (SE) to the standard deviations we find that the SE's are smaller than the SD's in every case, and the optimal design methods that have the smallest SE's are different than those with the smallest SD's. The optimal design methods with the smallest SE's are D-optimal followed by SE-optimal for K, and SE-optimal followed by E-optimal for r and x_0.

We also compare the methods based on the results from the GEE-1 algorithm for data simulated with $\alpha^* = 0.225$ (Table 7.6). Comparing the optimal design methods based on which has average parameter estimates closest to the true value, we find that the uniform mesh is closest for K and r followed by D-optimal (for K) and SE-optimal (for r), for x_0 D-optimal is closest followed by the uniform mesh, and for α^* the closest is SE-optimal followed by D-optimal. Comparing the optimal design methods based on which has smallest SD's, we find that for (K, r, x_0) the uniform mesh is smallest followed by D-optimal (same as in the OLS case, Table 7.5), and E-optimal followed by the uniform mesh for α^*. The average asymptotic standard errors are again smaller than the SD's. Comparing the optimal design methods based on which has the smallest SE's, we find that SE-optimal is the smallest followed by E-optimal for K and x_0 and followed by D-optimal for r.

Comparing the OLS results (Table 7.5) to the GEE-1 results (Table 7.6) we find that their average parameter estimates and their standard deviations are similar for the common parameters (K, r, x_0). In general the average approximate asymptotic standard errors are larger from the GEE-1 algorithm compared to the OLS case.

Discussion for $\alpha^* = 0.45$:

Comparing which optimal design methods had average parameter estimates closest to the true value from OLS with autocorrelated data $\alpha^* = 0.45$ (Table 7.7), we find that the uniform mesh is closest for K and r followed by SE-optimal, and SE-optimal is closest for x_0 followed by D-optimal. Comparing the optimal design methods based on their standard deviations (SD),

we find that the uniform mesh has the smallest SD's followed by D-optimal (for K and r) and SE-optimal (for x_0). In general, the average approximate standard errors (SE) are smaller than the SD's. The optimal design method with the smallest SE is E-optimal followed by SE-optimal for K, and SE-optimal followed by E-optimal for r and x_0.

Comparing the optimal design methods based on the results using the GEE-1 (Table 7.8), we find that the uniform mesh has closest average parameter estimate to the true values for K and r followed by D-optimal (for K) and SE-optimal (for r). The average estimate closest to the true value for x_0 was SE-optimal followed by D-optimal, and for α^* SE-optimal was closest followed by E-optimal. Comparing the optimal design methods based on which has the smallest SD, we find that for (K, r, x_0) the uniform mesh is smallest followed by D-optimal, and for α^* E-optimal has the smallest SD followed by the uniform mesh. In every case the average approximate asymptotic standard errors are smaller than the standard deviation, more so with SE-optimal and E-optimal than with D-optimal and the uniform mesh. Comparing the methods based on which has the smallest SE, we find that SE-optimal is the smallest followed by E-optimal (for K and r) and D-optimal (for x_0).

The average parameter estimates and standard deviations are very similar for (K, r, x_0) comparing the OLS and GEE-1 results (Tables 7.7 and 7.8). The average approximate asymptotic standard errors are smaller in the OLS results compared to GEE-1 results.

Discussion for $\alpha^* = 0.675$:

Table 7.9 contains the results from OLS with $\alpha^* = 0.675$ autocorrelated data. Comparing the optimal design methods based on which has average estimates closest to the true value, we find that for K E-optimal is closest followed by the uniform mesh, and for r and x_0 SE-optimal is closest followed by D-optimal. Comparing the optimal design methods based on which has the smallest standard deviation we find that for K the uniform mesh is the smallest followed by D-optimal, and for r and x_0 SE-optimal is the smallest followed by D-optimal. Again, we find that the average standard errors are smaller than the standard deviations. The optimal design method with the smallest average asymptotic standard error is E-optimal for K followed by SE-optimal, and SE-optimal followed by E-optimal for r and x_0.

