Inference for Optimal Dynamic Treatment Regimes Through a Bayesian Lens

Daniel Rodriguez Duque

Doctor of Philosophy

Department of Epidemiology, Biostatistics and Occupational Health

McGill University
Montréal, Québec
June 2022

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Acknowledgements

I am beholden to several individuals who supported me in the process of writing this work. First, I am grateful to my co-supervisors professors Erica E.M. Moodie and David A. Stephens; I have benefited greatly from their knowledge, attentiveness, availability, dedication, and effectiveness. I have learned a lot from them, and this thesis would not have been possible without them. I am also thankful to my committee member Dr. Marina B. Klein for her thoughtful comments and ideas.

I am thankful to professors who broadened my perspective throughout my time at McGill, including Dr. James Hanley, Dr. Alexandra Schmidt, Dr. Shirin Golchi, Dr. David Wolfson, Dr. Mireille Schnitzer, Dr. Andrea Benedetti, and Dr. Abbas Khalili. I am also grateful to the administrative staff in the department of epidemiology and biostatistics who were always so re-assuring and attentive, specially André Yves Gagnon, Katherine Hayden, and Dolores Coleto. I am also thankful to my McGill friends and colleagues, including Joanne, Armando, Victoire, Terry, Vanessa, Janie, Steve, and Larry. I am also indebted to my other friends who provided so much support during some challenging pandemic years, including Gregorio, Maher, and Mariami. I am especially thankful to Nassib for bringing sunshine to my life when I needed it most.

Lastly, I am thankful to my family who have done their best to support me throughout this process; they are the inspiration for all I do.

Funding from the National Sciences and Engineering Research Council, the Fonds de recherche du Québec– Nature et technologies, as well as from my supervisor Dr. Erica Moodie allowed me to carry out this work. The computational infrastructure and support provided by Compute Canada also allowed for the completion of this work.

Preface

This thesis seeks to develop methods for the identification of treatment rules that allow for optimal decision-making in delivering medical care, using a Bayesian perspective. Six chapters are developed, with chapters 1 and 2 containing and introduction and a literature review, chapters 3-5 containing original, inter-related manuscripts advancing the central aims of this thesis, and chapter 6 containing a conclusion and discussion of future work. In writing this thesis, I benefited from discussions and feedback from my supervisors and committee member; this is detailed in what follows.

Chapters 1 and 2 contain the introduction and literature review, both of which were written by Daniel Rodriguez Duque (DRD) and revised by Erica E.M. Moodie (EEMM) and David A. Stephens (DAS).

The general problem in Chapter 3 was suggested by EEMM and DAS, and further conceptualized and concretized by DRD. The methodological developments were performed by DRD with the guidance of EEMM and DAS. The simulations were designed and coded by DRD with troubleshooting suggestions and advice form EEMM and DAS. The illustrative example in HIV therapy was developed in discussions with EEMM, DAS, and Marina B. Klein (MBK); DRD performed all analyses in discussion with EEMM, DAS, and MBK. DRD wrote the manuscript with EEMM, DAS, and MBK providing revision and suggestions.

Chapter 4 was conceptualized through discussion between DRD, EEMM, and DAS. The framing, contextualization, and methodological developments were done by DRD with input from EEMM and DAS. All simulations were designed and programmed by DRD with guidance from EEMM and DAS. The data analysis was designed by DRD, with input from EEMM and DAS. The manuscript was written by DRD, with EEMM and DAS providing feedback about the content.

Chapter 5 was conceptualized in discussions between DRD, EEMM, and DAS. DRD conducted all analyses, wrote the manuscript, and wrote all programs; EEMM and DAS provided feedback and guidance throughout all these stages.

The conclusion and discussion of future work in chapter 6 was conceived and written by DRD with feedback and corrections provided by EEMM and DAS.

Abstract

The availability of health data, access to powerful computing, and development of sophisticated inferential tools now makes optimal sequential decision-making a realistic goal in medicine. Consequently, with the aim of advancing evidence-based medicine, researchers aim to identify decision rules that tailor patient care to patient specific characteristics through time; these rules are termed dynamic treatment regimes (DTRs). Many of the established methods for DTRs rely on the frequentist paradigm, however Bayesian methods have much to offer as they allow for the appropriate representation and propagation of uncertainty, including the facility to make probabilistic statements about optimal treatment strategies. This thesis seeks to develop Bayesian methods to perform inference of optimal DTRs in order to advance the aims of precision medicine.

The first manuscript in this thesis develops a methodology for identifying optimal DTRs by proposing Bayesian dynamic marginal structural models (MSMs), models representing the expected outcome under adherence to DTRs in a family indexed by a parameter ψ. To infer about these models, Bayesian decision theory is used to motivate the maximization of an expected posterior utility with respect to an unknown parameter. Singly and doubly robust Bayesian inference for DTRs is also explored using posterior predictive inference and the Bayesian bootstrap. The consequences of this Bayesian approach in quantifying uncertainty about the optimality of a decision rule for specific patients, thereby allowing for personalized decision-making, are also examined. These methods are studied using simulations and through an example in HIV therapy, where we seek to tailor treatment, with the aim of minimizing liver scarring.

Although (Bayesian) dynamic MSMs allow for the identification of optimal regimes, these methods have limitations. Importantly, they require modeling of the function that maps regime indices ψ to the expected outcome under regime adherence; this function is known as the value function. A misspecified model for the value function may lead to bias in the estimated optimal DTR. To avoid this, an estimator for the value of a DTR can be paired with a grid search for the optimal regime, however this may be computationally intractable, especially if Bayesian methods are used requiring draws from a posterior distribution. The second manuscript address these challenges by examining whether computer experiments with Gaussian processes can be used to identify optimal DTRs, a procedure that requires emulating the value function by fitting a working model on an initial set of design points and sequentially sampling new points that are most informative about the optimum. This methodology allows for robust estimation of optimal DTRs by permitting more flexible models of the value function, all while using data more efficiently than a grid-search. We find that accounting for variability in the estimated value function can yield improved performance over a grid-search, particularly when the value function exhibits multimodality. We illustrate the use of these methods by analyzing trial data to determine if HIV therapy can be tailored on patient-specific characteristics.

The aim of the third manuscript is to present an application of the methods developed in the first two manuscripts and illustrate implementation using an open-source software package. For this, we make use of the trial data on HIV therapy, as in the second manuscript, but additionally incorporate a simulated component to the data to arrive at a two-stage sequential decision-making problem. We explore decision rules that tailor therapy based on patients' baseline and 20-week CD4 cell count, with the aim of maximizing CD4 cell count at 90 weeks. For this analysis all methods considered yield similar inference regarding the optimal DTR. This manuscript also showcases the BayesDTR package, developed to perform the required analyses.

Abrégé

Dans le but de faire progresser la médecine fondée sur les preuves, les chercheurs visent à identifier des règles de décision qui adaptent les traitements du patient aux caractéristiques spécifiques de celui-ci, à travers le temps. Ces règles sont appelées régimes de traitement dynamiques (RDT). Plusieurs des méthodes établies pour les RDT reposent sur le paradigme fréquentiste, pourtant les méthodes bayésiennes ont beaucoup à offrir car elles permettent de propager l'incertitude et la représenter adéquatement, tout en rendant possibles des déclarations probabilistes sur les stratégies de traitement optimales. Cette thèse vise à développer des méthodes bayésiennes qui peuvent inférer sur des RDT optimaux afin de faire avancer la médecine de précision.

Le premier manuscrit de cette thèse développe une méthodologie pour identifier les RDT optimaux en proposant des modèles structurels marginaux (MSM) dynamiques bayésiennes, qui sont des modèles représentant le résultat espéré après adhésion aux RDT dans une famille indexée par un paramètre ψ. Pour inférer à propos de ces modèles, la théorie de la décision bayésienne est utilisée pour motiver la maximisation d'une utilité espérée a posteriori par rapport à un paramètre inconnu. L'inférence bayésienne simplement robuste et doublement robuste pour les RDT est également explorée. Les conséquences de cette approche bayésienne sur la quantification de l'incertitude de l'optimalité d'un RDT pour des patients spécifiques sont également examinées; ceci permet une prise de décision personnalisée. Ces méthodes sont étudiées à l'aide de simulations et à travers un exemple qui cherche à adapter le traitement des patients avec un objectif de minimiser les cicatrices hépatiques.

Les MSM dynamiques (bayésiens) nécessitent la modélisation de la fonction qui associe chaque régime indexé par ψ au résultat espéré après adhésion au régime, cette fonction est connue comme la fonction de valeur. Un modèle mal spécifié pour cette fonction peut induire un biais dans l'estimation du RDT optimal. Pour éviter cela, un estimateur de la valeur d'un RDT peut être jumelé avec une recherche par quadrillage pour trouver le régime optimal, mais cela peut exiger des calculs insolubles quand des méthodes bayésiennes sont utilisées. Le deuxième manuscrit vise à relever ces défis en étudiant l'usage de processus gaussiens pour identifier le RDT optimal, une procédure qui nécessite d'émuler la fonction de valeur en estimant un modèle préliminaire se basant sur un ensemble de points initiaux et en échantillonnant séquentiellement de nouveaux points. Cette méthodologie permet une estimation robuste des RDT optimaux à travers des modèles plus flexibles de la fonction de valeur. Nous constatons que la prise en compte de la variabilité dans la fonction de valeur estimée peut engendrer des gains de performance par rapport à une recherche par quadrillage, en particulier quand la fonction de valeur est multimodale. Nous illustrons l'utilisation de ces méthodes en analysant les données d'un essai clinique pour déterminer si le traitement anti-VIH peut être adapté aux caractéristiques du patient.

Le troisième manuscrit présent une application des méthodes développées dans cette thèse et illustre l'implémentation de ces méthodes en utilisant un logiciel a source ouvert. Pour cela, nous utilisons les données de traitement pour le VIH, comme dans le deuxième manuscrit, mais aussi nous incorporons un composant simulé pour obtenir un problème de prise de décisions séquentiel à deux étapes. Nous explorons des règles qui adaptent le traitement en fonction du nombre de cellules CD4 au début de l'étude et à la 20ème semaine, avec un but de maximiser le nombre de cellules CD4 à la 90ème semaine. Ce manuscrit présent aussi la librairie R BayesDTR, qui a été développée pour effectuer les analyses requises. Nous concluons que toutes les méthodes considérées mènent à une inférence similaire sur le RDT optimal.

Table of contents

List of Tables

List of Figures

Abbreviations

ART antiretroviral therapy

CDF cumulative distribution function

$\mathcal{DP}$ Dirichlet process

DR doubly robust

DTR dynamic treatment regime

dWOLS dynamic weighted ordinary least squares

EI expected improvement

$\mathcal{GP}$ Gaussian process

HE$\mathcal{GP}$ Gaussian process model with heteroskedastic additive noise

HIV human immunodeficiency virus

HM$\mathcal{GP}$ Gaussian process model with homoskedastic additive noise

Int$\mathcal{GP}$ Gaussian process model with no additive noise

IPW inverse probability of treatment weighting

ITR individualized treatment rule

MAP maximum a posteriori

MCMC Markov Chain Monte Carlo

MSM marginal structural model

NA-ACCORD North American AIDS cohort collaboration on research and design

SMART sequential multiple assignment randomized trials

SMD standardized mean difference

SUTVA stable unit treatment value assumption

Chapter 1

Introduction

Precision medicine seeks to tailor treatments to patients based on their individual characteristics in order to improve health outcomes. The main challenge in this field is to identify sets of variables, called tailoring variables, that allow us to distinguish which patients would benefit most from which types of therapies and at what time. The practice of precision medicine can be understood to be feasible by first noting that patients differ due to a variety of factors including genetic, demographic, and clinical measures in addition to their medical history. This observation, in conjunction with the understanding that, crucially, patients' outcomes are not solely associated with the treatments they receive but also with their individual characteristics, clarifies why the aim of tailoring therapy is realistic. Indeed, the study of precision medicine begins with studying how these factors modify the effect of treatment. However, identifying effect modification is not sufficient to identify variables on which to tailor treatment. Generally, we require that for some values of the tailoring variables, the expected benefit under one therapy is greater than the expected benefit under other therapies, and vice versa for other levels of the tailoring variables. That is, if a variable modifies the effect of treatment but the modification is such that all patients would still benefit most from only one level of therapy, then this effect modification does not allow for tailoring [Gunter et al., 2007].

Precision medicine is guided by meeting clinicians' decision-making needs when treating (chronic) diseases, where decisions can be made at different times point, for example during fixed treatment schedules, or at turning points during disease progression. At each decision point, all relevant treatment options and all available information should be considered, while aiming to deliver optimal care. To meet clinicians' decision-making needs in clinical practice, we must formalize a framework for the study of this decision-making process. This is achieved through the study of statistical methods for dynamic treatment regimes (DTRs) [Murphy et al., 2001, Robins, 2004], which are sets of decision rules that inform a decision maker how to treat patients based on all pertinent data on a patient. Most often, we are interested in estimating optimal DTRs, meaning DTRs that optimize the expected outcome if everyone in the population under study received treatment according to the optimal decision rule.

Fundamentally, questions about DTRs are causal questions as they require asking about the effects of interventions, for example, "what is the effect of following DTR A vs. DTR B?". In keeping with the majority of statistical literature, the conception of causality in this thesis comes from understanding the effects of causes, not the causes of an effect [Holland, 1986]. That is, first we identify an intervention to study and then we examine the consequences it has on an outcome of interest. Historically, data from clinical trials have been the most accepted sources of evidence for answering causal questions, but studying DTRs involves complex decision rules, with possibly time-varying treatment. Running trials that allow us to ask causal questions about DTRs can be very costly, and we must often look to non-experimental or observational data to gain insights about the effects of DTRs, while acknowledging that these data are susceptible to a variety of biases.

The majority of methodological developments in the study of DTRs have been within the frequentist inferential framework, a framework which views probabilities as being limiting frequencies under an infinite set of identical experiments. This approach to inference views parameters, the possible targets of inference, as fixed but unknown; probability statements can be made for estimators of these parameters, but not for the parameters themselves. However, the study of DTRs may also benefit from the use of a Bayesian lens. This framework views probabilities as dependent on available information, thereby allowing for differences in belief about a quantity of interest. This approach to inference has a more expansive view of what uncertainties can be quantified using probabilities. More practically, this means that Bayesian methods view unknown parameters as random variables, and inference about these parameters is based directly on their probability distribution. This in turn means that a Bayesian framework is a more natural approach for answering some scientific questions; for example, it can naturally incorporate model uncertainty in the inferential process. A Bayesian framework has additional advantages by allowing us to incorporate prior knowledge about quantities of interest, like treatment effects, or by not requiring asymptotic assumptions in order to perform inference. Generally, both the frequentist and Bayesian inferential approaches have merit, and it is a worthy pursuit to advance each.

This thesis seeks to advance Bayesian inference for optimal DTRs through the development of three manuscripts. It begins with a literature review in chapter 2, contrasting some of the current methods, both frequentist and Bayesian, available for inference about DTRs. Chapter 3 then presents a manuscript that builds on work by Saarela et al. [2015b] to achieve Bayesian estimation of dynamic marginal structural models. These models, originally proposed in a frequentist framework, have the benefit of requiring that only a mean model be specified, all while working within an infinite-dimensional set of distributions. A Bayesian analogue of this procedure is proposed allowing for Bayesian semiparametric inference of optimal DTRs. Bayesian predictive inference is also explored for singly and doubly robust estimators, thereby allowing us to consider individualized inference by ascertaining the probability that a specific patient is following an optimal therapy. An example in HIV care is examined to illustrate the use of these methods. Chapter 4 is motivated by the acknowledgment that frequentist and Bayesian dynamic MSMs have some limitations, and addresses these by focusing on an approach that allows for flexible modeling of the value function, i.e. the function that maps a regime of interest to the expected outcome under regime adherence. The proposed methodology uses the structure of computer experiments, where an initial set of design points are used to fit a working model and further points are gathered sequentially, with the aim of efficiently identifying an optimal function value. The model fits are predicated on prior assumptions on the form of the value function, using Gaussian processes. Further discussion is provided as to how the computer experiment approach is a natural example of how Bayesian uncertainty quantification can be used to express uncertainty about a possibly deterministic system. Lastly, the use of these methods is examined by using clinical trial data for HIV therapy. Chapter 5 demonstrates the use of the proposed methods using the HIV trial data from chapter 4, with an additional simulated component that allows for considering a two-stage sequential decision-making problem. Some of the methods discussed in chapters 3 and 4 are nuanced, and this chapter allows us to consider in detail how they should be applied. The analysis of these data is further facilitated by the `BayesDTR R` package, developed to accompany the methods in this thesis. Chapter 6 provides a conclusion for the work in this thesis, where a review of the contained contributions is provided, possible limitations are discussed, and future work is considered.

Note that chapters 3-5 are written as self-contained manuscripts, with a preamble detailing the contributions of each work. Although the same notation is largely maintained across all manuscripts, it is occasionally the case that notation is altered slightly from one manuscript to another; these variations are noted and clarified in the preamble to each manuscript. Chapter 3 has been published in *Biostatistics*, chapter 4 will be submitted to a statistical journal soon after the submission of this thesis, and chapter 5 is under review in a statistical journal.

Chapter 2

Literature review

This thesis focuses on statistical methods for conducting scientific studies that explore how medical interventions can be assigned to patients using all relevant information. These methods contribute toward the goal of delivering tailored care by moving away from a one-size-fits-all approach in medical therapy. The aim of delivering precision care is of importance in both acute care and chronic care settings. Generally, this thesis focuses on the development of methods that allow for the tailoring of care for treatments that vary over time, for patients requiring chronic care. However, time-invariant treatments, likely arising from providing acute care, are seen as a special case. In the context of tailoring time-varying treatments, further distinctions between treatment regimes should be made. First, a *static treatment regime* is a sequence of treatments that varies through time but that can be specified at study start; it does not change with changing patient covariates and can only depend on baseline covariates. In contrast, a *dynamic treatment regime* (DTR) involves a treatment sequence that cannot be determined at study start as it depends on patients' changing covariates. Although the treatment sequence is not determined at study start, the treatment rules are fixed, and so the information used to allocate components of the sequence are known in advance. These rules can depend on baseline and time-varying covariates. Murphy

[2003] defines a DTR as a function that takes treatment and covariate history, including baseline covariates, and that outputs a treatment recommendation. Of importance is the identification of an optimal DTR, that is, the DTR that optimizes the expected outcome under adherence to a regime; we refer to this expectation as the value of a DTR. The study of DTRs is the study of sequential decision-making, where at each stage of the decision process we seek to use all available information to arrive at an (optimal) decision or therapy. Variables that help guide which treatment is right for whom are generally termed tailoring variables. Tailoring therapy in a single-stage setting is also an important goal, and methods for DTRs can be useful in addressing these problems as well; treatment decision rules in a single-stage setting are termed individualized treatment rules (ITRs).