Table 7.10 contains the corresponding results from GEE-1, using the correct assumptions about the errors. Comparing the average parameter estimates to the true values, we find that for K E-optimal is the closest followed by the uniform mesh, for r and x_0 SE-optimal is the closest followed by the uniform mesh (for r) and D-optimal (for x_0), and for α^* the uniform mesh is closest followed by D-optimal. Comparing the optimal design methods based on which has the smallest SD, we find that for K the uniform mesh followed by D-optimal is the smallest, for r and x_0 SE-optimal is the smallest followed by D-optimal, and for α^* E-optimal is the smallest followed by the uniform mesh. The average asymptotic standard errors are smaller

than the standard deviations. Comparing the optimal design methods based on which has the smallest SE, we find that for (K, r, x_0) SE-optimal is the smallest followed by E-optimal.

Comparing Tables 7.9 and 7.10, the average parameter estimates and their standard errors are similar. The average asymptotic standard errors are smaller from OLS compared to the GEE-1 results.

Discussion for $\alpha^* = 0.9$:

We compare the optimal design methods based on their OLS results (Table 7.11) for data simulated with autocorrelation level $\alpha^* = 0.9$. Comparing the optimal design methods based on how close their average estimates are to the true values, we find that for parameters K and r D-optimal is the closest followed by E-optimal (for K) and the uniform mesh (for r), and for x_0 SE-optimal is the closest followed by D-optimal. Comparing the optimal design methods based on which has smallest standard deviation we find for K D-optimal and the uniform mesh are the smallest, for r D-optimal is the smallest followed by SE-optimal, and for x_0 SE-optimal is the smallest followed by D-optimal. Comparing the average approximate asymptotic standard errors (SE) to the standard deviations we find that the SE's are smaller than the SD's in every case. The optimal design methods with the smallest SE's are E-optimal followed by SE-optimal for K, and SE-optimal followed by E-optimal for r and x_0.

We also compare the methods based on the results from the GEE-1 algorithm (Table 7.12). Comparing the optimal design methods based on which has average parameter estimates closest to the true value, we find that for K and r D-optimal is closest followed by E-optimal (for K) and the uniform mesh (for r), for x_0 SE-optimal is closest followed by the uniform mesh, and for α^* the closest is the uniform mesh followed by D-optimal. Comparing the optimal design methods based on which has smallest SD's, we find that for K the uniform mesh is smallest followed by D-optimal, for r D-optimal is the smallest followed by SE-optimal, for x_0 SE-optimal is the smallest followed by D-optimal, and for α^* the uniform mesh followed by E-optimal is smallest. The average asymptotic standard errors are smaller than the SD's. The optimal design methods with the smallest SE's are E-optimal followed by SE-optimal for K, and SE-optimal followed by E-optimal for r and x_0 (same as in the OLS case above).

Comparing the OLS results (Table 7.11) to the GEE-1 results (Table 7.12) we find that their average parameter estimates and their standard deviations are similar for parameters (K, r, x_0). In general the average approximate asymptotic standard errors are larger from the GEE-1 algorithm compared to the OLS case.

In conclusion, no optimal design method is consistently favorable among the OLS or GEE-1 results. Comparing the average estimates for (K, r, x_0) and their standard deviations from OLS to those from GEE-1, we find very similar results. For both inverse problem methods (OLS and

GEE-1), we find that when autocorrelation is such that $\alpha^* > 10^{-4}$, within any set of results the average asymptotic standard error is less than the standard deviation. From this Monte Carlo analysis, we find that using OLS with autocorrelated data may give too optimistic asymptotic standard errors.

Here we used the optimal design methods assuming an independent error statistical model (Fig. 5.3). For these optimal design points, using GEE-1, the averages estimates of α^* were lower than the true value in all cases with true autocorrelation $\alpha^* > 10^{-4}$, and large standard deviations (compared to the model parameters of similar magnitude). Also, for this analysis, no optimal design method (or the uniform mesh) was favorable for estimating α^* for all levels of autocorrelation in the data.