Both experimental and observational data can be used to infer about DTRs. To produce experimental data for causal inference of DTRs requires the use of sequential multiple assignment randomized trials (SMARTs) [Murphy, 2005a]. These trials can be costly to perform; consequently, we cannot solely rely on data arising from SMARTs to study DTRs. Alternatively, observational data can be used, but care should be taken as these data are susceptible to a variety of biases. One well known bias that can arise in the use of observational data is confounding bias, which can distort the estimate of the true effect of a treatment on an outcome. This bias is a result of not controlling for confounders, variables that are both a cause of treatment and of the outcome. In cross-sectional problems, confounding bias can be accounted for using standard regression methods, however this is not always the case in the data dependencies that are of interest in sequential decision-making problems.

To identify optimal DTRs requires asking "what if" types of questions, thereby emphasizing the need for methods to draw causal inference about these DTRs. Most commonly, we are concerned about performing causal inference about the value of a given DTR or about the optimal DTR over all classes of DTRs or over a specific set of regimes. Section 2.3, provides the formal assumptions required to perform causal inference using the methods described in this thesis. Informally, these assumptions require that 1) we consider a problem where the outcome of one individual is not affected by the treatment assignment of other individuals (known as SUTVA), 2) that there are no unmeasured confounders, and 3) that all treatment patterns can be observed for all types of patients (known as positivity).

To formalize the decision-making process, we consider a sequence of observed treatments $\bar{z} = (z_1, ..., z_K)$, where K is the number of decision points in the sequential problem and where $z_k \in \{0, 1\}$ are binary treatments. The case of non-binary treatment can also be considered, although this involves some additional modeling challenges. Prior to each decision point, we observe covariates x_k, which we assume contain all relevant confounders or tailoring variables. The entire covariate history is given by $\bar{x} = (x_1, ..., x_K)$. By $\bar{z}_k$ and $\bar{x}_k$, we refer to treatment and covariate histories up to and including stage k. We use the vector-valued function $g(\cdot)$ to denote DTRs, a function that takes as input patient information and that outputs a treatment recommendation for each stage of the process. We refer to the DTR-enforced treatments history by $g(\bar{x}) = (g_1(x_1), ..., g_K(\bar{x}))$. These are the treatment recommendations that are provided by a DTR, and can differ from z_k which were assigned according to standard practice. To denote treatment recommendations up to stage k, we write $\bar{g}_k(\bar{x}_k) = (g_1(x_1), ..., g_k(\bar{x}_k))$. We focus only on deterministic DTRs, that is DTRs that assign treatment using deterministic rules of how patient covariates are mapped to treatment recommendations. Lastly, we use y to denote the final outcome of the decision-process. It is this outcome that we are seeking to optimize (without loss of generality, this review focuses on maximization). In particular, we seek g_{opt} that, when a population of interest is treated according to this rule, yields the optimal expected outcome. Counterfactuals are denoted by superscripts, for example the expected counterfactual outcome under adherence to a regime g can be denoted by $E[Y^g]$; in Bayesian settings, draws from posterior distributions are denoted with *. Throughout, we will also refer to the function mapping a regime g to the expected outcome under adherence to g as the value function, with the expected outcome under adherence to g being referred to as the regime value.

Some of the methods to be discussed in this review require adapting the notation above. As noted above, at stage k a regime g recommends a treatment by using patient covariate information up to stage k and treatment information up to stage $k-1$. If we define a history vector $h_k = (\bar{x}_k, \bar{z}_{k-1})$, note that stage k treatment is not included in this history, then at stage k, this regime should take information h_k to output a treatment recommendation as $z_k = g(h_k)$. In the context of a patient following a regime g through stage k, then it is enough to use the covariate history $\bar{x}_k$ to output a treatment recommendation as $z_k = g(\bar{x}_k)$, as previous treatments were assigned based on patient covariates. Further discussion on this point can be found in Tsiatis et al. [2019].

With the required notation, it is now possible to examine Figure 2.1 which shows one possible relational structure, known as a directed acyclic graph [Greenland et al., 1999], for data that may be available to estimate a DTR. We measure covariates x_1, x_2 which can help decide on treatment assignment z_1, z_2. There may be additional unobserved variables in the problem u. Of course, in practice this is not the only structure possible, for example x_1 can be a direct cause of x_2 and of y etc. Chakraborty and Moodie [2013] provide a diagram of all possible causal paths available in this decision problem. In a longitudinal setting, like in Figure 2.1, concerns about confounding increase as now we are also concerned about confounders that depend on previous treatment, a phenomenon known as treatment-confounder feedback that makes the confounder a mediatior for previous stage treatments. In the example in Figure 2.1, x_2 is a confounder that mediates z_1's effects on y. When causal structures like this arise, using regression modeling to adjust for confounders is not sufficient as it can lead to *over-adjustment bias* [Schisterman et al., 2009], a bias that occurs when conditioning on downstream confounders leads to blocking the effect of past treatments thereby underestimating their effect. Another bias that can arise when conditioning on confounders in a longitudinal setting is *collider-stratification bias* [Greenland, 2003] which says that conditioning on a confounder that shares a common cause, u, with the outcome will confound the relationship between previous treatments and the outcome of interest. Further discussion of these issues can be found in Chakraborty and Moodie [2013], but the conclusion that we arrive at is that to infer about optimal DTRs, we must look beyond traditional regression methods. In the following sections, we examine specific methods designed for inference of DTRs, created by the need to identify other tools for inferring about time-varying treatments and DTRs.

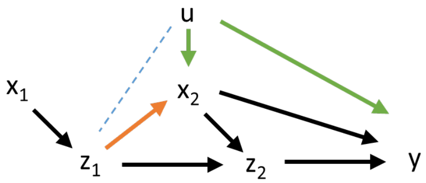


Figure 2.1: Time-varying confounding with treatment confounder feedback in a two-stage decision process.

The structure of the remaining literature review is now laid out. In section 2.1 we discuss marginal structural models — one tool that can be used to infer about time-varying treatments, including with regard to DTRs. These models benefit from the fact that they yield interpretable decision rules, a challenge for other frequentist and Bayesian models. In section 2.2, we discuss general frequentist methods for DTRs and contrast some of the approaches. Section 2.3 allows us to formally state the required assumptions for causal inference. Section 2.4 introduces some concepts in Bayesian inference, which further leads to the Bayesian DTR literature in section 2.5. Section 2.6 discusses methods for Bayesian optimization.

2.1 Marginal Structural Models (MSMs)

One common set of models used for inferring about time-varying treatments are MSMs. As these are of substantial focus in this thesis, we begin with an overview of their use to specify treatment effects in longitudinal static regime settings, and we then examine their use for DTRs. Indeed, MSMs were first developed to parameterize the effect of static regimes [Robins et al., 2000]; specifically, an MSM is a model for a counterfactual $y^{\bar{z}}$ under a sequence of treatments $\bar{z}$. It is marginal because it models a marginal quantity that is not conditional on patient covariates, except possibly for baseline covariates; it is structural because it posits a model for a counterfactual. These models are important because, when correctly specified, they yield meaningful estimands for the causal effect of a sequence of treatments $(z_1, z_2, ..., z_K)$. Most often, we are interested in marginal structural mean models, and to correctly specify these requires correctly specifying a model for $E[Y^{\bar{z}}]$ under different treatment patterns $\bar{z}$. An example of such a model is $E[Y^{\bar{z}}] = \bar{z}^T\beta$, where β is an unknown parameter of dimension p. In this thesis, the term MSM refers exclusively to marginal structural mean models. In this approach, one is required to specify the structure of the mean of $Y^{\bar{z}}$, while leaving the rest of the distribution unspecified.

Now, positing a model for an MSM is relatively straightforward, however, what is not immediate is understanding how to go about estimating the model parameters. The most common way by which to estimate parameters in MSMs is via inverse probability of treatment weighting (IPW). For this approach to work, the treatment assignment mechanism $\bar{z}|\bar{x}$ should be known or consistently estimated. However, no additional assumptions are placed on the distribution of $Y^{\bar{z}}$. If the counterfactual mean is modeled as $\bar{z}_i^T\beta$, then the estimating equations that must be solved to use IPW are given by

$$\sum_{i=1}^{n} \frac{\partial(y_i - \bar{z}_i^T\beta)}{\partial\beta} \frac{1}{p(\bar{z}_i|\bar{x}_i)} (y_i - \bar{z}_i^T\beta) = \mathbf{0}_{p\times 1}.$$

The solution $\hat{\beta}_n$ to these equations benefits from asymptotic normality and consistency, so long as the treatment assignment model is consistently estimated and the counterfactual mean model is correctly specified; the subscript n references the fact that the solution depends on the observed data. Furthermore, for the solution of these equation to represent the causal parameters of interest, we require all three causal assumptions: SUTVA, no unmeasured confounders, and positivity. Treatment assignment probabilities may be decomposed into stage-specific contributions. For example, in a two-stage setting, we obtain

$p(z_2, z_1|x_1, x_2) = p(z_2|z_1, \bar{x}_2)p(z_1|x_1)$. It is possible to ask questions about optimal static regimes, and these may be identified by finding the fixed sequence of treatments that optimize the expected counterfactual outcome.

Although straightforward to implement, IPW estimators can exhibit high variability, especially if the treatment probabilities are small, thereby yielding large weights. A more efficient estimator that can be considered is the augmented IPW estimator, also known as a doubly robust (DR) estimator [Bang and Robins, 2005, 2008]. Briefly, a DR estimator requires modeling: 1) the outcome as a function of exposure and covariates, and 2) the treatment assignment process. What is useful about DR estimators is that so long as either the outcome or treatment process is consistently modeled, consistent estimation of the causal treatment effects is attained. Other methods to infer about MSM parameters include a Bayesian approach proposed by Saarela et al. [2015b] and an approach using targeted maximum likelihood estimation proposed by Rosenblum and van der Laan [2010].

MSMs are useful for estimating the effect of static regimes, but they do not allow us to ask about dynamic regimes. For MSMs that infer about DTRs, Murphy et al. [2001] first showed how to estimate the marginal mean outcome under adherence to a single DTR. Additional work by van der Laan and Petersen [2007] and Orellana et al. [2010a,b] extend inference for MSMs to allow for joint estimation of the value of regimes in a family, thereby yielding dynamic MSMs. Interest lies in using these dynamic MSMs to compute the average causal effect of being assigned and adhering to a specified regime in a family $\mathcal{G} = \{g^{\psi}; \psi \in \mathcal{I}\}$, indexed by a parameter ψ. Then, the first challenge in this modeling approach is to connect the mean counterfactual outcome under adherence to regimes in $\mathcal{G}$ via a parameter β in a model $h(\psi; \beta), \beta \in \Re^p$. We begin with an account of how to estimate the parameter β in a singly robust setting, and we then examine doubly robust estimation. To denote counterfactual outcomes in which a patient receives regime g^{ψ}, we omit the g for ease of notation; that is, we define $E[Y^{g^{\psi}}] = E[Y^{\psi}]$. As we are dealing with deterministic DTRs then following Orellana et al. [2010a] the counterfactual distribution $(\bar{Z}^{\psi}, \bar{X}^{\psi})$ is such that the treatment probability at stage k is

$$p(Z_{\psi}^{\psi} = z_k | \bar{X}_k^{\psi} = \bar{x}_k, \bar{Z}_{k-1}^{\psi} = \bar{z}_{k-1}) = \mathbb{1}_{g_k^{\psi}(\bar{x}_k)}(z_k),$$

where the indicator function is 1 when $z_k = g_k^{\psi}(\bar{x}_k)$ and zero otherwise. This leads to the following weight definition for information up to stage k:

$$w_k^{\psi}(\bar{x}_k, \bar{z}_k) = \frac{\prod_{j=1}^{k} \mathbb{1}_{g_k^{\psi}(\bar{x}_k)}(z_k)}{\prod_{j=1}^{k} p(z_k | \bar{x}_k, \bar{z}_{k-1})}.$$

This weight is the ratio whose numerator is an indicator function, which is 1 when patients receive treatment according to the DTR g^{ψ} up to stage k and 0 otherwise, and whose denominator is the treatment probability in the study population through stage k. Effectively, w_k truncates or censors subjects who are not adherent to regime g^{ψ} through stage k. Note that the superscript in the weight does not indicate a counterfactual, but rather it simply references the fact that the form of the weight depends on the regime under consideration.

With the weight definition in place, we can consider estimating the parameters in $h(\psi; \beta)$. An example of such an MSM can be $h(\psi; \beta) = \beta_0 + \beta_1 \psi + \beta_2 \psi^2$. We see that dynamic MSMs allow for the possibility of parameter sharing across regimes. Assuming a correctly specified model with true parameter β^{+}, in addition to SUTVA, no unmeasured confounders, and positivity, the following identity is obtained:

$$E\left[w_k^{\psi}(Y - h(\psi; \beta^{+}))\right] = 0 \tag{2.1}$$

which motivates the the estimating function

$$S(\psi, \beta, b) = \sum_{\psi \in \mathcal{I}_0} \sum_{i=1}^{n} w_{i,K}^{\psi} b(\psi) \left(y_i - h(\psi; \beta)\right), \tag{2.2}$$

where $b(\psi)$ is a vector-valued function of equal dimension to β and commonly taken to be $\frac{\partial h(\psi;\beta)}{\partial \beta}$, and where $\mathcal{I}_0 \subset \mathcal{I}$ is a finite set of regimes for which the positivity condition holds. Following the frequentist semiparametic inferential approach, we search for the IPW esimator $\hat{\beta}_n$ such that $S(\psi, \hat{\beta}_n, b) = 0$. Although solving this equation seems challenging, it can be done using a data-augmentation technique described in Cain et al. [2010], whereby a new row of data is created for every regime to which a patient adheres. The solutions of this estimating equation are asymptotically normal with a variance formula that can be arrived at using a sandwich variance estimator [Orellana et al., 2010a,b]. The use of these dynamic MSMs has been explored in Cain et al. [2010] for HIV therapy and in Shortreed and Moodie [2012] for psychiatric therapy, for example.

A doubly robust estimator for the marginal counterfactual outcome under adherence to a DTR and for parameters in $h(\psi;\beta)$ has also been developed. The former was first proposed by Murphy et al. [2001] who consider the doubly robust estimator for the marginal mean; later this was extended to a family of regimes by Orellana et al. [2010a,b]. The DR estimation requires identifying a series of conditional outcome models, so that consistent inference is attained when either the treatment models or the outcome models are correctly specified. With this aim in mind, consider a sequence of conditional outcomes ϕ_k defined as

$$\phi_K^{\psi}(\bar{x}_K) = E[Y|\bar{X}_K = \bar{x}_K, \bar{Z}_K = \bar{g}_K^{\psi}(\bar{x}_K)] \text{ for } k = K \text{ and as}$$

$$\phi_k^{\psi}(\bar{x}_k) = E[\phi_{k+1}^{\psi}(\bar{x}_{k+1})|\bar{X}_k = \bar{x}_k, \bar{Z}_k = \bar{g}_k(\bar{x}_k)] \text{ for } k = 1, ..., K-1.$$

Orellana et al. [2010a] show that for each k, $\phi_k^{\psi}(\bar{x}_k) = E\left[Y^{\psi}|\bar{X}_k = \bar{x}_k, \bar{Z}_k = \bar{g}_k(\bar{x}_k)\right]$. Thus, the ϕ_k^{ψ}s can be interpreted as the expected counterfactual outcome had a patient followed regime g^{ψ} throughout the entire study period conditional on information up to time k. Modeling these outcomes with a parameter τ such that $\phi_k^{\psi}(\bar{x}_k) = \phi_k^{\psi}(\bar{x}_k;\tau)$ allows us to consider the β-specific estimating function,

$$\begin{aligned} S_\beta(\psi;\beta,\gamma,\tau) =& \{\phi_1^\psi(\bar{x}_1;\tau) - h(\psi;\beta)\} \\ & + \sum_{k=1}^{K-1} w_k^\psi(\gamma)(\phi_k^\psi(\bar{x}_k,\tau) - \phi_{k-1}^\psi(\bar{x}_{k-1},\tau)) + w_K^\psi(\gamma)(y - \phi_K^\psi(\bar{x}_k,\tau)). \end{aligned}$$

Orellana et al. [2010a] show that under a correctly specified model $h(\psi;\beta)$, $S_\beta(\psi,\beta,\hat{\gamma}_n,\hat{\tau}_n)$ is an unbiased estimator of zero either when the set of treatment models or when a set of outcomes models is correctly specified and consistently estimated, where $\hat{\gamma}_n$ and $\hat{\tau}_n$ are the parameters estimated in the treatment or outcomes models, respectively. This characteristic provides us with doubly robust estimation for the β^+ that we seek by solving $\sum_i^n \sum_{\psi\in\mathcal{I}_0} b_i(\psi) S_{i,\beta}(\psi,\beta,\hat{\gamma}_n,\hat{\tau}_n) = 0$. Standard arguments show that $\hat{\beta}_n$ is consistent and asymptotically normal [Orellana et al., 2010a]. Tsiatis et al. [2019] have an account of the several procedures that can be used to fit these models in practice. A discussion on the use of singly versus doubly robust estimators can be found in Kang and Schafer [2007] and in the commentaries to that article.

Dynamic MSMs lend themselves to inference with interpretable decision rules, however accurate inference is only guaranteed when $h(\psi;\beta)$ is correctly specified as well as when treatment or outcome models are correctly specified. These offer an approach to inference that does not require specifying parametric likelihoods, unlike a typical, fully parametric Bayesian approach which would usually require that full distributions for the data-generating mechanism to be specified. In what follows, we now examine other frequentist methods for DTRs. In particular, we distinguish between value-search approaches and regression-based methods.