7.6.3 The Effect of the Assumptions made within the Optimal Design methods about Correlation in the Errors.

In this analysis we explore the effect that the assumptions made when constructing the optimal mesh have. We repeat the analysis from the previous section only for the case of having data with an autocorrelation level of $\alpha^* = 0.9$ now using the optimal design points which assumed a statistical model with autocorrelation (Fig 7.1(b), with $N = 15$ and constraint implementation $(C3)$). In practice, this situation could arise if one person assumed autocorrelation in the optimal design methods, and another person got the data corresponding to the optimal design points, not knowing the underlying assumptions, and assumed the errors were independent and used the OLS formulation of the inverse problem.

Now using optimal design points which assumed an autocorrelated statistical model, we repeat $M = 250$ Monte Carlo trial using both OLS and GEE-1 for each data set which contains $\alpha^* = 0.9$ autocorrelation in the errors. We will compare the results to Tables 7.11 and 7.12, which is a similar Monte Carlo analysis with the only difference being the underlying optimal design points.

Table 7.13 contains the results from OLS (incorrect assumption) solving the inverse problem with autocorrelated data corresponding to optimal design points which assumed a statistical model with autocorrelation. Table 7.14 contains the results from the GEE-1 algorithm (correct assumption) for the same autocorrelated data.

Discussion

In the discussion below, we will start by discussing the results tables individually, and then compare tables to each other and to the tables in the previous section.

Table 7.13 contains the results from OLS (incorrect assumption) from optimal design methods (assuming autocorrelation in the errors), and autocorrelated data with $\alpha^* = 0.9$. Comparing

Table 7.13: Average estimates obtained from OLS ($\hat{\theta}_{Avg}$) with their standard deviations ($\hat{\theta}_{SD}$) and average asymptotic standard errors (7.3) ($SE(\hat{\theta})_{avg}$) from $M = 250$ Monte Carlo trials. Autocorrelated data corresponding to the optimal time points with autocorrelation assumption was simulated using true values $(K, r, x_0) = (17.5, 0.7, 0.1)$, $N = 15$, 10% noise ($\sigma_0^2 = 0.10$), and $\alpha^* = 0.9$ autocorrelation.

	SE-opt	D-opt	E-opt	Uniform
$\hat{K}_{avg}$	17.4832	17.4994	17.5099	17.5322
$\hat{K}_{SD}$	2.679×10^{-1}	2.749×10^{-1}	2.641×10^{-1}	2.589×10^{-1}
$SE(\hat{K})_{avg}$	9.390×10^{-2}	6.822×10^{-2}	5.839×10^{-2}	7.022×10^{-2}
$\hat{r}_{avg}$	0.7007	0.7030	0.6998	0.7009
$\hat{r}_{SD}$	2.812×10^{-2}	2.779×10^{-2}	3.269×10^{-2}	3.235×10^{-2}
$SE(\hat{r})_{avg}$	1.222×10^{-2}	1.455×10^{-2}	1.885×10^{-2}	1.958×10^{-2}
$\hat{x}_{0,avg}$	0.1020	0.1005	0.1034	0.1028
$\hat{x}_{0,SD}$	2.507×10^{-2}	2.274×10^{-2}	2.708×10^{-2}	2.733×10^{-2}
$SE(\hat{x}_0)_{avg}$	8.252×10^{-3}	1.058×10^{-2}	1.326×10^{-2}	1.458×10^{-2}

Table 7.14: Average estimates obtained from GEE-1 ($\hat{\theta}_{Avg}$) with their standard deviations ($\hat{\theta}_{SD}$) and average asymptotic standard errors (7.2) ($SE(\hat{\theta})_{avg}$) from $M = 250$ Monte Carlo trials. Autocorrelated data corresponding to the optimal time points with autocorrelation assumption was simulated using true values $(K, r, x_0) = (17.5, 0.7, 0.1)$, $N = 15$, 10% noise ($\sigma_0^2 = 0.10$), and $\alpha^* = 0.9$ autocorrelation.

	SE-opt	D-opt	E-opt	Uniform
$\hat{K}_{avg}$	17.4808	17.4992	17.5082	17.5312
$\hat{K}_{SD}$	2.695×10^{-1}	2.728×10^{-1}	2.592×10^{-1}	2.516×10^{-1}
$SE(\hat{K})_{avg}$	1.195×10^{-1}	1.006×10^{-1}	1.024×10^{-1}	1.219×10^{-1}
$\hat{r}_{avg}$	0.7011	0.7027	0.6995	0.7014
$\hat{r}_{SD}$	2.694×10^{-2}	2.690×10^{-2}	3.041×10^{-2}	2.784×10^{-2}
$SE(\hat{r})_{avg}$	1.555×10^{-2}	1.678×10^{-2}	2.068×10^{-2}	2.119×10^{-2}
$\hat{x}_{0,avg}$	0.1013	0.1004	0.1035	0.1010
$\hat{x}_{0,SD}$	2.285×10^{-2}	2.178×10^{-2}	2.559×10^{-2}	2.315×10^{-2}
$SE(\hat{x}_0)_{avg}$	1.137×10^{-2}	1.250×10^{-2}	1.575×10^{-2}	1.587×10^{-2}
$\hat{\alpha}^*_{avg}$	0.4450	0.4783	0.5709	0.6272
$\hat{\alpha}^*_{SD}$	3.367×10^{-1}	3.637×10^{-1}	2.948×10^{-1}	2.405×10^{-1}

the optimal design methods based on how close the average parameter estimate is to the true value, we find that for K and x_0 D-optimal is the closest followed by E-optimal (for K) and SE-optimal (for x_0), and for r E-optimal is closest followed by SE-optimal. Comparing the optimal design methods based on their standard deviations, we find that for K the uniform mesh is smallest, followed by E-optimal, and for r and x_0 D-optimal is smallest followed by SE-optimal. Comparing methods based on their average asymptotic standard errors, we find that for K E-optimal is smallest followed by D-optimal, and for r and x_0 SE-optimal is smallest followed by D-optimal.

The results for GEE-1 are reported in Table 7.14, such that the optimal design assumptions and the inverse methodology (GEE-1) make the correct assumption about autocorrelation in the data. Comparing the optimal design methods based on how close their average estimate is to the true value, we find that for K and x_0 D-optimal is the closest followed by E-optimal (for K) and the uniform mesh (for x_0), for r E-optimal is closest followed by SE-optimal, and for α^* the uniform mesh was closest followed by E-optimal. Comparing the optimal design methods based on which has the smallest standard deviation, we find that for K and α^* the uniform mesh is smallest followed by E-optimal, and for r and x_0 D-optimal is smallest followed by SE-optimal. Comparing the optimal design methods based on their average asymptotic standard errors, we find that for K D-optimal is smallest followed by E-optimal, and for r and x_0 SE-optimal is the smallest followed by D-optimal.

Comparing the OLS and GEE-1 results (Tables 7.13 and 7.14) based on the model parameters (K, r, x_0) we find that the average estimates are similar, and the standard deviations from GEE-1 are slightly smaller than those from OLS. In addition, the average standard errors from GEE-1 are larger than those from OLS, which we expect based on equations (7.2) and (7.3) which correspond to their respective assumptions. Though GEE-1 produces larger average asymptotic standard errors, we know in this example that they are more realistic given the nature of our simulated autocorrelated data.

Comparing the OLS estimates from the optimal design methods assuming independent errors (Table 7.11) to the OLS estimates from the optimal design methods assuming autocorrelated errors (Table 7.13) for autocorrelated data with $\alpha^* = 0.9$, the average estimates are similar, and the standard deviations are similar and often slightly smaller in Table 7.13. The average standard errors are often slightly larger in Table 7.13 comparing to Table 7.11. The optimal design assumptions (independence vs autocorrelation) do have a subtle effect on the results using OLS (incorrect assumption).