2.2 Frequentist Methods for DTRs

A variety of other frequentist methods for estimating the effect of dynamic treatment regimes have been proposed. For example g-methods including g-computation [Robins, 1986] and g-estimation of structural nested models [Robins, 1993] may be used for this purpose. Other ways by which to identify optimal DTRs include Q-learning [Murphy, 2005b], outcome weighted learning [Zhao et al., 2012], dynamic weighted ordinary least squares [Wallace and Moodie, 2015], and residual weighted learning [Zhou et al., 2017]. Bayesian methods have also been proposed, although this inferential approach has received significantly less attention in the DTR literature. For example, Arjas and Saarela [2010] performed backward induction using Bayesian nonparametric regression models, Saarela et al. [2015a] took a parametric approach using Bayesian predictive inference, Xu et al. [2016] used Bayesian nonparametrics in a survival context, where patients could randomly transition between disease states, and Murray et al. [2018] adapted Q-learning to a Bayesian setting. In this section we examine commonly used frequentist methods, and defer discussion on Bayesian methods to section 2.5 so that we can first introduce concepts in Bayesian inference in section 2.4.

Ways by which to estimate the value of DTRs and to identify optimal DTRs are commonly placed in two categories: value-search approaches or regression-based approaches. Value-search methods look to directly optimize the value $E[Y^g]$, like the previously discussed dynamic MSMs, whereas regression-based approaches model outcomes conditional on stage-specific patient information which can subsequently be used to identify optimal DTRs. We begin with a discussion of value-search methods, and we then examine regression-based methods.

A related approach to that of dynamic MSMs is that of Zhang et al. [2013], who propose that in the family of regimes of interest $\mathcal{G}$, an estimator for $E[Y^{g^\psi}]$ be chosen and subsequently maximized as a function of ψ; this avoids positing a mean model $h(\psi;\beta)$. The authors suggest using a genetic algorithm to maximize the IPW estimator and the augmented IPW estimator.

Another related approach that seeks to optimize a worst-case value function is that of Mo et al. [2021], who estimate distributionally robust ITRs in the realistic setting where data used to estimate optimal DTRs, termed training data, are not necessarily distributed exactly like data where the ITR will be deployed. Consequently, the authors seek to maximize a worst-case value function in a set of distributions that are similar to the training data distribution.

Methods for value-search estimation have also been cast as a weighted classification problem, thereby allowing for the use of statistical learning methods. These weighted learning approaches include outcome weighted learning (OWL) [Zhao et al., 2012] for individualized treatment rules and residual weighted learning [Zhou et al., 2017] as an improvement to OWL. OWL was later extended to a multistage setting by Zhao et al. [2015] with methods termed backwards outcome weighted learning and simultaneous outcome weighted learning. We begin by describing OWL in a single-stage setting, as many of the core ideas of this group of methods can be understood from this setup. For the purposes of exploring these methods, we temporarily change treatment coding to $z \in \{-1, 1\}$, which differs from the $\{0, 1\}$ coding used in the rest of this thesis. Like with other value-search methods, an importance sampling argument yields the following expression for the expected outcome under adherence to an ITR g:

$$E[Y^g] = E\left[\frac{\mathbb{1}(Z = g(X))}{p(Z|X)} Y\right].$$

Interest lies in directly identifying $g_{opt}(x)$ that satisfies $g_{opt} = \arg\max_g E[Y^g]$. It can be shown that this maximization problem is equivalent to finding g_{opt} that minimizes

$$\mathcal{R}(f) = E\left[\frac{\mathbb{1}(Z \neq sign(f(X)))}{p(Z|X)} Y\right], \tag{2.3}$$

where we have re-expressed $g(x)$ as $sign(f(x))$ for some decision function f. Consequently, the problem now centers around finding f_{opt} and this can be seen as a weighted classification problem: when the treatment that a patient receives does not match the treatment suggested by the DTR, then this is an instance of misclassification and contributes to the total missclassification error. Zhang et al. [2012] have also studied value-search estimation though a classification lens. Empirically, the objective function in equation 2.3 can be approximated by $\sum_{i=1}^{n} \mathbb{1}(z_i \neq \mathbb{1}(f(x_i) > 0))y_i/p(z_i|x_i)$. It is evident that to minimize this function, the DTR should suggest treatments that match the observed treatments for patients with small weighted outcomes and vice-versa. Minimizing this objective function is challenging due to the indicator function, which here represents a 0-1 loss. Consequently, it is replaced with a hinge loss $\phi(t) = (1-t)^+ = \max(0, 1-t)$, and the complexity of the decision function is penalized by aiming to minimize the following regularized expression

$$\frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{p(z_i|x_i)}(1 - z_i f(x_i))^+ + \lambda_n||f||^2 \tag{2.4}$$

with respect to f. Under some assumptions regarding the class of functions to which f belongs, flexible decision rules can be estimated by minimizing expression (2.4) via support vector machine techniques. Some consistency guarantees are described in Zhao et al. [2012], who show that asymptotically minimizing the regularized problem is equivalent to solving the original value maximization problem. Generally, these methods require the SUTVA, no unmeasured confounders, and positivity assumptions to arrive at inference. Using this theme of minimizing the missclassification error, Zhao et al. [2015] extend OWL to a multi-stage setting. Backwards outcome weighted learning identifies an optimal stage k decision rule using a weighted classification problem assuming patients have followed the optimal treatment rule from stake $k+1$ onward. Contrastingly, simultaneous outcome weighted learning considers the weighted classification problem across all stages and simultaneously solves for all stage-specific optimal decision rules by adding a regularization term for each stage-specific decision function f_k. OWL faces several limitations; most importantly, although the optimal decision rule g_{opt} is invariant to constant shifts of the outcome, $y+c$, the estimated optimal decision rule $\hat{g}_{opt}$ is not. Zhou et al. [2017] examine the optimal constant c, or function $s(x)$, by which to shift y and propose residual weighted learning as a way of obtaining improved finite sample performance. For this purpose, authors show that for any measurable function $s(x)$,

$$E\left[\frac{Y - s(X)}{p(Z|X)}\mathbb{1}_{(Z\neq g(X))}\right] = E\left[\frac{Y}{p(Z|X)}\mathbb{1}_{Z\neq g(X)}\right] - E[s(X)]. \tag{2.5}$$

Thus, shifting y by $s(x)$ does not change the minimizer g_{opt} of the problem. Zhou et al. [2017] advocate for choosing $s(x) = E\left[Y/(2p(Z|X))|X = x\right]$, which may be estimated via a regression model and can reduce the estimator variance. A regularized formulation of this problem again leads to consistent inference. These weighted learning approaches yield the possibility of leveraging tools from the classification and prediction literature, but their main limitation is that they do not have a straightforward manner by which to quantify uncertainty about the obtained decision rule, a crucial element of any statistical analysis.

In contrast to value-search methods, regression-based approaches require that we model the outcome of interest directly, conditional on stage-specific information. To examine these methods, we return to using the $\{0, 1\}$ coding for treatments. A commonly used regression based approach is Q-learning, introduced into the statistical literature by Murphy [2005b] and Chakraborty et al. [2010] with Moodie et al. [2012] further exploring its use with observational data. This method begins by defining the stage-specific Q-functions, which in a two-stage setting are given by:

$$\begin{aligned}
Q_2(h_2, z_2) &= E[Y|H_2 = h_2, Z_2 = z_2],\\
Q_1(h_1, z_1) &= E[\max_{z_2} Q_2(H_2, z_2)|H_1 = h_1, Z_1 = z_1].
\end{aligned}$$

If these quantities were known, then the decision problem could be solved backward by identifying the optimal rule at stage 2, and then identifying the optimal rule at stage 1 assuming that the optimal stage two treatment decision was followed. That is, the optimal treatment is given by, $z_k^{opt} = \arg\max_{z_k} Q_k(h_k, z_k)$. This procedure is known as backwards induction. However, as these Q-functions are unknown, they must be modeled and their parameters estimated; they are expected outcomes conditional on stage-specific information, thus can be tackled via the familiar tool of regression.

To fit models for the Q-functions requires considering that at each stage k there are covariates and previous treatments that may be predictive of the outcome, and that there are also those that may modify the effect of stage k treatment. We can separate these two histories, with h_{k0} denoting previous treatments and covariates that are predictive of the stage k outcome and h_{k1} those that modify stage k treatment. These two sets of variables do not have to be mutually exclusive; indeed, it is typically the case that all variables in h_{k1} are contained in h_{k0} to ensure strict hierarchy (i.e all main effects of interactions are retained in a model). Consequently, models for the Q-functions can be given by: $Q_k(h_k, z_k) = \beta_{k0}^T h_{k0} + (\beta_{k1}^T h_{k1}) z_k$, $k = 1, ..., K$. These models can be fit using standard ordinary least squares regression. Of course the fitting procedure also requires estimating the outcomes for stages $k = K-1, ..., 1$, known as pseudo-outcomes and given by $\tilde{y}_k = \max_{z_k} Q_k(h_k, z_k)$. Then, the fitting process proceeds backward: estimate a model for Q_K, use it to predict a pseudo-outcome $\tilde{y}_K$ to be used as the outcome to fit a model for Q_{K-1}, etc. In a binary treatment setting, the optimal regime for the models considered above says to treat when $\beta_{k1}^T h_{k1} > 0$. Thus, the pseudo-outcomes can be computed in a straightforward manner as $\tilde{y}_1 = \beta_{k0}^T h_{k0} + (\beta_{k1}^T h_{k1}) \mathbb{1}(\beta_{k1}^T h_{k1} > 0)$. Consistent estimation requires that these models be correctly specified, which may mean specifying complex models that result in complex decision rules that are hard for clinicians to interpret. Additionally, the no-unmeasured confounders and SUTVA assumptions are also required. Q-learning benefits from the fact that it is straightforward to implement in practice. An example of a practical application of Q-learning can be found in Krakow et al. [2017] who employ the method to identify treatment strategies for graft-versus-host disease with the aim of maximizing survival.

A related approach that additionally benefits from a double robustness property is dynamic weighted ordinary least squares (dWOLS) [Wallace and Moodie, 2015]. This is a method that allows for the use of a variety of weights including inverse probability of treatment weights and that, like Q-learning, requires performing stage-specific regressions, however these regressions are now weighted. The double robustness property arises from its similarity to g-estimation, a method to estimate optimal DTRs that has seen limited use in practice due to its complex formulation. Consequently, dWOLS offers the usability of methods like Q-learning with added robustness. This approach makes use of backward induction in the same way that Q-learning does, however it differs in the way that the pseudo-outcomes are computed. To understand this, we must define the blip functions $\gamma_k(h_k, z_k)$ [Robins, 2004] which represent the difference in expected outcome between patients who 1) receive treatment z_k at stage k as compared to patients who receives a reference treatment ($z_k = 0$) at stage k, 2) have the same treatment history through stage $k-1$, and 3) receive optimal treatment subsequently. Mathematically, this can be written as

$$\gamma_k(h_k, z_k) = E[Y^{\bar{z}_k, z^{opt}_{k+1},...,z^{opt}_K} - Y^{\bar{z}_{k-1},0,z^{opt}_{k+1},...,z^{opt}_K}|H_k = h_k], \text{ for } k = 1, ..., K, \qquad (2.6)$$

where z_j^{opt} are optimal treatments and z_j are the treatments actually received by the patient. Note that $\gamma_k(h_k, z_k)$ corresponds to the model $\beta_k h_{k1}$ previously discussed in the Q-learning approach, where the blip function was modeled linearly in that case. With this definition, the pseudo-outcome at stage k can be defined as $\tilde{y}_k = y + \sum_{j=k+1}^{K}(\gamma_j(h_j, z_j^{opt}) - \gamma_j(h_j, z_j))$. Effectively, this operation removes the effect of observed treatments from stage $k+1$ onward and adds the effect of optimal treatment, to arrive at the desired pseudo-outcomes. Models for stage specific (pseudo) outcomes can be set as $E[\tilde{Y}_k|H_k = h_k, Z_k = z_k] = f_k(h_k) + \gamma_k(h_k, z_k)$, where $f_k(h_k)$ is termed the treatment-free part of the model. This model is the same as the model for the expected pseudo-outcomes in Q-learning.

So long as the blip function is correctly specified, inference for the blip parameters will be consistent if the treatment assignment models or if the treatment-free part of the outcome models are correctly specified. To arrive at causal inference, the SUTVA, no unmeasured confounders, positivity, and strict hierarchy assumptions are required. Wallace et al. [2017] explored model assessment and selection for dWOLS, a subject that has seen little attention in the DTR literature; dWOLS has seen extensions to other settings, including for continuous treatments [Schulz and Moodie, 2021] and for survival outcomes [Simoneau et al., 2020].

In this section, we have reviewed frequentist methods to estimate optimal DTRs. We saw value-search methods often require us to restrict ourselves to a family of regimes. In contrast, regression-based methods do not necessarily require restricting to a family of regimes but they require correct modeling of the outcome process. Models leading to interpretable optimal DTRs may face a high risk of misspecification; flexible models can lead to optimal decision rules that are challenging to interpret. The next section presents in more detail the assumptions required to draw causal inference.

2.3 Assumptions for Causal Inference

We now expand on the assumptions needed to draw causal inference. The *stable unit treatment value assumption* is an assumption that requires (causal) *consistency* in the sense that observed covariates and outcomes under assigned treatment equal their counterfactual counterparts under the observed treatments. Mathematically, we may say that for $k = 1, ..., K$, $X_k^{\bar{z}_{k-1}} = X_k$ if $\bar{Z}_{k-1} = \bar{z}_{k-1}$ and that $Y^{\bar{z}} = Y$ if $\bar{Z} = \bar{z}$. That is, all observed quantities equal their corresponding counterfactual quantities [Rubin, 1980, Hernán and Robins, 2020]. A consequence of this assumption is that the outcome of a given individual is not affected by the treatment assignment of other individuals, known as the no interference assumption. The *no unmeasured confounders (sequential randomization)* assumption can be understood in two ways, one that makes use of counterfactuals and one that does not. These definitions are given by:

- Definition 1: For each stage of the decision process, conditional on confounders and treatment history up to time k, treatment assignment at time k is independent of the counterfactual outcomes $Y^{\bar{z}}$ and all other future intermediary outcomes $\bar{X}_{k+1}^{\bar{z}_k}$. Mathematically, we may write that for $k = 1, ..., K$, $Y^{\bar{z}} \perp Z_k | \bar{X}_k, \bar{Z}_{k-1}$ and $(\bar{X}_{k+1}^{\bar{z}_k}, ..., \bar{X}_K^{\bar{z}_{K-1}}) \perp Z_k | \bar{X}_k, \bar{Z}_{k-1}$ [Robins, 1986].

- Definition 2: As presented in Arjas [2012], consider the unobserved history up to time k, $L_k = \{(Y, Z_t, X_t, U_t), t = 1, ..., k\}$, where U_t are unobserved covariates. These contrast observed covariates X_t. Furthermore, defining the observed history up to time k by $O_k = \{(Y, Z_t, X_t), t = 1, ..., k\}$. Then, the sequence of treatments $\{Z_t\}$ is unconfounded relative to latent variables $\{U_t\}$ if for each k, Z_k and $\{U_t, t = 1, .., k\}$ are conditionally independent given $(O_{k-1} = o_{k-1}, X_k = x_k)$. Mathematically, this may be written as $p(z_k | l_{k-1}, u_k, x_k) = p(z_k | o_{k-1}, x_k)$, $k = 1, ..., K$. Throughout this thesis, we assume that x_k contains all the necessary information in order to ensure unconfoundedness, consequently we omit the variables u_k from our notation.

Lastly, the *positivity (absolute continuity)* assumption requires that for $k = 1, ..., K$ if $f_{\bar{X}_k, \bar{Z}_{k-1}}(\bar{x}_k, \bar{z}_{k-1}) \neq 0$ then $P(z_k | \bar{x}_k, \bar{z}_{k-1}) > 0$. From a practical viewpoint, this says that we should be able to observe all types of patients receive all types of treatments. As stated by Hernán and Robins [2020], when considering a specific DTR g, then positivity needs to hold only for treatment and covariate histories consistent with g.

In the next section, we discuss topics in Bayesian inference, so that we can further discuss Bayesian methods for DTRs in section 2.5.

2.4 Bayesian Approaches to Inference

In this section, we touch on some important elements of Bayesian inference that are drawn upon throughout this thesis. We first discuss the traditional approach to Bayesian inference, and we then examine some of the characteristics of Bayesian nonparametric inference. We focus on the Dirichlet process model and additionally discuss some elements of Bayesian decision theory. The methods discussed in this section are not restricted to a sequential decision-making problem, so we consider only a random variable Y, which could be univariate or vector-valued.

Much of the statistical literature focuses on performing inference having observed a sequence of independent and identically distributed (IID) random variables. The De Finetti representation theorem [De Finetti, 1931, Hewitt and Savage, 1955] motivates modeling infinitely exchangeable sequences of observations, which include IID observations, as the product of a likelihood that is conditionally independent on a parameter θ and a prior probability for θ. In parametric inference, the parameter θ belongs to a finite dimensional space and it is assumed that this parameter encodes everything that is unknown about this distribution. Then, by combining the likelihood for the data with a prior probability for θ, it becomes possible to perform statistical modeling and inference. In the next sections, we examine Bayesian nonparametric inference, where the parameter is infinite dimensional, like the data-generating distribution itself. We also examine contemporary literature that allows us to move away from the prior times likelihood paradigm, when inferring about a parameter of interest θ.

2.4.1 Bayesian Nonparametric Inference and the Dirichlet Process

Focusing on distributions that are conditionally independent based on a finite dimensional parameter θ is a restrictive modeling choice; if the distribution is misspecified, then inference for the estimand of interest, represented by the parameter θ, has no guarantee of being consistent. Bayesian nonparametrics reduces the restrictions in Bayesian modeling by taking the entire distribution as the parameter necessary to fully specify a likelihood. Nonparametric approaches offer greater flexibility and robustness to model misspecification [Müller et al., 2015]. Effectively, the distribution (parameter) belongs to an infinite dimensional space, hence the term nonparametric. Like before, we may combine a likelihood with a prior for the space of distribution functions $\Pi(f)$ in order to perform inference. More specifically, we have have the following:

1. A prior model $\Pi(f)$ econding beliefs about the class of distributions giving rise to our data.

2. A likelihood such that conditional on f, $Y_1, Y_2, ..., Y_n \sim f$, where n is the sample size.

Generally, if we choose to make use of nonparametric priors, we may think of Bayes rule as operating in the following manner [Walker, 2010] in order to yield the posterior:

$$\Pi(df|y_1, ..., y_n) = \frac{\Pi_{i=1}^{n} f_Y(y_i)\Pi(df)}{\int \Pi_{i=1}^{n} f_Y(y_i)\Pi(df)}$$

One possible choice of nonparametric prior is the Dirichlet process ($\mathcal{DP}$) prior, not to be confused with the Dirichlet distribution. Before examining the properties of this prior, let us define the Dirichlet process.

Consider a base measure (probability distribution) G_x and a constant $\alpha > 0$. A distribution f with sample space Ω is defined as being distributed according to a $\mathcal{DP}(\alpha, G_x)$ if for every measurable partition of Ω, $(A_1, ..., A_k)$, $(f(A_1), ..., f(A_K)) \sim Dir(\alpha G_x(A_1), ..., \alpha G_x(A_K))$, where Dir corresponds to the Dirichlet distribution. Ferguson [1973] shows that processes with these properties can actually exist. We may consider the $\mathcal{DP}$ a distribution on discrete distributions. However, as we will see next, these discrete distributions cannot be described using a finite number of parameters, hence the term nonparametric.

The first important question that arises from the $\mathcal{DP}$ definition is how to sample distributions from the $\mathcal{DP}$. This may be done via what is known as the stick-breaking construction [Sethuraman, 1994] which begins by considering $V_1, V_2, V_3, ... \overset{iid}{\sim} Beta(1, \alpha)$ and defining $\pi_1, \pi_2, ...$ as following a stick-breaking process with parameter α if $\pi_1 = v_1$, $\pi_2 = (1 - v_1)v_2$, $\pi_3 = (1 - v_1)(1 - v_2)v_3, ...etc$, with $\sum_{i=1}^{\infty} \pi_i = 1$. Then, $f_Y(y)$ is said to be a random discrete distribution following a $\mathcal{DP}(\alpha, G_x)$ if it is of the form

$$f_Y(y) = \sum_{i=1}^{\infty} \pi_i \mathbb{1}_{x_i}(y),$$

with x_i being observed quantities drawn form G_x. We note that f_Y is in the scale of π_is, not of x_is. A common result is that for large values of α, the process variance is smaller and so $f_Y(y)$ will concentrate around G_x [Teh, 2017]. The canonical form of a distribution sampled from the $\mathcal{DP}$ involves an infinite sum. A sample from a $\mathcal{DP}$ can be approximated by truncating the sum at a large index (i.e. when π_i are small).

Having these definitions in place, we can now consider the consequence of placing a $\mathcal{DP}(\alpha, G_x)$ prior on data generating distributions. This prior is a conjugate prior, so that the posterior is also a $\mathcal{DP}$ [Ghosal, 2010] of the form

$$\mathcal{DP}(\alpha + n, \frac{\alpha}{\alpha + n} G_x + \frac{1}{\alpha + n} \sum_{i=1}^{n} \mathbb{1}_{y_i}(y)).$$

We may be interested in the posterior predictive distribution

$$f(y^*_{n+1}|y_1, ..., y_n) = \int f(y^*_{n+1}|f)\Pi(df|y_1, ..., y_n).$$

A well known result from Blackwell and MacQueen [1973] is that the posterior predictive distribution is given by:

$$f(y^*_{n+1}|y_1, ..., y_n) = \frac{\alpha}{\alpha + n} G_x(y^*_{n+1}) + \frac{1}{\alpha + n} \sum_{i=1}^{n} \mathbb{1}_{y_i}(y^*_{n+1}).$$

Under a specific choice of α, the $\mathcal{DP}$ prior allows us to arrive at a procedure known as the Bayesian bootstrap, first discussed by Rubin [1981] . The Bayesian bootstrap uses the same nonparametric assumptions discussed above, meaning it places a $\mathcal{DP}$ prior on the data generating distribution. In the context of a Bayesian bootstrap, we allow $|\alpha| \to 0$. In practice, we can think of α as being effectively zero by taking it to be such a small number that it has negligible effects on inference. This modeling choice leads us to the following posterior:

$$\Pi(df|y_1, ...y_n) = \mathcal{DP}(n, \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{y_i}(y)). \tag{2.7}$$

In this case, given that the base measure is discrete and has finite support, we may write a distribution sampled from the posterior as

$$f(y) = \sum_{i=1}^{n} \pi_i \mathbb{1}_{y_i}(y), \text{ where } (\pi_1, ..., \pi_n) \sim Dir(1, ..., 1) \tag{2.8}$$

[Gasparini, 1995]. This is a key fact as now a distribution sampled from the posterior $\mathcal{DP}$ is uniquely represented by a finite vector of weights. The reason as to why expression (2.8) is termed the Bayesian bootstrap is because it can also be seen as the frequentist bootstrap when each weight π_i takes values in $\{0, 1/n, ..., n/n\}$ representing the proportion of times that y_i appears in a given bootstrapped sample [Mitra and Müller, 2015]. This Bayesian bootstrapping procedure allows us to perform a Bayesian analysis, all while utilizing tools that are familiar to the broader statistical community. Although Bayesian nonparametrics allows for flexible modeling, sometimes the estimand of interested is a finite-dimensional parameter. In the next section we examine how parameters can be incorporated into an inferential scheme, even when they are not embedded in a parametric likelihood.

2.4.2 Bayesian Decision Theory

Much of Bayesian inference focuses on inferring about parameters embedded in a likelihood. However, there may be scenarios where interest is not in these parameters but in parameters embedded in utility or loss functions. Walker [2010] lays out a framework to justify decision-making based on optimizing expected posterior utilities/losses.

To define the benefit of an action, Walker [2010] proceeds by example in considering a model selection problem where we seek the optimal choice of θ for a parametric family of densities $\mathbb{F} = \{h(y;\theta); \theta \in \Re\}$ when modeling data generated by a distribution $f(y)$, which may or may not be contained in the set $\mathbb{F}$. A natural manner in which to do this is to define a utility that expresses the benefit of selecting a function $h(\cdot)$ and to search for the action that maximizes this utility; alternatively a loss function can be defined and minimized.

For example, if the Kullback-Leibler (KL) divergence ($d(f,h) = \int f \log(\frac{f}{h})$) is used as a measure of divergence between two distributions, and if we take the expectation of this divergence with respect to the posterior predictive distribution, then we are interested in finding θ^* that maximizes $E[\log(h(Y^*;\theta))|y_1,...,y_n]$. Regardless of whether $\mathbb{F}$ contains the true posterior distribution or not, the optimal choice of θ is the minimizer of the posterior expected loss, as it allows us to update our believe of the loss-minimizing parameter in light of new data. If a nonparametric model leading to the Bayesian bootstrap is used, then minimizing the KL divergence is equivalent to maximizing

$$E[\log(h(Y^*;\theta))|y_1,...,y_n] = E_\pi[\sum_{i=1}^{n} \pi_k \log(h(y_i;\theta))] = \frac{1}{n}\sum_{i=1}^{n} \log(h(y_i;\theta)),$$

leading to a θ^* that is the maximum likelihood estimator. Walker [2010] goes on to explain how this setup allows for linking random distributions from the posterior distribution with draws from θ^* in order to construct a posterior distribution for the utility/loss optimizing parameter, θ^*. Furthermore, this procedure can be regarded as a case of semiparametric inference in the sense that inference about a finite dimensional parameter is being made, all while working with an infinite dimensional space of distributions. Semiparametric inference has an expansive literature in the frequentist setting, see for example Tsiatis [2007], and the utility/loss optimization framework discussed above has promise in allowing for the benefits of frequentist semiparametric theory and methodology to be brought over into a Bayesian framework.

This utility/loss optimization framework was taken by Saarela et al. [2015b] who consider inference using utilities to infer about static treatment regimes of time-varying treatments using marginal structural models. Stephens et al. [2022] and Luo et al. [2022] have also considered the use of Bayesian inference via utilities in a causal setting. Non-causal work includes Bissiri et al. [2016], who propose inferential procedures in a loss/utility function rather than a likelihood, and Lyddon et al. [2019], who examine inference via loss functions by proposing the loss-likelihood bootstrap. A decision-making framework based on utilities is appealing as a means to perform Bayesian inference. It has the possibility of obviating some of the challenges with Bayesian inference that require specifying complex likelihoods and that may face a risk of model misspecification. In the next section, we examine some of the work that has been done on Bayesian inference for DTRs. As we will see, in contrast to the frequentist approach, there is a paucity of Bayesian approaches for DTRs and some of the existing methods must be carefully adapted to each inferential problem, thus discouraging their use in applied literature.

2.5 Bayesian Methods for DTRs

Bayesian inference has many appealing aspects: the flexibility of incorporating prior information into the inferential problem, the possibility of drawing inference without the necessity for asymptotic considerations, and the coherence of making probabilistic statements about the quantities of interest are but some of the advantages. However, Bayesian inference faces challenges too as the standard approach requires identifying likelihoods that face a risk of misspecification. Furthermore, the resulting inferential procedures are often complex and may have a heavy computational burden. In the following, we discuss some current approaches to Bayesian inference for DTRs and note that the current literature still requires further developments so that these methods can be more widely adopted in general practice.

Arjas and Saarela [2010] directly address the problem of Bayesian inference for DTRs by considering the analysis of data from the well-known Multicenter AIDS Cohort Study [Kaslow et al., 1987] to assess the effect of initiating therapy with AZT, an antiretroviral medication. They infer an optimal DTR in a two-stage setting with the aim of maximizing 12 month CD4 cell count, and they do so with a full likelihood based Bayesian analysis. The authors' approach requires them to model the intermediate covariate, x_2, distribution in addition to conditional outcome distributions using nonparametric regression models defined in Saarela and Arjas [2011]. In their application, a single stage two covariate needs to be included in the analysis, thereby making the modeling approach feasible. In particular, x_2 is the univariate variable CD4 cell count at six months. Backward induction is used to identify optimal regimes by first estimating posterior predictive expectations using Markov Chain Monte Carlo (MCMC) and Monte Carlo integration. The authors argue that the entirely probabilistic framework of their approach is an asset, however it is unclear how the proposed approach could be applied in settings with other data characteristics, for example with more tailoring covariates or in observational studies that require adjusting for a variety of confounders, possibly time-varying. These added characteristic would require modeling a multivariate x_2 distribution. Furthermore, the resulting optimal DTRs identified with this method have no analytic expression or clear interpretable form. Saarela et al. [2015a] follow a similar likelihood approach but make use of fully parametric models. Zajonc [2012] also explore full likelihood based methods, with the aim of sampling from a posterior predictive distribution of counterfactual outcomes and intermediary covariates, in order to evaluate the value under a small set of regimes and consequently identify the optimal DTR. Lee et al. [2015] explore Bayesian methods requiring the specification of a likelihood in a clinical trial design that adaptively optimized patient doses.

Other work using Bayesian methods in the DTR realm includes that of Xu et al. [2016], who are motivated to evaluate chemotherapeutic regimes for acute myelogenous leukemia. They compare several candidate dynamic regimes with the aim of maximizing mean overall survival time in a setting where patients can transition to several disease states, namely death, resistant disease, complete remission, or progressive remission. Bayesian nonparametric survival regression models for transition times between states are used. In particular they propose a dependent Dirichlet process mixture (of normals) model [MacEachern, 1999] with a Gaussian process prior on the mean function of the normal distributions. Having described a sampling strategy for these models, the authors make use of G-computation [Robins, 1986] to arrive at an estimate for the mean survival time for each regime. This computation is as in Wahed and Thall [2013], who examine the same question in cancer therapy but who use a frequentist likelihood-based regression approach. The authors compare the proposed Bayesian nonparametric approach with IPW and augmented IPW in simulation and identify that their proposed approach yields improved performance by decreasing uncertainty around the estimated regime value. Although this approach works well, the estimation procedure leads to a general lack of interpretability, which may be of importance when addressing clinical questions. Questions remain regarding the adaptability of this approach to settings where the optimal regimes is chosen among a large, possibly infinite, set of regimes as opposed to a small discrete set of candidate regimes.

A Bayesian approach to identifying optimal ITRs that may have broader applicability is proposed by Logan et al. [2019], who use Bayesian additive regression trees (BARTs) [Chipman et al., 2010] to model $E[Y|X = x, Z = z]$. With this model in place, solving $\arg\max_z E[Y|X = x, Z = z]$ permits for the maximization of the marginal outcome $E[E[Y|X, Z = g(X)]]$. Their approach models the outcome Y using a sum of Bayesian regression trees, each represented by a function $h(x, z; T, M)$, where T denotes a tree structure consisting of interior and terminal nodes, with branches representing decision rules based on covariates, and where M denotees a list of function values at the terminal nodes. Mathematically, this model posits that

$$Y = f(x,z) + \epsilon, \text{ with } \epsilon \sim N(0,\sigma^2) \text{ and}$$
$$f(x,z) = \sum_{j=1}^{m} h(x,z;T_j,M_j).$$

The authors then move to place a BART prior on f, using notation $f \sim BART$. More specifically, this is a prior for the constituents of the tree; nodes in the tree are assigned probabilites of having a child node; covariates that partition a node, the partitioning value, and the values at the terminal nodes are also assigned priors. With this prior structure, authors use MCMC to obtain trees $\{T_j^d, M_j^d\}$ for $j = 1,...,m$ and MCMC iterations $d = 1,...,D$. The BART procedure is viewed as yielding draws $\{f_d; d = 1,...,D\}$ from the posterior distribution of f [Logan et al., 2019]. Across draws from the posterior distribution, $E[Y|X = x, Z = z]$ can be approximated by using $\hat{E}[Y|X = x, Z = z] = 1/D\sum_{d=1}^{D} E[Y|X = x, Z = z, f_d]$. The BART ITR is then defined as $g_{BART}(x) = \arg\max_z \hat{E}[Y|X = x, Z = z]$. This means that for a given covariate x, the optimal treatment can be identified. Logan et al. [2019] also discuss how to compute the value at the optimal ITR and its associated uncertainty by drawing from the posterior distribution. They mention that an advantage of this methodology is its ability to handle complex functional forms in the outcome process as well as covariate-treatment interactions, which is of pertinence in the study of ITRs. One main limitation of this approach is the lack of interpretable decision rules that result.

A Bayesian Q-learning approach proposed by Murray et al. [2018], termed by authors a Bayesian machine learning method, takes a similar approach to Logan et al. [2019] but focuses on a sequential setting. This approach is best understood by considering a two-stage problem. In the authors' setup, stage-specific rewards Y_1, Y_2 are available, and an optimal decision rule is one that maximizes the expected cumulative outcome of $Y = Y_1 + Y_2$. Backward induction is used to arrive at the stage-specific optimal strategies, which requires positing stage-specific outcome models. At stage 2, a stage 2 model is posited:

$$\begin{aligned}
&Y_2|H_2 = h_2, Z_2 = z_2, \theta_2 \sim f_2(y_2|h_2, z_2, \theta_2)\\
&\theta_2 \sim p_{20}(\theta_2),
\end{aligned}$$

with θ_2 being an unknown parameter with prior p_{20}. From this model, a posterior predictive distribution $f_{2n}(y_2^*|h_2, z_2, \theta_2, \mathcal{D}_n)$ can be obtained and used to evaluate the posterior optimal decision rule $\hat{g}_{2opt}(h_2) = \arg\sup_{z_2} E[Y_2^*|H_2 = h_2, Z_2 = z_2, \mathcal{D}_n]$; note that for each θ the optimal decision rule is given by $g_{2opt}(h_2; \theta_2) = \arg\sup_{z_2} E[Y_2|H_2 = h_2, Z_2 = z_2, \theta_2]$. The last-stage decision rule is straightforward to identify; the challenge comes at the first stage where a likelihood and prior should be placed on the cumulative counterfactual outcome having received optimal stage 2 treatment:

$$\begin{aligned}
&Y^{z_1, g_{2opt}}|X_1 = x_1, Z_2 = z_2, \theta_1 \sim f_1(y|x_1, z_2, \theta_1)\\
&\theta_1 \sim p_{10}(\theta_1).
\end{aligned}$$

Unfortunately, these counterfactual outcomes are not observed for all patients; if a patient follows the optimal stage two treatment, then the counterfactual of interest is observed, $y^{z_1, g_{2opt}} = y_1 + y_2^{z_1, g_{opt}} = y$. However, for patients where z_2 is not the optimal therapy then $y^{z_1, g_{2opt}} = y_1 + y_2^{z_1, g_{opt}}$ is not observed. For these patients, $y_2^{z_1, g_{opt}}$ can be considered missing and denoted by y_2^{mis}. The aim here is to obtain the stage 1 posterior predictive distribution, $f_{1n}(y^*|z_1, g_{2opt}, \mathcal{D}_n)$, as this will allow for computing the stage 1 posterior expected outcome. For this, the posterior distribution of θ_1, $p_{1n}(\theta_1^*|d_{2opt}, \mathcal{D}_n)$, must be computed; to address the fact that some counterfactual outcomes are not observed, Murray et al. [2018] arrive at the following computation:

$$\begin{aligned}
&p_{1n}(\theta_1^*|d_{2opt},\mathcal{D}_n) \propto p_{10}(\theta_1) \\
&\times \int \Bigg(\prod_{i:z_{2i}=g_{2opt}(h_2;\theta 2)} f_1(y_{1i}+y_{2i}|x_{1i},z_{1i},\theta_1) \\
&\times \int \Bigg[\prod_{i:z_{2i}\neq g_{2opt}(h_2;\theta 2)} f_1(y_{1i}+y_{2i}^{mis}|x_{1i},z_{1i},\theta_1) \Bigg] \\
&f_{2n}(y_2^{mis}|\mathcal{D}_n)dy_2^{mis} \Bigg) p_{2n}(\theta_2^*|\mathcal{D}_n)d\theta_2^*.
\end{aligned}$$

A standard Bayesian computation for the posterior distribution of θ_1 would take the product of the prior p_{10} on the first line, with the stage 1 likelihood for the counterfactual outcomes of interest. Unfortunately as some of these values are missing, the likelihood must be split into a component with observed values for those who follow the optimal stage two rule and with missing values for those who do not follow it. Marginalization can then be performed over the missing values using the f_{2n} distribution. Lastly, as there is also uncertainty around the optimal stage 2 rule, the splitting of the likelihood into unobserved and observed contributions must also marginalized over all values of the optimal stage two treatment; this is the role of p_{2n} in the equation. Murray et al. [2018] provide an algorithm for sampling from this posterior, which can then be used to compute a posterior predictive distribution to ultimately compute:

$$\hat{g}_{1opt}(x_1) = \arg\sup_{z_1} E[Y^{z_1,g_{2opt}*}|X_1=x_1,Z_1=z_1,\mathcal{D}_n].$$

Note that the priors in this Bayesian Q-learning approach were not specified; authors advocated for using Bayesian nonparametric regression models like BART.