Comparing the GEE-1 estimates from the optimal design methods assuming independent errors (Table 7.12) to the GEE-1 estimates from the optimal design methods assuming autocorrelated errors (Table 7.14), we find that the average estimates are similar for (K, r, x_0). The average estimates of α^* are closer using the autocorrelated optimal design mesh (Table 7.14)

for SE-optimal, E-optimal (much closer) and the uniform mesh. The standard deviations from Table 7.14 were often slightly smaller than the standard deviations in Table 7.12. Comparing the average asymptotic standard errors we find that for D-optimal and the uniform mesh were smaller in Table 7.14 than in Table 7.12. The optimal design assumptions (independence vs autocorrelation) effect the results from GEE-1, such as the average estimates of α, and the standard deviation.

7.7 Conclusions

Just as residual analysis is useful in determining if the data has constant or non-constant variance, it is also useful in determining if autocorrelation is present. If the data is autocorrelated it is important to incorporate a suitable autocorrelation model into the statistical model. It is simple to incorporate the new statistical model into the formulation of the FIM, which is essential in computing the optimal design points. Data taken at these optimal design points can be used with the GEE-1 algorithm to estimate the model parameters as well as the variance parameters: $\gamma = (\sigma^2, \alpha)$.

We compared the optimal design methods based on the logistic model corresponding to a statistical model with autocorrelation in a Monte Carlo analysis with $M = 1000$. No optimal design was favorable based on the results for the model parameters (K, r, x_0). Estimation of the autocorrelation parameter was more difficult (estimates had a wide range of values, and high standard deviation). The uniform mesh was the best for estimation of the autocorrelation parameter, followed by E-optimal. On average, the autocorrelation parameter was underestimated.

We examined what happens if you incorrectly assume that the data is independent for the optimal design point and use OLS to solve the inverse problem. For each data set, the inverse problem was solved using OLS (incorrect assumption) and GEE-1 (correct assumption). In general the average estimates, from $M = 250$ Monte Carlo trials, and standard deviation were similar comparing OLS and GEE-1. For both OLS and GEE-1, when autocorrelation in the data is such that $\alpha^* > 10{-}4$, the average standard error is smaller than the standard deviation. The average asymptotic standard errors were smaller from OLS compared to GEE-1. Hence, OLS should not be used for autocorrelated data. Though the OLS asymptotic standard errors are smaller, the standard error estimates from GEE-1 are more realistic. In this analysis no optimal design method is favorable over the others. The autocorrelation parameter was only estimated from GEE-1. The average estimates of the autocorrelation parameter tended to be lower than the true value and the standard deviation was relatively high. No optimal design method was consistently favorable for estimation of the autocorrelation parameter.

Finally, we compared OLS and GEE-1 based on optimal design points which assumed an autocorrelated statistical model. We find similar results to the previous analysis with optimal

design point which assume independent errors. The optimal mesh which account for autocorrelation (compared to the results from the optimal design points which assumed independence) result in OLS estimates with smaller standard deviation and larger average standard errors. For the GEE-1 estimates, the optimal design points which account for autocorrelation result in better average estimates of the autocorrelation parameter, and slightly smaller standard deviations for all parameters. Again, for this analysis, no optimal design methods is favorable.

In general, all the optimal design methods perform similarly for the model parameters. Estimation of the autocorrelation is much more difficult and inconsistent. Though the GEE-1 algorithm is one way to solve this type of inverse problem, it would be worth examining other options. Often the uniform mesh was the best for estimating the autocorrelation parameter, but even those results had relatively large standard deviation and an average estimate lower than the true value. If estimation of the autocorrelation parameter was more precise, it may be easier to compare the optimal design methods with autocorrelation.

From this analysis we have learned a useful protocol for determining whether or not to include autocorrelation into our optimal design assumptions or not. Since the uniform mesh often was the best for estimating the autocorrelation parameter, start by collecting data on the optimal mesh. That data can be examined for the presence of autocorrelation (as well as constant or non-constant variance). State the statistical model with underlying assumptions that can either include autocorrelation or not depending on your findings. Then compute the optimal design points using the formulation of the FIM which corresponds to your statistical model. Parameters can be estimated using either GEE, OLS, or GLS, depending on your statistical model, from data corresponding to your optimal design points. This methodology should result in optimal design points with correct statistical assumptions, and more accurate parameter estimates.

Chapter 8

Concluding Remarks

8.1 Research Conclusions

We compared D-optimal, E-optimal and SE-optimal design methods for the logistic population model, the harmonic oscillator model, and a model for glucose regulation. D-optimal and E-optimal design methods are more established in the literature. Our comparisons test the performance of SE-optimal design, which is a relatively newer method. In every example, we found that the final time T, the number of points N, and the constraint implementation can effect the optimal time points for each of the methods, and the resulting parameter estimates and standard errors.

Here we give a brief overview of our finding from Chapters 5 and 6, where we assume independent errors with constant variance.

For the logistic example, there were similarities between the results in Section 5.1 and Section 6.2, both using Monte Carlo analysis but with slight differences. Comparing the average and median parameter estimates to their true values, we find that SE-optimal has closest parameter estimates for $N = 10$ time points. For $N = 15$, no method had estimates that were always closest to the true values. There were common findings about the smallest average and median standard errors, and the smallest standard deviations: for K, E-optimal was the smallest, and for r and x_0 SE-optimal was the smallest followed by D-optimal.

Given what we know about the cost function, it makes sense why E-optimal has the smallest standard errors and standard deviations for K. Since the true value of the parameter K is larger than the true values of r and x_0 ($(K, r, x_0) = (17.5, 0.7, 0.1)$) we expect that its standard error would also be larger. E-optimal minimizes the largest principle axis of the confidence ellipse, which in this example will most likely always be the axis corresponding to K. SE-optimal and D-optimal have smaller standard errors for r and x_0 since their cost functionals are minimizing the variance in all the parameters, in different ways.

Comparing the optimal design methods based on which has the best coverage of the true parameter value based on the 95% confidence ellipsoid, we find that SE-optimal is the best, followed by D-optimal or the uniform mesh, with E-optimal being the worst for this logistic example.

For the harmonic example, based on the parameter estimates none of the optimal design methods were consistently favorable, and all optimal design methods were close to the true values. Comparing the optimal design methods based on standard errors and standard deviations, D-optimal or SE-optimal tended to be smallest for parameter C, and for parameter K D-optimal and the uniform mesh tended to be the smallest. No optimal design method was best based on the 95% confidence ellipsoids coverage of the true values, except that often E-optimal was the worst.

For the glucose regulation model, comparing the approximate asymptotic standard errors, we found that for parameters (p_1, p_2, p_3) either SE-optimal or E-optimal had the smallest standard errors. D-optimal tended to have the smallest standard errors for p_4. Our results from the inverse problem, using asymptotic theory and bootstrapping to compute standard errors, were less conclusive. Comparing the parameter estimates to their true values, none of the optimal design methods were consistently closer. Comparing the optimal design methods based on who had the smallest standard errors and covariances we found that no method was preferable over the others. In general, the optimal design methods had smaller standard errors and covariances than the uniform mesh for this example.

In Chapter 7, we compared the optimal design methods for autocorrelated data. From our Monte Carlo analysis, we found that no method was favorable for the model parameters. The parameter estimation for the autocorrelation parameter, though less consistent, was often best using the uniform mesh followed by E-optimal based on the average estimates and the corresponding standard errors. We also examined the consequences of assuming the data is independent, when it is in fact autocorrelated. There wasn't a noticeable difference in the average or standard deviation of the estimates from the inverse problem using either OLS or GEE-1. However, the asymptotic standard errors from OLS were deceivingly smaller than the standard errors computed under the correct assumption. It is important check your assumptions and make sure you are using the correct statistical model.