Bayesian methods for DTRs have progressed slowly; generally the inferential problem is challenging and not amenable to "likelihood times prior" approaches, as it becomes necessary to model intermediary covariates, or to posit likelihoods that face a high risk of model misspecification, unless nonparametric approaches are utilized. The Bayesian Q-learning approach developed by Murray et al. [2018] shows promise though resulting optimal treatment rules are not necessarily interpretable. There is a gap in the literature in developing Bayesian methods that can yield interpretable optimal DTRs all while being robust to model misspecification. Additionally, we note that beyond methods that estimate the value for a small set of regimes and then identify an optimum among these, the value-search approach to inference has not been exploited in the Bayesian setting. In the next section, we examine another approach to function optimization, which has not been explored in the DTR literature.

2.6 Function Optimization using Computer Experiments

As discussed, value-search approaches involve maximizing the value function directly; these can be implemented by specifying parsimonious models for the value function with dynamic MSMs and consequently maximizing the value function, using a genetic algorithm to maximize an estimator of the value function, or changing the value function optimization problem into a classification problem that minimize a weighted misclassification error. These are all promising approaches, however there are other methods for function maximization that have not yet been examined in the DTR literature. In particular, the area of computer experiments, has placed much focus in solving problems of the form $\arg\max_{\psi} f(\psi)$. These methods are usually motivated by settings where $f(\psi)$ is challenging to evaluate, and so a working model for the objective function should be fit using a set of design points and sequentially sampling more experimental points based on a selection criterion. Generally, this criterion, known as an infill criterion, is such that it selects points most informative about where an optimum may be.

Usually, an approximation for the objective function is formed using a Gaussian process ($\mathcal{GP}$) assumption, which says that any set of values of f evaluated at any arbitrary $\{\psi_1, ..., \psi_m\}$ have an m-variate Gaussian distribution Quadrianto et al. [2017]. These processes are uniquely determined by their mean function $m(\psi) = E[f(\psi)]$ and their covariance function $k(\psi_i, \psi_j) = E[(f(\psi_i) - m(\psi_i))(f(\psi_j) - m(\psi_j))]$.

Models obtained through the $\mathcal{GP}$ assumption have been used extensively in regression and classification tasks [Williams and Rasmussen, 2006]. When the aim is to emulate a function using a set of experimental points, models arising from the $\mathcal{GP}$ assumption are often termed kriging models [Krige, 1951]. In particular, these models seek to model a function $f(\psi)$ for ψ in some domain of interest in order to make predictions about f, when only knowing values at a set of experimental points $\{\psi_1, ..., \psi_m\}$. The $\mathcal{GP}$ assumption allows for the quantification of uncertainty about f even if it is not the result of a random phenomenon. Rather, uncertainty in this problem arises, at least in part, as a result of only partially observing a deterministic phenomenon; this kind of uncertainty is termed epistemic [Roustant et al., 2012]. Kriging methods can be combined with sequential sampling strategies that evaluate the objective function at new experimental points; usually, the reason for sequential sampling is to identify new points that inform the model about the location of a maximizer.

Kriging methods, including those used to optimize unknown functions, have a long history. For example, Kushner [1964] maximize a multipeaked curve in the presence of noise, using a specific Gaussian process assumption and the maximum probability of improvement as an infill criterion. Jones et al. [1998] introduce the efficient global optimization (EGO) algorithm for identifying function optima in deterministic computer experiments. Their approach makes use of an infill criterion known as the expected improvement criterion which is now very popular in applications. O'Hagan [2006] make a case for Bayesian kriging, and discussed why Bayesian methods are more appropriate for many kriging tasks. Lizotte [2008] acknowledge that a Bayesian approach is appropriate, but also that a fully Bayesian analysis is complex and challenging to adopt; consequently the author used empirical Bayes or maximum a posteriori inference to estimate $\mathcal{GP}$ parameters and to perform optimization.

Kriging methods in settings with point-wise noisy observations are of substantial interest in some fields [Roustant et al., 2012], where values of f cannot be directly observed, even at the experimental points; only a noisy version of f is observable. In computer experiments, for example, Forrester et al. [2006] and Huang et al. [2006] proposed optimization methods using kriging in a homoskedastic noise settings. Kriging methods accommodating heteroskedastic noise have received less attention, but fully Bayesian approaches have been explored including by Goldberg et al. [1997] and Wang [2014]. With an attempt at reducing the computational burden, other Bayesian inspired approaches have also been proposed [Kersting et al., 2007, Zhang and Ni, 2020]. In optimization settings that involve noisy observations, the choice of infill criterion becomes important; Picheny et al. [2013] review infill criteria for noisy observation problems in settings where sampling at the same function input is informative and in settings where it is not.

The problem of identifying a value-maximizing DTR can be seen as a problem of function optimization. The computer experiments literature is vast, with methods that have seen applications in many fields. Surprisingly, no work has been done to explore the use of these optimization methods in the realm of identifying optimal DTRs.

2.7 Summary

In this literature review, we have discussed the goals of precision medicine and some of the methodological challenges that need to be overcome and the assumptions that need to be made in order to effectuate these objectives. We have discussed value-search and regression based methods, and generally we have observed that methods that yield interpretable decision rules are important; one challenge that we saw when examining Bayesian DTR methods is the interpretability of the decision rules and the fact that there is a scarcity specifically in Bayesian value-search approaches. Our discussion of Bayesian nonparametric methods allowed us to observe that these methods yield flexible modeling strategies and in the context of Gaussian process optimization, which is an optimization approach not explored in precision medicine, we see that Bayesian methods are required, as frequentist approaches are not always amenable to characterizing the uncertainties of interest; this consideration alone is reason to advocate for the development of Bayesian inferential methods in general, although the benefit of incorporating priors, and being able to make probabilistic statements about quantities of interest is also important.

Chapter 3

Semiparametric Bayesian Inference for Optimal Dynamic Treatment Regimes

Preamble to Manuscript 1. This chapter presents the first manuscript in this thesis; it builds on work by Saarela et al. [2015b] who show how to infer about marginal structural models for static regimes using a utility maximization framework. It is also motivated by work in Saarela et al. [2016] who use Bayesian predictive inference and a doubly robust estimator in a cross-sectional setting.

This manuscript analogizes frequentist developments for inference about MSMs in a Bayesian framework: first an estimation procedure was developed for static MSMs [Robins et al., 2000], and later this was extended to dynamic MSMs [Orellana et al., 2010a]. This manuscript develops the latter within a Bayesian paradigm such that estimation and inference can be achieved with the benefits of the probabilistic statements and interpretation that the Bayesian framework permits.

Bayesian literature for optimal dynamic treatment regimes (DTRs) is scarce, in part because conventional Bayesian approaches requiring a likelihood pose a challenge in this setting, where model misspecification is of considerable concern. Observational studies compound the difficulty of performing causal inference, as confounding must be addressed. The work in this chapter demonstrates how to bypass these difficulties through a decision theoretic Bayesian approach.

Specifically, the contributions in this chapter include i) a demonstration of how inference for dynamic MSMs may be cast as a Bayesian problem of utility maximization thereby allowing for the estimation of the expected outcome under adherence to a DTR and of optimal DTRs, ii) the defining of a probability measure that facilitates performing causal inference about the DTRs of interest and that allows for a Bayesian interpretation of the data-augmentation procedure needed to infer about optimal DTRs, iii) a demonstration of how to perform individualized inference using a novel computation for the probability that a specific patient should receive a given treatment, based on what has been learned from data about the optimal treatment strategy, iv) an examination of how Bayesians may regard singly and doubly robust inference for DTRs via a nonparametric prior, and v) the exploration how these methods may be applied to answer a substantive question in HIV care.

This manuscript has been published in *Biostatistics* [Rodriguez Duque et al., 2022b]. Note that the online supplementary material that is mentioned in section 9 of this chapter is also given in this thesis' Appendix A.

Semiparametric Bayesian Inference for Optimal Dynamic Treatment Regimes

Daniel Rodriguez Duque[1], David A. Stephens[2], Erica E.M. Moodie[1], Marina B. Klein[3].

[1] *Department of Epidemiology, Biostatistics, and Occupational Health, McGill University*

[2] *Department of Mathematics and Statistics, McGill University*

[3] *Department of Medicine, McGill University*

Abstract

Considerable statistical work done on dynamic treatment regimes (DTRs) is in the frequentist paradigm, but Bayesian methods may have much to offer in this setting as they allow for the appropriate representation and propagation of uncertainty, including at the individual level. In this work, we extend the use of recently developed Bayesian methods for Marginal Structural Models (MSMs) to arrive at inference of DTRs. We do this 1) by linking the observational world with a world in which all patients are randomized to a DTR, thereby allowing for causal inference and then 2) by maximizing a posterior predictive utility, where the posterior distribution has been obtained from non-parametric prior assumptions on the observational world data-generating process. Our approach relies on Bayesian semiparametric inference, where inference about a finite-dimensional parameter is made all while working within an infinite-dimensional space of distributions. We further study Bayesian inference of DTRs in the doubly robust setting by using posterior predictive inference and the non-parametric Bayesian bootstrap. The proposed methods allow for uncertainty quantification at the individual level, thereby enabling personalized decision making. We examine the performance of these methods via simulation and demonstrate their utility by exploring whether to adapt HIV therapy to a measure of patients' liver health, in order to minimize liver scarring.

3.1 Introduction

Precision medicine is a research area that seeks to tailor patient care to improve health outcomes, all while reducing over-treatment. For conditions that require sustained therapy through time, assigned treatments may vary through stages of the treatment process. To identify treatment strategies that follow the principles of precision medicine, stage-specific treatments must be allowed to change with patients' evolving characteristics. These treatment strategies are termed dynamic treatment regimes (DTRs). DTRs contrast static treatment regimes, where time-varying treatments are assigned at study-start. One tool employed to infer about time-varying treatments are marginal structural models (MSMs). These models were developed to evaluate the effect of static regimes [Robins et al., 2000] and later extended to evaluate adherence to DTRs [Murphy et al., 2001], and to identify optimal DTRs [Orellana et al., 2010a, van der Laan and Petersen, 2007]. MSMs rely on an appealing estimation strategy; they allow scientists to target a finite set of causal estimands without requiring restrictive assumptions about the family of data generating distributions. Semiparametric methods like these have mostly been studied from a frequentist viewpoint.

Semiparametric methods are enviable as they avoid specifying fully parametric probabilistic models that face a high risk of misspecification. These methods may be contrasted with the conventional Bayesian approach to inference, which seeks to multiply a parametric likelihood with a prior. In simple settings, this approach works well, but in more complex settings, like in sequential decision-making, the correct specification of a likelihood is highly suspect. Some work has been done examining the effects of model misspecification in Bayesian inference. For example, Walker [2013] shows that under some conditions, parameters in the misspecified model converge to the minimizers of the Kullback-Leibler divergence. Although this is reassuring, it does mean that inference cannot be guaranteed to be consistent and consequently, treatment recommendations based on misspecified models could be suboptimal. Furthermore, in a setting with time-varying confounding and mediation, the correct specification of a likelihood with parameters representing causal treatment effects will not yield fruitful results; this is because only confounded data are available and these data follow a different probability law. Now, one approach that may guarantee consistency is Bayesian inference via completely non-parametric specifications. In the DTR setting Bayesian non-parametrics have been used to estimate the effect of a small number of dynamic regimes [Xu et al., 2016], but when the family of regimes grows, this approach may not be feasible to identify optimal regimes, due to computational limitations. Generally, it is unresolved how Bayesians may best capitalize on semiparametric approaches to inference about DTRs, and this is one of the challenges that our work addresses.

A variety of other methods for estimating the effect of DTRs have been proposed. For example g-methods including g-computation [Robins, 1986], and g-estimation of structural nested models [Robins, 1993]. Other ways by which to identify optimal DTRs include Q-learning [Zhao et al., 2009], and outcome weighted learning [Zhao et al., 2012]. In a Bayesian setting, a standard parametric approach to inference requires specifying the full dynamics of the data generating process in order to learn about dynamic regimes. For example Saarela et al. [2015a] use a predictive Bayesian approach that requires the specification of parametric distributions for outcomes and intermediate covariates in order to identify optimal DTRs. Murray et al. [2018] propose a Bayesian adaptation to Q-learning that utilizes machine learning methods for flexible modeling, however the approach still relies on likelihoods for stage-specific rewards/outcomes. Exceptionally, a few researchers have explored the use of Bayesian non-parametric methods in the DTR setting; Arjas and Saarela [2010] take this approach, however their method is not computationally feasible as the number of confounders increases.

Ideally, Bayesians would target a finite dimensional estimand that indexes a large family of regimes, all while working within an infinite dimensional class of data generating distributions. Recent work has elucidated ways in which semiparametric inference may be viewed through a Bayesian lens. First, let us review the frequentist setup. Frequentist semiparametrics begins with an estimating function, which under certain modeling assumptions (e.g. for the mean) is an unbiased estimator of zero. For finite samples, setting the estimating function equal to zero and solving for a parameter of interest β yields an estimator $\hat{\beta}_n^*$ which, under regularity conditions, is consistent and asymptotically normal. A framework for Bayesian semiparametric inference should allow us to take a similar approach. It was not until recently that MSMs for static regimes were provided with a Bayesian motivation by considering the maximization of an expected posterior predictive utility [Saarela et al., 2015b], which required solving for β in a manner analogous to the frequentist procedure. Later, using a similar flavor, Bayesian doubly robust inference was motivated [Saarela et al., 2016]. Other similar recent approaches have further considered inference via utility functions [Bissiri et al., 2016] and through the loss-likelihood bootstrap [Lyddon et al., 2019]. What is particularly liberating about these inferential procedures is that Bayesian methods can be used to infer about parameters that are not necessarily embedded in a likelihood, which would undoubtedly be misspecified. However, none of these approaches have examined causal inference for optimal DTRs.

Our work builds on the general framework developed by Saarela et al. [2015b] for performing Bayesian causal inference with MSMs. Those authors focused on inferring about stage-specific causal treatment effects of static regimes. As it is well established that MSMs can also be used to infer about (optimal) DTRs, our work seeks to examine how to use this general framework to perform Bayesian causal inference of DTRs. This requires us to carefully interpret the estimands of interest, so that we may conceive of a counterfactual world that allows for causal inference. In the doubly robust setting, we explore posterior predictive inference for DTRs. This approach to inference was proposed by Saarela et al. [2016], but it has only been studied in the cross-sectional setting. We transparently lay out the use of this new framework for Bayesian causal inference, and with this in mind, we explore the performance of this approach via simulations with treatment rules like "assign treatment when a covariate value x exceeds a threshold θ", with the aim of identifying θ_{opt} that optimizes a final outcome. Additionally, with the purpose of illustrating how this methodology may be used in practice, we consider an analysis of HIV therapy using data from the North American AIDS Cohort Collaboration on Research and Design (NA-ACCORD) where we aim to learn about whether to tailor on FIB4, a measure of liver scarring, in order to decide when to switch antiretroviral therapies, with the aim of minimizing long term liver damage.

In addition to the above-mentioned contributions, we note that frequentist uncertainty quantification does not allow for decision-makers to ask if a new patient will benefit from therapy suggested by an optimal DTR. As we will elaborate, Bayesian posterior predictive inference allows for decision-makers to assess the probability that therapy is optimal for a specific patient, thereby allowing for individualized care. To our knowledge, no other approach quantifies uncertainty at the patient-level decision-making process.

The approach to inference presented here uses the posterior predictive distribution in order to answer causal questions about DTRs; there is no need to model counterfactual outcomes directly. The advantages and detriments of counterfactuals has been studied by, for example, Dawid [2000]. Arjas [2012] presents an approach similar to the one taken here, where the quantities of interest are expected conditional outcomes.

3.2 Estimation Strategy

In this section, we first describe the inferential setting and motivate Bayesian inference via a utility maximization framework. We follow this by a precise definition and formulation for connecting two probability laws: the observational world law and the law that allows us to draw causal inference about optimal DTRs by eliminating confounding. We then provide a prior that facilitates robust inference in the developed framework. Lastly, we examine specific utilities that allow for causal inference about optimal DTRs. Some of the developments parallel Saarela et al. [2015b], but require some specific considerations for our context; we also take the opportunity to emphasize some of the nuanced arguments present in this framework.

3.2.1 Inferential Setting

We consider a sequential decision problem with K decision points and a final outcome y to be observed at stage $K+1$. Decisions taken up to stage k give rise to a sequence of treatments $\bar{z}_k = (z_1, ..., z_k)$, $z_j \in \{0, 1\}$. At each stage k, a set of covariates x_k is available for decision-making and it is assumed that these consist of all time-fixed and time-varying confounders. To denote covariate history up to time k, we write $\bar{x}_k = \{x_1, ..., x_k\}$. Subscripts are omitted when referencing history through stage K. We denote a DTR-enforced treatment history by $g(\bar{x}) = (g_1(x_1), ..., g_K(\bar{x}_K))$. Our focus is restricted to deterministic DTRs. Throughout, we will consider a family of DTRs, which will be indexed by $r \in \mathcal{I}$ to give $\mathcal{G} = \{g^r(\bar{x}); r \in \mathcal{I}\}$. The index is omitted when it is clear that our focus lies on a single DTR. Treatment and covariate histories may be considered under the probability laws in two worlds: the *observational* world $\mathcal{O}$ which denotes the law giving rise to the data in the study population, and the *experimental* world $\mathcal{E}$, which denotes a world in which causal inference may be performed. In the next sections, the definition of $\mathcal{E}$ will be made more precise. Lastly, variables sampled from a posterior distributions are shown with *.