The best choice of optimal design method depends on the complexity of the model, the type of constraint one is using, the subset of parameters you are estimating, and even the choice of N and T. The examples in this comparison provide evidence that SE-optimal design is competitive with D-optimal and E-optimal design.

8.2 Future Work

Some work remains to be done with optimal design methods applied to autocorrelated data. In Sections 7.6, we discuss the implications of making the wrong assumptions: using OLS to estimate parameters from autocorrelated data. For future work, it would be nice to quantify using model comparison techniques which method (OLS, or GEE-1) gives a better fit to the data for various levels of autocorrelation. It would also be informative to see if the autocorrelation parameter α^* is significantly different from 0 in the cases where its estimate from GEE-1 is very small. Overall, the estimates of the autocorrelation parameters had high variance. In future investigations it would be valuable to try other approaches to the inverse problem for the estimation of model and variance (including autocorrelation) parameters, with the hope that the autocorrelation parameter estimates may be more precise.

Much of this dissertation has focused comparing the optimal design methods based on which has the smallest standard errors. A future goal would be minimizing correlation between parameters using optimal design methods. Between D-optimal, E-optimal, and SE-optimal, does one optimal design method reduce correlation between the parameters more than the others? In Chapter 2, we observed that correlation between the parameters seemed to be related to the tilt of the confidence ellipsoid. Investigating this hypothesis further may give insight into an optimal design criteria that would be focused on reducing the correlation between parameters, or help us understand why one of the current optimal design methods may be ideal for this problem.

REFERENCES

[1] A.C. Atkinson and R.A. Bailey, One hundred years of the design of experiments on and off the pages of Biometrika, *Biometrika*, **88** (2001), 53–97.

[2] A.C. Atikinson and A.N. Donev, *Optimum Experimental Designs,* Oxford University Press, New York, 1992.

[3] H. T. Banks and K. L. Bihari, Modeling and estimating uncertainty in parameter estimation, *Inverse Problems*, **17** (2001), 95–111.

[4] H.T. Banks and C.A. Carter, Mathematical modeling of the glucose homeostatic system in humans, CRSC Technical Report, CRSC-TR10-09, NCSU, Raleigh, May, 2010.

[5] H.T. Banks, M. Davidian, J.R. Samuels Jr., and K.L. Sutton, An inverse problem statistical methodology summary, CRSC Technical Report, CRSC-TR08-01, NCSU, January, 2008; Chapter 11 in *Statistical Estimation Approaches in Epidemiology*, (edited by Gerardo Chowell, Mac Hyman, Nick Hengartner, Luis M.A. Bettencourt and Carlos Castillo-Chavez), Springer, Berlin Heidelberg New York, 2009, 249–302.

[6] H.T. Banks, Sava Dediu, S.L. Ernstberger and F. Kappel, A new approach to optimal design problems, CRSC-TR08-12, September, 2008, (Revised), November, 2009; *J. Inverse and Ill-posed Problems*, **18** (2010), 25–83.

[7] H.T. Banks, K. Holm and D. Robbins, Standard error computations for uncertainty quantification in inverse problems: Asymptotic theory vs. bootstrapping, CRSC-TR09-13, June, 2009; Revised, August 2009; *Mathematical and Computer Modeling*, **52** (2010), 1610–1625.

[8] H.T. Banks and H.T. Tran, *Mathematical and Experimental Modeling of Physical and Biological Processes,* Chapman and Hall/ CRC, Boca Raton, FL, 2008.

[9] M.P.F. Berger and W.K. Wong (Editors), *Applied Optimal Designs,* John Wiley & Sons, Chichester, UK, 2005.

[10] R.N. Bergman, L.S. Phillips and C. Cobelli, Physiologic evaluation of factors controlling glucose tolerance in man: measurement of insulin sensitivity and beta-cell glucose sensitivity from the response to intravenous glucose, *J Clin Invest,* **68(6)** (1981), 1456–1467.