As in Saarela et al. [2015b], we assume that for each $i = 1, ..., n, n+1, ...,$ $b_i = (y_i, \bar{x}_i, \bar{z}_i)$ are infinitely exchangeable sequences to deduce the de Finetti representation (as in Bernardo and Smith [2009]) in the observational world:

$$\begin{aligned} p_{\mathcal{O}}(b_1, ..., b_n) = \int_{\tau,\phi,\gamma} \prod_{i=1}^{n} p_{\mathcal{O}}(y_i|\bar{x}_i, \bar{z}_i, \tau) \\ \prod_{j=1}^{K} p_{\mathcal{O}}(x_{ij}|\bar{z}_{i(j-1)}, \bar{x}_{i(j-1)}, \phi_j) p_{\mathcal{O}}(z_{ij}|\bar{z}_{i(j-1)}, \bar{x}_{ij}, \gamma_j) p(\tau, \phi, \gamma) d\tau d\phi d\gamma. \end{aligned} \tag{3.1}$$

In Appendix A.1, we provide a more general representation in cases where there may be unmeasured causes u of both intermediary and the final outcome. Outcomes do not inform the treatment assignment mechanism, characterized by a parameter γ (i.e. $p_{\mathcal{O}}(\gamma|\bar{b}) \propto p(\gamma|\bar{x}, \bar{z}))$[Saarela et al., 2015b]. The no unmeasured confounders assumption allows us to model treatment assignment probabilities in equation (3.1) with observed covariates only as $p_{\mathcal{O}}(z_{ij}|\bar{z}_{i(j-1)}, \bar{x}_{ij}, \gamma_j)$. This assumption is not often encountered outside the counterfactual framework, so we provide it in Appendix A.1.

3.2.2 Bayesian MSMs for Dynamic Regimes

Saarela et al. [2015b] have previously considered Bayesian MSMs to estimate the stage-specific effect of static regimes. However, in a precision medicine setting, it is not immediately clear how to employ this method of inference to infer about DTRs. In what follows, we adapt their work to the dynamic MSM setting for DTRs, attempting in the process to clarify the nuances in this general framework. To allow for MSMs to make Bayesian inference of optimal DTRs, we must make several considerations. First, consider a utility function $U(\bar{b}, g, \beta)$; which represents a patient's utility as a function of patient covariates and regime assignment, parameterized by an unknown parameter β. This utility may take any form relevant to the decision-maker (further details about this decision-theoretic approach may be found in Walker [2010]). We will see that some specific utilities allow us to infer about the causal parameters of interest. As Bayesian decision-makers, we are interested in finding the value of β that maximizes the posterior expected utility $E_{\mathcal{E}}[U(b^*, g, \beta)|\bar{b}]$. This is an expectation taken with respect to the experimental measure in which patients are randomized to regimes in $\mathcal{G}$ at study start, with probability $p(g)$. When we consider a finite set of regimes in which patients have equal probability of randomization, we may replace this probability with $1/C_g$, where $C_g = |\mathcal{I}|$. In the experimental setting consider $v_i = (b_i, g_i) \equiv (x_i, z_i, y_i, g_i)$, and assume infinite exchangeability to obtain:

$$p_{\mathcal{E}}(v_1,\ldots,v_n) = \int \prod_{i=1}^{n} p_{\mathcal{E}}(y_i|\bar{x}_i, \bar{z}_i, g_i, \tau)$$
$$\prod_{j=1}^{K} p_{\mathcal{E}}(x_{ij}|\bar{z}_{i(j-1)}, \bar{x}_{i(j-1)}, g_i, \phi_j) p_{\mathcal{E}}(z_{ij}|\bar{z}_{i(j-1)}, \bar{x}_{ij}, g_i, \alpha_j) p(g_i) p(\tau, \phi, \alpha) d\tau d\phi d\alpha. \qquad (3.2)$$

Note $p_{\mathcal{E}}(z_{ij}|z_{i(j-1)}, \bar{x}_{ij}, g_i, \alpha_j) = \mathbb{1}_{g(\bar{x}_{ij})}(z_{ij})$, as treatment is deterministically assigned conditional on regime. For convenience, we re-express the product across all stages as $\prod_{j=1}^{K} \mathbb{1}_{g_j(\bar{x}_{ij})}(z_{ij}) = \mathbb{1}_{g(\bar{x}_i)}(\bar{z}_i)$. This representation differs from that presented in Saarela et al. [2015b], as the experimental world here differs. Now, we seek to link $\mathcal{E}$ and $\mathcal{O}$. In particular, we make this link with respect to the posterior predictive distribution. Note that considering measures $\mathcal{E}$ and $\mathcal{O}$ under a predictive inferential setting allows us to bypass the use of counterfactual quantities and allows us to directly consider the conditional distributions of Y given Z [Arjas, 2012]. For any utility, an importance sampling argument yields

$$\begin{aligned} E_{\mathcal{E}}[U(b^*, g, \beta)|\bar{b}] =& E_{G_{\mathcal{E}}}\left[E_{b^*_{\mathcal{E}}|g}[U(b^*, g, \beta)|g, \bar{b}]\middle|\bar{b}\right] \\ =& E_{G_{\mathcal{E}}}\left[\int_{b^*} U(b^*, g, \beta) p_{\mathcal{E}}(b^*|g, \bar{b}) \frac{p_{\mathcal{O}}(b^*|\bar{b})}{p_{\mathcal{O}}(b^*|\bar{b})}\middle|\bar{b}\right] \\ =& E_{\mathcal{O}}\left[\frac{1}{C_G} \sum_{\{r \in \mathcal{I}\}} w^{*r} U(b^*, g^r, \beta)\middle|\bar{b}\right]. \end{aligned} \qquad (3.3)$$

Randomization to regime g^r is equiprobable for all regimes in our experimental world; this is captured by the constant C_G (See Appendix A.1 for more details). The weights w^r in equation (3.3) are given by

$$w^{*r} = \frac{\mathbb{1}_{g^r(\bar{x}^*)}(\bar{z}^*)}{\prod_{j=1}^{K} p_{\mathcal{O}}(z_j^*|\bar{z}_{j-1}^*, \bar{x}_j^*, \bar{b})}.$$

The denominator is the well-known treatment probability in the observational measure; the numerator is the probability of a sequence of treatments conditional on regime assignment. Note that this weight formula differs from that presented in Saarela et al. [2015b], though the general procedure is the same. For equation (3.3) to hold for the entire support of the data, we require that for each g, $p_{\mathcal{E}}(b^*|g,\bar{b})$ be absolutely continuous with respect to $P_{\mathcal{O}}$; this is equivalent to the positivity condition cited in the causal inference literature. Practically, this means that if a patient following regime g has recorded history $(\bar{x}_k, \bar{z}_{k-1})$ and receives treatment z_k, then in the observational world we should be able to find patients of this sort. Note that as in the frequentist setting, these dynamic MSM weights are *not* stabilized, and the above argumentation clarifies why the usual stabilization is not possible in the DTR framework. Although importance sampling can motivate inverse probability of treatment weighting (IPW) – a classical approach to estimating MSMs in the frequentist setting – the inferential machinery must still come from semiparametric theory. In Bayesian inference, importance sampling and an appropriate prior lead to a method of inference. In the frequentist literature, the linking of two measures is not usually termed importance sampling; this is done via a Radon-Nykodym derivative. This derivative was first used by Murphy et al. [2001] to connect the observational distribution with the distribution in which all patients follow a DTR, and it has been further adapted in works like Johnson and Tsiatis [2004, 2005], Orellana et al. [2010a], and Hu et al. [2018].

Now that we know how to link the expected utility in the experimental worlds with the observational world, we must consider how to infer about the parameter of interest β. Recall that as Bayesian decision makers, our best estimate for β is one that maximizes the posterior expected utility. This requires a posterior distribution to characterize the uncertainty of this maximizer. Consequently, before specifying the utility of choice and before performing the necessary maximization, we must specify a prior. The prior we consider is not placed on β as is done in Bayesian parametric inference; the prior is placed on the family of data generating distributions in the observational world $P_{\mathcal{O}}$, and denoted by $P_{\mathcal{F}}$. In fact, this prior induces a prior on β as $P_B(\beta \in \Omega) = P_{\mathcal{F}}(\{P_{\mathcal{O}} : \beta(P_{\mathcal{O}}) \in \Omega\})$. A robust, non-informative choice of prior in the observational measure is the non-parametric Dirichlet process ($\mathcal{DP}$) prior, which asymptotically concentrates around the true data generating distribution. Stephens et al. [2022] explore in detail the consequences of what the Dirichlet process prior implies for a prior on a functional, like β. Now, when $\mathcal{DP}(\alpha, G_x)$ is chosen such that $|\alpha| \to 0$, we obtain the non-parametric Bayesian bootstrap as the posterior predictive distribution. This Bayesian bootstrap is the same as that employed by Saarela et al. [2015b], however we have been explicit about the assumptions needed to utilize it. This bootstrap is analogous to the Bayesian bootstrap presented in Rubin [1981]. Under this specification, one sample drawn from the posterior $\mathcal{DP}$ is given by $p(b^*|\bar{b}, \pi) = \sum_{i=1}^{n} \pi_i \mathbb{1}_{b_i}(b^*)$, where $\pi = (\pi_1, ..., \pi_n)$ is a sample from $\Pi \sim Dir(1, ..., 1)$, a Dirichlet distributed random variable with all concentration parameters equal to one. Note that under the Bayesian bootstrap assumptions, any distribution sampled from the posterior $\mathcal{DP}$ is uniquely determined by Π. To compute functionals of the posterior predictive, we require $p(b^* \in A|\bar{b}) = E_\Pi[p(b^* \in A|\bar{b}, \Pi)]$ which are estimated by resampling weights $(\pi_1, ..., \pi_n)$ from $Dir(1, ..., 1)$, and computing the average over samples. Consequently, under Bayesian bootstrap assumptions, we compute the expected posterior experimental world utility via:

$$E_{\mathcal{E}}[U(b^*, g, \beta)|\bar{b}] = E_\Pi[E_{\mathcal{E}}[U(b^*, g, \beta)|\bar{b}, \Pi]] = E_\Pi\left[\frac{1}{C_G}\sum_{i=1}^{n}\sum_{r \in \mathcal{I}} \pi_i w_i^{*r} U(b_i, g^r, \beta)\right]. \tag{3.4}$$

β_{opt}, the true maximizer of the expected utility, can be expressed by maximizing the expected posterior utility: $\beta_{opt} = \arg\max_\beta \; E_\Pi\left[\sum_{i=1}^{n} \pi_i \sum_{r \in \mathcal{I}} w_i^{*r} U(b_i, g^r, \beta)\right]$. Furthermore, the uncertainty around β_{opt} may be characterized by noting that β_{opt} is a deterministic function of π, computed as

$$\beta_{opt}(\pi) = \arg\max_\beta \sum_{i=1}^{n} \pi_i \sum_{r \in \mathcal{I}} w_i^{*r} U(b_i, g^r, \beta).$$

Thus, uncertainty in the posterior distribution reflects uncertainty in β_{opt}; this approach to Bayesian inference is discussed by Walker [2010]. We may disregard C_G for the purposes of predictive inference. Modulo Monte Carlo error, this is an exact Bayesian procedure, regardless of the sample size. In work by Saarela et al. [2015b], simulations show that multiplying π_i with importance sampling weights dampens the effect of extreme weights thereby leading to improved variance estimators as compared to those relying on asymptotic approximations, the latter tending to underestimate variance. From equation (3.4), we note that to draw inference in the experimental world, we require an analytic expression for the weight w; this leads us to modeling the treatment assignment probabilities. We touch on this in Section 2.3. Furthermore, we note that inverse probability weighting methods may not be adequate in settings with many stages, as these require us to take the product of many probabilities, thereby leading to large weights and yielding both bias and imprecision [Robins et al., 2008, Scharfstein et al., 1999]. We now present some utilities that allow for causal inference of DTRs.

Utility as Negative Squared Error Loss:

An appealing choice of utility is the negative square error loss given by: $U(b^*, g^r, \beta,) = -(y^* - h(\beta, r))^2$, where $h(\beta, r)$ models $E[y^*|g^r, \bar{b}]$. This leads to solving:

$$\beta_{opt}(\pi) = \arg\max_{\beta} \left[- \sum_{i=1}^{n} \pi_i \sum_{r \in \mathcal{I}} w_i^{*r} (y_i - E[y|g^r, \beta])^2 \right]. \tag{3.5}$$

Again, over repeated draws from Π, this is an exact Bayesian procedure for finite samples, modulo Monte Carlo variation. This procedure allows us to leverage the possibility that patients adhere to multiple DTRs, thereby contributing to the objective function multiple times. Orellana et al. [2010a] show that solving for $\beta_{opt} = \arg\max_{\beta} [- \sum_{i=1}^{n} \sum_{r \in \mathcal{I}} w_i^{*r} (y_i - E[y|g^r, \beta])^2]$ yields a consistent estimator for β when the mean model is correct. We note that dynamic MSMs are not impacted by issues of non-regularity that arise in methods like Q-learning and g-estimation. See Appendix A.2. Analogously, our procedure can be seen to be consistent for β, by computing the posterior expected utility:

$$\begin{aligned}
E_{\mathcal{E}}\left[-(y^* - h(\beta, r))^2|\bar{b}\right] &= -\int_{b^*} \sum_{r\in\mathcal{I}} w^{*r}(y^* - E[y^*|g^r, \beta])^2 p_{\mathcal{O}}(b^*|\bar{b}) db^* \\
&= -\int_{b^*} \sum_{r\in\mathcal{I}} w^{*r}(y^* - E[y^*|g^r, \beta])^2 \frac{1}{n}\sum_{i=1}^{n} I_{b_i}(b^*) db^* \\
&= -\frac{1}{n}\sum_{i=1}^{n}\sum_{r\in\mathcal{I}} w_i^{*r}(y_i - E[y^*|g^r, \beta])^2.
\end{aligned}$$

We see that the β_n which maximizes the equation above is the same as that which solves the estimating equation in Orellana et al. [2010a]. Indeed we see why our approach may be regarded as a way to unify Bayesian inference with dynamic MSMs. Now, we need not limit ourselves to a finite family of regimes. If the family of DTRs is indexed by a continuous parameter, then a relaxed positivity condition described in Orellana et al. [2010a] will allow us to perform inference on values of the index where positivity may not hold. This condition says that instead of requiring that we observe patients who followed all regimes of interest, we require for patients to follow a subset of regimes. More specifically, β in $h(\beta, r)$ may be identified $\forall r \in \mathcal{I}$ even when the positivity assumption fails for some r, and it suffices to observe r for sufficient points such that β is identifiable. For example, a model $h(\beta, r) = \beta_0 + \beta_1 r + \beta_2 r^2$ that is correctly specified is identifiable if positivity is met for at least three values of $r \in \mathcal{I}$. Of course, the model should be correct in the range of inference. For example, if the identified optimal r is far from the range of observed values, we should question the resulting inference. When searching for optimal DTRs via smooth modeling, we must keep in mind that we seek two optimal posteriors: the first is the posterior distribution of $\bar{\beta} = (\beta_{0,opt}, \beta_{1,opt}, \beta_{2,opt})$; the second is the posterior distribution of r_{opt} which is a deterministic function of $\bar{\beta}$.

Utility as Log Likelihood:

If we choose the utility as the log likelihood of the outcome conditional on regime assignment in $\mathcal{E}$, then for repeated samples of Π we can compute

$$\beta_{opt}(\pi) = \arg\max_{\beta} \sum_{i=1}^{n} \pi_i \sum_{r \in \mathcal{I}} w_{i,K}^{*r} \ell(y_i|g^r, \beta). \tag{3.6}$$

The choice of this utility is guided by aiming to minimize the Kullback-Leibler divergence between $\ell(y_i|g^r, \beta)$ and the data-generating distribution. β may describe the relationship between g^r and y for any $r \in \mathcal{I}$ thus making it a target for causal inference. Interestingly, this utility actually allows us to consider conventional parametric Bayesian inference (i.e. likelihood times prior) by making use of the weighted likelihood bootstrap [Newton and Raftery, 1994]. We show that $\sum_r w_{i,K}^{*r} \ell(y_i|g^r, \beta)$ can be regarded as a weighted likelihood in order to connect the Bayesian bootstrapping procedure with the weighted likelihood bootstrap. Denote $\mathcal{A}_i$ as the set of regimes to which patient i adheres, then for $r_1, r_2 \in \mathcal{A}_i$ we have that $w_{\mathcal{A}_i}^{*} = w_K^{*r_1} = w_K^{*r_2}$. These weights are zero otherwise. Then, we may write equation (3.6) as

$$\beta_{opt}(\pi) = \arg\max_{\beta} \sum_{i=1}^{n} \pi_i w_{\mathcal{A}_i}^{*} \sum_{r \in \mathcal{A}_i} \ell(y_i|g^r, \beta). \tag{3.7}$$

Note that $w_{\mathcal{A}_i}^{*} \sum_{r \in \mathcal{A}_i} \ell(y_i|g^r, \beta)$ is a weighted likelihood; in accordance with the weighted likelihood bootstrap, $\beta_{opt}(\pi)$ may be regarded as a sample from the posterior distribution of β under a flat prior. Thus, repeated sampling from this posterior allows for quantification of uncertainty around β. Other priors may be incorporated via sampling importance resampling, but this is not essential and is not the focus of our work.