[11] R.N. Bergman, Y.Z. Ider, C.R. Bowden and C. Cobelli, Quantitative estimation of insulin sensitivity, *Am. J. Physiol.,* **236** (1979), E667–E677.

[12] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, New York, NY, 1968.

[13] R. J. Carroll and D. Ruppert, *Transformation and Weighting in Regression*, Chapman & Hall, London, 1988.

[14] R. J. Carroll, C. F. J. Wu and D. Ruppert, The effect of estimating weights in Weighted Least Squares, *J. Amer. Statistical Assoc.*, **83** (1988), 1045–1054.

[15] M. Davidian and D. Giltinan, *Nonlinear Models for Repeated Measurement Data*, Chapman & Hall, London, 1998.

[16] M. Davidian, Marie Davidian's Nonlinear Models (ST 762) notes, Chapters 13 and 14.

[17] A. De Gaetano and O. Arino, Mathematical modeling of the intravenous glucose tolerance test, *J. Math. Biology,* **40** (2000), 136–168.

[18] Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G., *Longitudinal Data Analysis,* Chapman and Hall/CRC Press, Boca Raton, 2009.

[19] Diggle, P.J., Heagerty, P., Liang, K.-Y., and Zeger, S.L., *Analysis of Longitudinal Data, 2nd Edition*, Oxford University Press, Oxford, 2002.

[20] V. V. Fedorov, *Theory of Optimal Experiments*, Academic Press, New York and London, 1972.

[21] V.V. Fedorov and P. Hackel, *Model-Oriented Design of Experiments,* Springer-Verlag, New York, NY, 1997.

[22] P. J. Huber, *Robust Statistics*, John Wiley & Sons, Inc., New York, NY, 1981.

[23] A. Kuntsevich and F. Kappel. SolvOpt, Retrieved December 2009, from http://www.kfunigraz.ac.at/imawww/kuntsevich/solvopt/.

[24] Liang, K.-Y., Zeger, S.L., and Qaqish, G., Multivariate regression analysis for categorical data, *Journal of the Royal Statistical Society, Series B,* **54** (1992), 340.

[25] G. Pacini, and R.N. Bergman, MINMOD: a computer program to calculate insulin sensitivity and pancreatic responsivity from the frequently sampled intravenous glucose tolerance test, *Comput Methods Programs Biomed.,* **23(2)** (1986), 113–122.

[26] Prentice, R.L, Correlated binary regression with covariates specic to each binary observation, *Biometrics*, **44** (1988), 10331048.

[27] Prentice, R.L. and Zhao, L.P., Estimating equations for parameters in means and covariances of discrete and continuous responses, *Biometrics,* **47** (1991), 825838.

[28] Yu. V. Prohorov, Convergence of random processes and limit theorems in probability theory, *Theor. Prob. Appl.,* **1** (1956), 157–214.

[29] J.R. Schoott, *Matrix Analysis for Statistics*, John Wiley & Sons, Inc., Hoboken, NJ, 2005.

[30] G. A. F. Seber and C. J. Wild, *Nonlinear Regression,* John Wiley & Sons, New York, NY, 1989.

[31] G. Strang, *Linear Algebra and Its Applications*, Academic Press Inc., New York, 1976.

[32] G. Toffolo, R.N. Bergman, D.T. Finegood, C.R. Bowden and C. Cobelli, Quantitative estimation of beta cell sensitivity to glucose in the intact organism: a minimal of insulin kinetics in the dog, *Diabetes,* **29** (1980), 979–990.

[33] D. Ucinski and A.C. Atkinson, Experimental design for time-dependent models with correlated observations, *Studies in Nonlinear Dynamics and Econometrics*, **8(2)** (2004), Article 13: The Berkeley Electronic Press.

[34] Vonesh, E.F. and Chinchilli, V.M., *Linear and Nonlinear Models for the Analysis of Repeated Measurements,* Marcel Dekker, New York, 1997.