3.2.3 Implementation

To clearly lay out how to perform Bayesian causal inference using the proposed approach, we provide Algorithm 1. Here, the aim is to obtain a sample from the posterior distribution of $\bar{\beta}$. The algorithm is shown for when the utility is proportional to the squared error loss, or the Normal log likelihood, but it is straightforward to see how it may be adapted to other likelihoods. The data-augmentation procedure described can be further understood from Cain et al. [2010], where a new row of data is created for every regime to which a patient adheres. Recall that equation (3.4) leads us to requiring a model for the weights w. For a given draw of the posterior distribution, we consider the model $p_{\mathcal{O}}(z_j^*|\bar{z}_{j-1}^*, \bar{x}_j^*, \gamma_j(\pi)),\ j = 1, ..., K$. The parameters γ_j may be regarded as coming from a posterior utility maximization framework with the same non-parametric prior. When the utility is the log-likelihood, we solve:

$$\gamma_j(\pi) = \arg\max_{\gamma_j} \sum_i^n \pi_i \log p_{\mathcal{O}}(z_{i,j}|\bar{z}_{i,j-1}, \bar{x}_{i,j}, \gamma_j).$$

Then, for every draw of Π, we first fit the weighted treatment propensity model and use the resulting weight $w(\pi)$ in equation (3.5). By computing $E_{\Pi}\{E_{\mathcal{E}}[U(b^*, g, \beta)|\bar{b}, \Pi]\}$, we are indirectly incorporating the uncertainty about γ_j into the estimation procedure.

```
Data: $DATA_{\mathcal{O}}$
for $r \leftarrow 1$ to $C_G$ do                // Create $AUGDATA_{\mathcal{O}}$ based on regime adherence
    Replicate rows of $DATA_{\mathcal{O}}$ for patients adherent to regime $g^r$
end
Posit model for $h(r, \beta)$
for $i \leftarrow 1$ to $B$ do                  // B is number of posterior draws
    Draw $\pi = (\pi_1, ..., \pi_n)$ from $\sim Dir(1, ..., 1)$
    Estimate $p_{\mathcal{O}}(z_k|\bar{z}_{k-1}, \bar{x}_k, \gamma_j, \pi)\ \forall k$
    Compute weights $w_i(\pi),\ i = 1, ..., n$        // n is number of patients

    Add weights to $AUGDATA_{\mathcal{O}}$
    Run regression with mean $h(r, \beta)$ and with weights $\pi_i w_i^r(\pi)$
end
Output: Posterior distribution of $\beta^*$
$DATA_{\mathcal{O}}$ is input data with one row per patient and is used to fit treatment models.
$AUGDATA_{\mathcal{O}}$ is augmented data, where patients are duplicated for as many DTRs as
they adhere to. This dataset is used to run regression for $h(r, \beta)$.
```

Algorithm 1: Fitting procedure for Bayesian dynamic MSM.

3.3 Predictive Doubly Robust Bayesian Inference for DTRs

In the frequentist literature, inverse probability of treatment weighting (IPW) is known to be an inefficient semiparametric procedure; it also yields inconsistent inference if the treatment models are miss-specified. To gain efficiency and robustness, researchers can consider the doubly robust estimator for the marginal mean of a DTR. This requires identifying a series of conditional outcome models, so that consistent inference is attained when either a set of treatment models *or* a set of outcome models is correctly specified. We now use some of the inferential framework presented in the previous section, and first developed in Saarela et al. [2016], to arrive at Bayesian doubly robust inference for the expected outcome of a DTR g. Though the underlying mechanics hinge on the developments of Saarela et al. [2016], examining and evaluating the use of this doubly robust estimator in a sequential DTR setting is of scientific pertinence. For reasons that will be elaborated on in the following, we no longer seek to model in a unified manner the expected outcome for regimes in a family $\mathcal{G}$, and therefore no longer consider inference via utilities. To preserve the notation we have developed so far, it is enough to consider a family $\mathcal{G}$ containing a single DTR. Consequently, identifying optimal DTRs now requires evaluating the doubly robust estimator to be proposed at each DTR of interest and comparing the expect outcomes. Effectively, these are expectations in a regime-enforced world, where everyone in the study population follows a regime g; this contrasts the previously considered experimental world where patients are randomized to DTRs in a family. With this aim in mind, consider a sequence of conditional predictive outcomes ϕ^*_{k+1}, $k = 1, ..., K$. For $k = K$, these are defined as

$$\phi^*_{K+1}(\bar{x}^*_K) = E_{\mathcal{O}}[y^* | \bar{x}^*_K, \bar{z}^*_K = \bar{g}_K(\bar{x}_K), \bar{b}]. \tag{3.8}$$

For $k = K-1, ..., 1$, ϕ^*_{k+1} are defined as

$$\phi^*_{k+1}(\bar{x}^*_k) = E_{\mathcal{O}}[\phi^*_{k+2}(\bar{x}_{k+1})|\bar{x}_k, \bar{z}^*_k = \bar{g}_k(\bar{x}^*_k), \bar{b}]. \tag{3.9}$$

These are expected outcomes in the observational world, conditional on subjects who had covariate history $\bar{x}_k$ and that followed the regime g up to time k. It can be shown via a conditional expectation argument that $E_g[y^*|\bar{b}] = E_{\mathcal{O}}[\phi^*_2(x^*_1)|\bar{b}]$, the estimand of interest.

Next, we describe how models for ϕ^*_k may be fit in a Bayesian framework; following this, we motivate the doubly robust estimator when models for ϕ^*_{k+1} are correct or when models for w^*_k are correct. Based on the de Finetti representation in equation (3.2), we see that outcome models are parameterized by τ such that $\phi^*_{k+1}(\bar{x}_k) = \phi^*_{k+1}(\bar{x}_k; \tau)$. From equations (3.8) and (3.9) we see exactly how a model should be fit for the mean of the conditional outcomes. We should begin by fitting a model for time point $k = K$ and continue backward; the outcomes for stage k can be computed once a model for stage $k+1$ has been fit. We can treat uncertainty in τ analogously to how we treated uncertainty in γ, the parameter corresponding to the treatment assignment model in the observational world: we make it dependent on Π via a non-parametric, non-informative prior. However, instead of posing a likelihood model as was done for the treatment assignment mechanism, we consider the negative squared error loss utility and posit a model for the conditional outcomes. Then, for every draw of Π, we can estimate $\phi^*_{k+1}(\bar{x}_k, \pi) = E_g[y^*|\bar{x}^*_k = \bar{x}_k, \pi, \tau(\pi)]$. In Appendix A.3.1, we provide details as to how τ may be estimated and incorporated into the inferential procedure.

Ultimately, we seek to estimate $E_g[y^*|\bar{b}]$ unbiasedly either when the conditional outcome models are correct, or when the treatment models are correct. This may be achieved by noting the following equality, which follows directly from Orellana et al. [2010a]:

$$E_g[y^*|\bar{b}] = E_{\mathcal{O}}\left[\phi^*_2(\bar{x}^*_1) + \sum_{k=2}^{K} w^*_{k-1}(\phi^*_{k+1}(\bar{x}^*_k) - \phi^*_k(\bar{x}^*_{k-1})) + w^*_K(y^* - \phi^*_{K+1}(\bar{x}^*_K))\middle|\, \bar{b}\right]. \tag{3.10}$$

From (3.10), we see that when outcome models are correct the estimator is unbiased (see Appendix A.3.2). To see that it is an unbiased estimator when treatment models are correct, we change the form of the estimator. Define $h(\bar{b}) = E_g[y^*|\bar{b}]$ and add $0 = \sum_{k=1}^{K} w_{k-1}^*[h(\bar{b}) - h(\bar{b})]$ to obtain

$$E_g[y^*|\bar{b}] = E_{\mathcal{O}}\left[h(\bar{b}) + w_K^*\left(y^* - h(\bar{b})\right) - \sum_{k=1}^{K}(w_k^* - w_{k-1}^*)(\phi_{k+1}^*(\bar{x}_k^*) - h(\bar{b}))\middle|\, \bar{b}\right], \tag{3.11}$$

where $w_0 \doteq 1$. In Appendix A.3.2, we show how to arrive at this equation and that it is unbiased.

Now that we have identified our estimator of choice for any posterior distribution, let us use the same prior used in the singly robust case and obtain the Bayesian non-parametric bootstrap as the posterior. Then, conditional on a posterior draw, we write (3.10) as

$$E_g[y^*|\bar{b}, \Pi] = \sum_{i=1}^{n} \pi_i \left[\phi_{i2}^*(x_{i1}) + \sum_{k=2}^{K} w_{ik-1}^*(\phi_{ik+1}^*(\bar{x}_{ik}) - \phi_{ik}^*(\bar{x}_{ik-1})) + w_{iK}^*(y_i - \phi_{iK+1}^*(\bar{x}_{Ki}))\right]. \tag{3.12}$$

Models for the ϕs and ws now depend on Π and may be incorporated into the inferential process as in (3.2.3). Furthermore, we may compute $E_g[y^*|\bar{b}] = E_\Pi\left[E_g[y^*|\bar{b}, \Pi]\right]$ by resampling Dirichlet weights, thereby enabling us to obtain a doubly robust estimator for the value of a DTR, including its uncertainty. As mentioned, the doubly robust Bayesian estimator proposed is only for the marginal mean of a DTR, not for the parameters in a model for the marginal mean linking a family of DTRs (e.g $E[y^*|\bar{b}, g^r] = \beta_0 + \beta_1 r + \beta_2 r^2$). In order to obtain doubly robust estimators of the latter, an appropriate utility would have to be proposed so that when importance sampling is used to link the experimental world with the observational world, the obtained expression in the observational world is doubly robust. Then, to use the proposed estimator to identify optimal DTRs, we are required to perform a grid search. Murphy et al. [2001] suggested that outcome models should be coherently parameterized so that for $k_2 > k_1$, a model conditional on information up to time k_2 would yield a model conditional on information up to time k_1 when covariates between k_2 and k_1 are marginalized.

3.4 Individualized Decision Making

Now that we have developed the inferential approach, we turn our attention to examining how to incorporate this into an individualized decision-making scheme. This consideration is particular to the DTR setting that we explore. For illustrative purposes, we focus on the following class of regimes: treat if $x_k > \theta$ for $k = 1, ..., K$. Suppose that a new patient is observed with covariate value x_1^{new}. Our interest is in deciding whether this patient should be treated based on our belief about the optimal θ. To do this, we are interested in computing $P(\theta^*_{opt} < x_1^{new}|\bar{b})$. This may be done by taking a sample of size m from the posterior distribution and computing $p_1 = (1/m)\sum_\theta \mathbb{1}(\theta^*_i < x_1^{new})$. Indeed this can be done for all stages p_k. Effectively, this probability is informing the decision-maker about how certain they should be in switching treatment given the patient's current health status, if the aim is to select an optimal therapy. It is then up to the decision-maker to make a treatment decision given that probability. Note that a patient's decision about treatment at a given stage does not alter the optimality of consequent decision rules, though it may alter the optimality of the overall treatment course. This individualized approach may be taken with any optimal regime derived through the proposed methodology, and we elaborate on this in the simulations.

3.5 Simulations

In this section, we use simulations to evaluate how this Bayesian approach to inference can be used to infer about optimal DTRs. We focus on multi-stage problems with a sample size of $n = 500$. All results are presented over 500 Monte Carlo replications. For comparison, we also provide results for the frequentist approach. Generally the strategy was to induce time varying confounding with treatment-confounder feedback. All intermediary variables were Gaussian, and all treatment variables Bernoulli. We followed the approach in Stephens [2015] to generate outcomes that allowed for the analytic identification of the optimal regimes. The true value (expected outcome) under the optimal regime was obtained by generating a large sample of data in which patients adhered to the optimal regime. Further simulation details can be found in Appendix A.4, as well as results for other sample sizes and for when intermediary variables are Gamma-distributed.

For simulation I, we considered a family of regimes indexed by θ_1, θ_2 where treatment is assigned when x_k exceeds θ_k, $\theta_k \in [0, 1]$, $k = 1, 2$. The known optimum is $(\theta_{1opt}, \theta_{2opt}) = (0.4, 0.8)$ and the outcome $y = x_1 - (-\theta_{1opt} + x_1)(\mathbb{1}_{\theta_{1opt} > z_1} - z_1) - (-\theta_{2opt} + x_2)(\mathbb{1}_{\theta_{2opt} > z_2} - z_2) + \sqrt{0.5}\epsilon$, $\epsilon \sim N(0, 1)$. We evaluate the performance of both the IPW and doubly robust estimator thereby leading us to compute these estimators for discrete values of $\theta_k \in \{0, 0.1, 0.2, ..., 0.9, 1\}$. Table 3.1 shows the results of the estimation procedure. The first column indicates the type of estimation procedure that was used. The second refers to the model specification. For the doubly robust estimator "None" means that both treatment and outcome models are miss-specified ; "Treat" means the treatment models are correctly specified; "Outcome" means that outcome models are correctly specified; "Both" means all models are correctly specified. "IPW" refers to the IPW estimator with correctly specified treatment models. For incorrectly specified models, we use intercept-only regressions. For the Bayesian approach, point estimates are provided at the posterior mean. For simulation I, the mean outcome at the optimal regime can be seen (from the data-generating mechanism) to be 0.

In Table 3.1 we observe that estimators with at least one set of models correct are unbiased. As expected, when the treatment and outcome models are all correctly specified, efficiency is maximized. The coverage probability measures the proportion of time that the true optimum is inside a 95% credible interval, across replications. As far as we are aware, there is no way to obtain a confidence interval for the optimal threshold in the frequentist setup. This is because we have evaluated the estimator in a grid of thresholds θ and identified the $\hat{\theta}_{opt}$ that maximizes the mean outcome; for the Bayesian setup, we have sampled the posterior distribution of θ_{opt}. "Estimated Outcome Train Pop." refers to estimated expected outcome under the optimal regime, this is known to be 0; "Mean Outcome Test Pop." refers to the mean outcome under the optimal DTR, in a new population with a different distribution for intermediate covariates. Thinking about the mean outcome in a test population allows us to contemplate how the identified optimal DTR will perform once deployed in the real world. We see that the frequentist and Bayesian methods perform similarly, and surprisingly the "no models correct" scenario leads to good performance in the testing set, though this is due in part to the scale of the value function which has a narrow range (see Appendix A.4). The uncertainty measures for $\theta_{k,opt}$ appear to be slightly higher for the Bayesian analysis than for the frequentist analysis. One reason for this may be that the Bayesian method acknowledges uncertainty in the outcome and treatment models, whereas the frequentist method takes these as known. The coverage probability for θ_1 in the no models correct scenario is low, and surprisingly it is close to nominal for θ_2. For the other setups, the coverage probabilities are slightly higher than their nominal value. Of course, it is important to keep in mind that this was a discrete problem and the coverage probabilities depend on the coarseness of the exploration grid; we have observed in other simulations that finer grids lead to further tightening of the confidence intervals toward the nominal value (results not shown). However, this must be balanced with the computational costs of an estimation procedure on a fine grid.

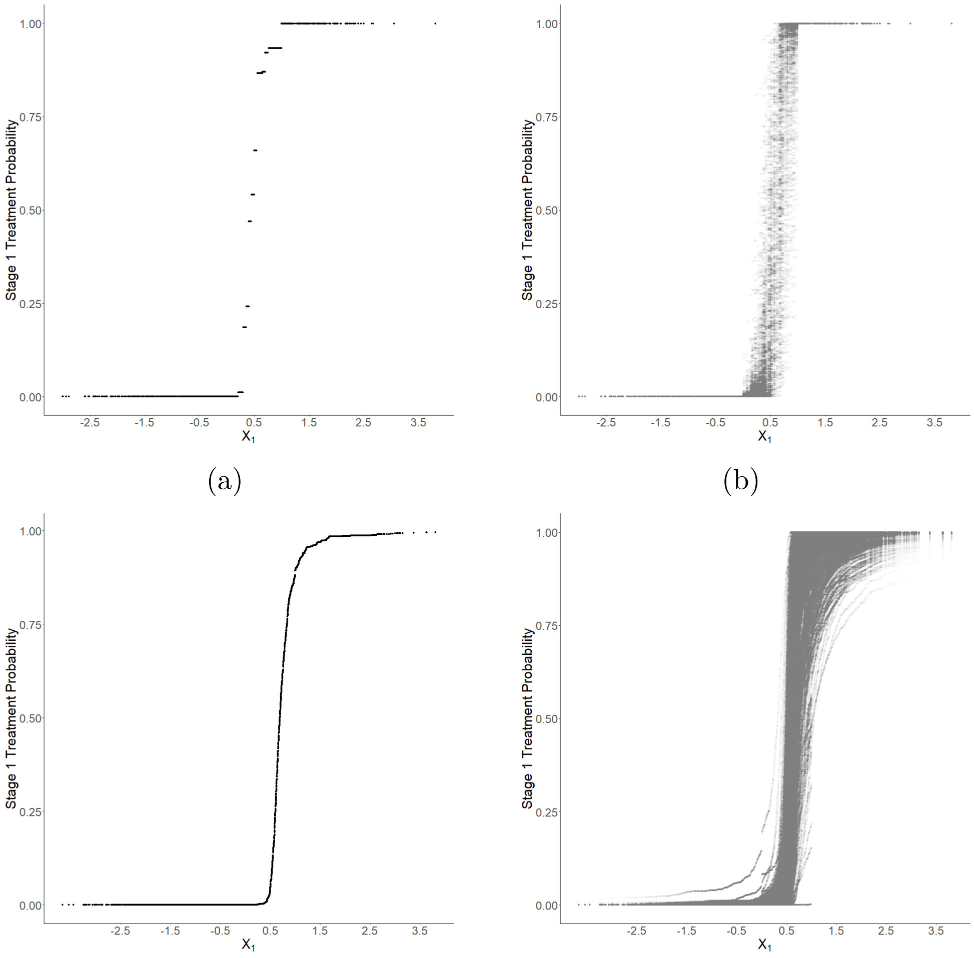
Now, we can ask whether newly observed patients will benefit from the estimated optimal rule. For illustration, we restrict the family of regimes to have a common threshold across periods: $\theta_1 = \theta_2 = \theta$, with $\theta_{opt} = 0.6$ (see Appendix A.4). Figure 3.1(a) shows the probability that a patient should receive treatment $z = 1$ at stage 1 for a single Monte Carlo replicate. This is a step function as θ was computed over a set of discrete values. Patients

Table 3.1: Results for simulation I (n=500; 500 Monte Carlo replicates). Point estimates are means across Monte Carlo replicates; standard deviations are Monte Carlo standard deviations.

Method	Model Correct	Posterior Mean of $\hat{\theta}_{1opt}$	Posterior Mean of $\hat{\theta}_{2opt}$	Estimated Outcome Train Pop.	Coverage Probability $\theta_{1opt}, \theta_{2opt}$	Mean Outcome Test Pop.
Frequentist	None	0.247 (0.116)	0.641 (0.183)	0.250 (0.120)	—	0.587 (0.012)
Frequentist	Treat	0.468 (0.232)	0.753 (0.207)	0.045 (0.066)	—	0.584 (0.017)
Frequentist	Outcome	0.385 (0.193)	0.735 (0.210)	0.022 (0.065)	—	0.588 (0.014)
Frequentist	Both	0.415 (0.182)	0.793 (0.162)	0.018 (0.056)	—	0.591 (0.011)
Frequentist	IPW	0.441 (0.205)	0.747 (0.209)	0.035 (0.064)	—	0.587 (0.014)
Bayesian	None	0.246 (0.124)	0.641 (0.192)	0.271 (0.119)	0.860, 0.914	0.586 (0.012)
Bayesian	Treat	0.480 (0.253)	0.759 (0.203)	0.070 (0.064)	0.990, 0.964	0.582 (0.019)
Bayesian	Outcome	0.371 (0.207)	0.737 (0.232)	0.037 (0.065)	0.974, 0.986	0.585 (0.015)
Bayesian	Both	0.414 (0.194)	0.797 (0.166)	0.029 (0.056)	0.978, 0.974	0.590 (0.012)
Bayesian	IPW	0.454 (0.218)	0.761 (0.214)	0.055 (0.063)	0.990, 0.964	0.585 (0.017)

with low and high values of x_1 experience high certainty as to whether they should receive optimal treatment or not. Patients whose covariate is near the true optimal threshold of 0.6 experience low certainty. Figure 3.1(b) shows the same result across 500 Monte-Carlo replicates, emphasizing that there is high uncertainty around the true value. It can also be useful to obtain a smooth decision curve. This may be done by evaluating the doubly robust estimator over a much finer grid of points or by modeling $E[y^*|\bar{b}, g^\theta]$ via a smooth function such as $\beta_0 + \beta_1\theta + \beta_2\theta^2$ (quadratic) and using IPW. Figure 3.1(c) shows the results of the individualized rule with the quadratic model and IPW estimator; the decision rule is much smoother and provides high certainty for most values of x_1, except for those closest to 0.6. Figure 3.1(d) shows the Monte Carlo variation around this curve; most uncertainty is around the true value of the threshold.

For simulation II, we explore a family of regimes indexed by ψ_1, ψ_2, ψ_3 such that $\psi_1 x_{k1} + \psi_2 x_{k2} > 0.5 - 3\psi_3 u; k = 1, ..., 4$; x_{k1}, x_{k2} are normally distributed intermediary covariates and u is a binary baseline covariate. This regime has an interpretation that treatment should be given if the weighted sum of x_{k1} and x_{k2} is above a threshold, and this threshold depends on patients' baseline covariate u. Increments of 0.05 were used for ψ_1, ψ_2 and of 0.1 for ψ_3.



(a)

(b)

(c)

(d)

Figure 3.1: Simulation I, stage 1 individualized treatment probabilities: (a) Individualized decision rule using doubly robust estimator with only the treatment model correct; (b) Same as (a) over 500 Monte Carlo replicates; (c) Individualized decision rule using IPW with a quadratic MSM; (d) Same as (c) over 500 Monte Carlo replicates.

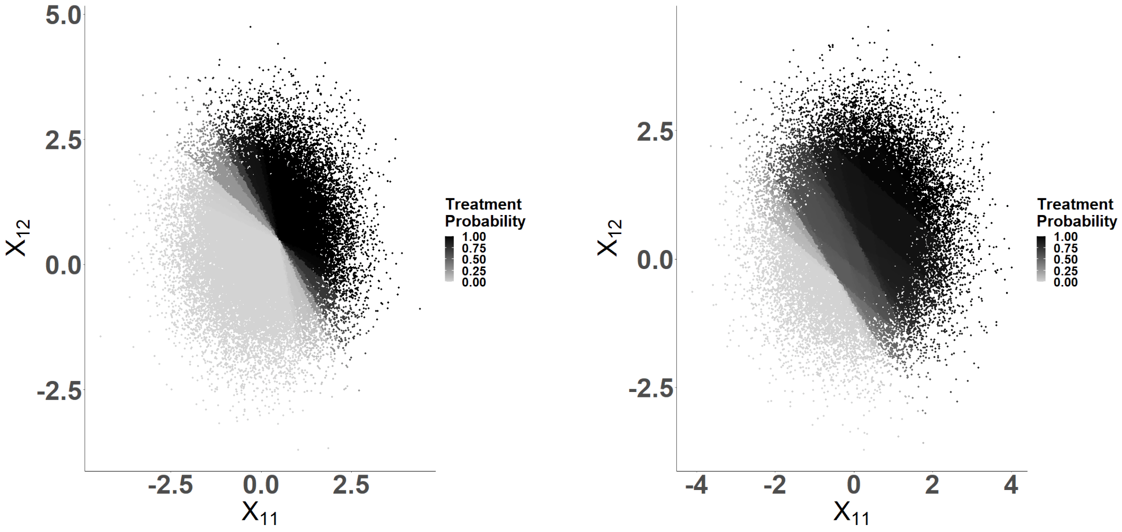
Appendix A.4.3 shows the data generating mechanism for this setup. The optimal regime is given by $\psi_{1opt} = \psi_{2opt} = 0.5$, $\psi_{3opt} = 0.1$, with a value of 1. We see from Table 3.2 that all scenarios, except the no models correct scenario are unbiased, with the all models correct scenario yielding the best results. Correctly specifying the outcome model provides improvement in the estimation of the value at the optimum over just getting the treatment model correct. We do not include a ψ_2 column in the table, as the constraint $\psi_1 + \psi_2 = 1$ makes this redundant. We note again that the coverage probabilities are high, recall that this is driven by the coarseness of the exploration grid; a finer grid in this problem would be very computationally intensive. Appendix A.4.2 presents a similar simulation without the binary covariate.

In Figure 3.2 we further illustrate how the Bayesian framework can be leveraged for individualized inference. We observe, for one replicate, the probability that a patient should be treated under the optimal decision rule, given a set of covariates. These probabilities are computed by using the posterior distribution of $\psi_{1opt}, \psi_{2opt}, \psi_{3opt}$ via $P(\psi^*_{1opt}x_{11} + \psi^*_{2opt}x_{12} + \psi^*_{3opt}u > 0.5)$. There are regions of high certainty that indicate patients should or should not receive treatment according to the optimal rule; there are also regions with more uncertainty regarding the choice of optimal treatment. In fact, patients with baseline covariate $u = 0$ face higher uncertainty overall than those with $u = 1$.

Table 3.2: Results for simulation II (n=500; 500 Monte Carlo replicates). Point estimates are means across Monte Carlo replicates; standard deviations are Monte Carlo standard deviations.

Method	Model Correct	Posterior Mean of $\hat{\psi}_{1opt}$	Posterior Mean of $\hat{\psi}_{3opt}$	Estimated Outcome Train Pop.	Coverage Probability ψ_{1opt}, ψ_{3opt}	Mean Outcome Test Pop.
Freq.	None	0.590 (0.126)	0.103 (0.104)	2.003 (0.355)	—	0.526 (0.064)
Freq.	Treat	0.479 (0.157)	0.101 (0.125)	1.160 (0.155)	—	0.530 (0.057)
Freq.	Outcome	0.503 (0.048)	0.102 (0.020)	1.004 (0.068)	—	0.581 (0.010)
Freq.	Both	0.499 (0.031)	0.100 (0.004)	1.000 (0.065)	—	0.585 (0.004)
Freq.	IPW	0.464 (0.157)	0.089 (0.134)	1.198 (0.184)	—	0.529 (0.055)
Bayes.	None	0.589 (0.123)	0.094 (0.106)	2.200 (0.351)	0.952 0.996	0.549 (0.022)
Bayes.	Treat	0.481 (0.165)	0.089 (0.124)	1.254 (0.150)	0.992 1	0.539 (0.025)
Bayes.	Outcome	0.498 (0.050)	0.101 (0.016)	1.008 (0.066)	0.994 1	0.587 (0.005)
Bayes.	Both	0.497 (0.029)	0.100 (0.004)	1.001 (0.064)	1 1	0.591 (0.003)
Bayes.	IPW	0.468 (0.163)	0.072 (0.130)	1.317 (0.198)	0.992 1	0.537 (0.024)

There is some debate in the literature on choice of doubly versus singly robust estimators; see e.g. Kang and Schafer [2007] and Bang and Robins [2005]. Our simulations emphasize that a lot is to be gained, in precision and accuracy, if we correctly specify the outcome models, when compared to the doubly robust estimator with only treatment models correct or the IPW estimator. Efficiency is maximized when all models are correct, thereby clarifying that



(a)

(b)

Figure 3.2: Simulation II individualized treatment probabilities using IPW estimator; (a) Stage 1 treatment probability for those with $u = 0$ (b) Stage 1 treatment probability for those with $u = 1$.

these considerations are not just theoretical; they also impact analyses with finite sample size. When deciding whether to use the singly robust or the doubly robust estimator, it is important to ask what is better understood: the treatment assignment process, or the outcome process.

3.6 Case Study: Analysis of the NA-ACCORD

Treatment for HIV infection with antiretroviral therapy (ART) must be lifelong to maintain control of HIV viral replication and improve immune function. Consequently, there is concern that some combinations of drugs may cause long-term harm. The multi-drug nature of this therapy allows for some flexibility in treatment course. Research by Klein et al. [2016] is consistent with the possibility that some ART agents contribute to long term liver damage in patients with chronic hepatitis C (HCV) infection. ART agents, like protease inhibitors (PI), may also help reduce adverse liver outcomes by providing virologic control [Macías et al., 2006], while also having some detrimental effects on liver health [Young et al., 2021]. We examine how to tailor ART therapy to reduce liver damage by exploring the use of Bayesian dynamic MSMs for tailoring therapy to patients' FIB4 score, an age-adjusted score that quantifies liver fibrosis; higher values indicate greater damage [Sterling et al., 2006]. We aim to identify whether there is an optimal FIB4 score at which patients should switch therapy, in order to minimize subsequent FIB4. In particular, for the purposes of demonstrating the use of the proposed methods, we explore the effect of switching into PI (z=1) and away from any other ART regimen (z=0) when FIB4 score surpasses a level θ, and when all patients start out on a non-PI based therapy. This is a thresholding regime, where we search for the optimal θ in the DTR: switch when FIB4$> \theta$.

We use data from the NA-ACCORD to identify a cohort of patients who initiated ART therapy from 2004 onwards, the period in which modern ART treatments were approved. Patients in this cohort may or may not have other viral infections, such as HCV and hepatitis B (HBV). Study initiation (time zero) is the first instance of ART treatment, after which patients are followed-up for a 12 month exposure ascertainment period. It is in this period where we may examine which DTRs patients follow. Lastly, outcomes are taken to be the first FIB4 measurement 18 to 30 months after study initiation. The outcome observation period is as defined because liver measurements are not taken at every follow-up visit, though they should occur at least annually as per standard of care. Patients are lost to follow-up if they stopped receiving ART, had missing ART records, or if they did not have an observed outcome. The range of thresholds is determined by the fifth and ninety-fifth quantile of FIB4 scores at baseline. We identify patient records every six months and record the treatment that patients received. Potential confounders included were: time-varying CD4 cell count, time-varying viral load, and the following baseline variables: insurance status, indicator of risky alcohol consumption, drug use, HCV status, HBV status, race, and sex.

Based on the six-month observation intervals, there were a total three decision points, each requiring a set of models. Potential confounders were identified a priori through discussions with a subject matter expert. Stage-specific propensity scores were then fit to achieve balance across treatments at each time point. Censoring weights were incorporated to eliminate selection bias. For the doubly robust estimator, it was assumed that the variables in outcome models explained both confounding and/or selection. The models that were fit can be found in Appendix A.5. Sensitivity analyses were performed in order to determine whether results were sensitive to model specifications. Balance from the propensity scores was assessed using SMD and by using a frequentist fit of the propensity scores. Balance was examined at all stages. Outcome models were examined to ensure the predicted distribution did not differ from the observed.

For a fixed value of θ, patients are indicated to switch treatments when their FIB4 measurements surpass θ. Accordingly, patients in the study could be categorized into five groups for each regime (g^{θ}) considered: those 1) indicated to switch but did not switch (ISNS), or switched at the wrong time; 2) indicated to switch and switched (ISS); 3) not indicated to switch and did not switch (NISNS); 4) not indicated to switch and switched (NISS); and 5) those who were assigned to PI at baseline (NR). Patients indicated to switch were given six months to do so (a grace period). To improve the properties of the estimators, we normalized the weights in the analysis and assessed positivity for each candidate regime by checking whether the distribution of the propensity scores at each interval for the modeled treatment are similar in the regime adherent group and the regime non-adherent group. The propensity to switch treatment was generally small, highlighting that relatively few individuals contribute to the estimation of our regime of interest – a limitation that must be acknowledged; more details can be found in Appendices A.5.3 and A.5.4. Only patients in the ISS and NISNS groups could adhere to a regime for the full study period. Consequently, patients in the other groups were artificially censored when they deviated off the specified regime. 95% credible intervals were calculated for all point estimates, approximated using 500 draws from the posterior distribution; point estimates were reported at the posterior mean. Details of the analysis plan can be found in Appendix A.5. We evaluated the estima-

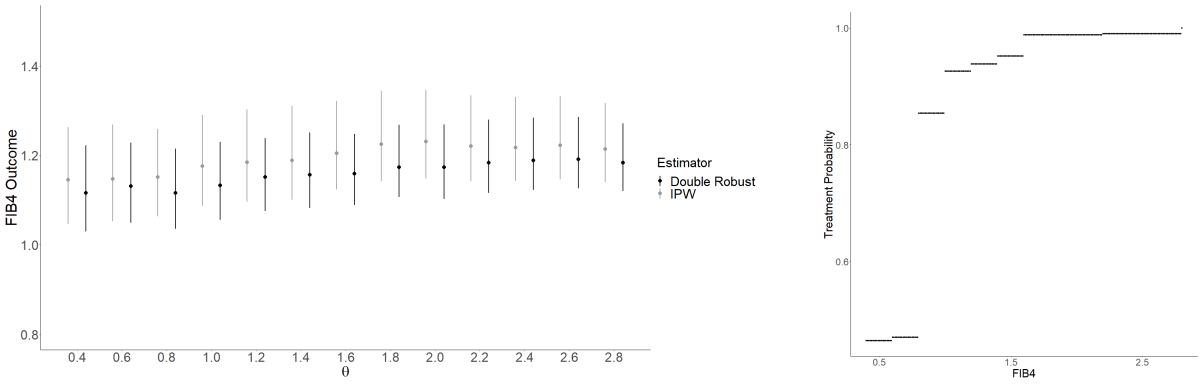
Table 3.3: NA-ACCORD case study: follow-up information for a subset of regimes ($n = 22,768$).

θ	ISNS	ISS	NISNS	NISS	NR	Uncensored	Uncensored
						ISS	NISNS
0.4	12172	611	244	8	9733	412	244
1.0	6798	398	5618	221	9733	276	5618
1.6	3194	213	9222	406	9733	143	9222
2.2	1732	143	10684	476	9733	89	10684
2.8	1136	111	11280	508	9733	73	11280

Note: ISNS=Indicated to switch & did not switch; ISS=Indicated to switch & switched; NISNS=Not indicated to switch & did not switch; NISS=Not indicated to switch and switched; NR=Received PI at baseline.

tors at thresholds of 0.4 to 2.8 in units of 0.2; the minimum and maximum threshold value correspond to the 5^{th} and 95^{th} percentile of the FIB4 distribution. In Table 3.3, we present follow-up information for a subset of these regimes. We did not posit a marginal structural model as a function of θ (e.g. a quadratic form) as we wanted to make use of both the IPW and doubly robust estimators. Although our overall sample size is large, we see that only half of patients follow a non-PI ART regimen at study start. Additionally, roughly 30% of ISS and NISS patients are censored or artificially censored. The number of NISNS patients varies strikingly across regimes. However, this is to be expected: for a threshold of 0.5, only a small proportion of patients are not indicated to switch, and a relatively large proportion of patients switch in the first year of the study. The sample size in the ISS group is generally low, which is unfortunate. In part, this is due to the fact that when patients are indicated to switch, not only should they switch, but they should switch within the indicated time. The sample size in the ISS group is further reduced for large values of θ as for these values, only a small number of patients would be indicated to switch.

From Figure 3.3 (a), we confirm that we are underpowered to detect any differences in final FIB4 scores, and that the doubly robust estimator provided some gains in efficiency. It is noteworthy that FIB4 scores drop overall at the end of the study, compared to the baseline values. We note that from this figure, there is no interior point that clearly minimizes FIB4



(a) (b)

Figure 3.3: (a) Mean FIB4 score under each DTR based on Bayesian IPW and doubly robust analyses, with 95% credible intervals (from 500 posterior draws). (b) Individualized treatment probability using doubly robust estimator. Note that in (a) the points corresponding to each method are presented out of phase for illustrative purposes. In reality, points are on top of each other starting at 0.4 and continuing in increments of 0.2.

score, thereby suggesting that there is no benefit to tailoring. A threshold of $\theta = 0.4$ yields a DTR that is very close to the static treatment always switch into PI. Though this may raise the question as to why patients would be given a drug other than PI, we remind the reader that there are a variety of other ART treatments, some of which may be more beneficial and some which may be more detrimental. From Table 3.4 we can examine the expected outcomes for a subset of regimes. We note that the IPW and doubly robust estimator yield very similar point estimates across most regimes; both estimators point to the same conclusions. In addition, Figure 3(a) also leads us to question the utility of individualized inference in this scenario. Though the figure shows a relatively flat relationship between the value function and the threshold (with considerable uncertainty), the value function under adherence to each candidate regime is not flat, as is shown in Appendix Figure A.4. Consequently, we can ask the probability that a patient's FIB4 value is greater than the optimal threshold. This results in Figure 3.3(b), which indicates that when a patients FIB4 score is at 0.8 or greater, they have a high probability of being above the optimal threshold. We discuss this further in Appendix A.5.6.

This analysis had several limitations. First, the follow-up may have been too short for

Table 3.4: NA-ACCORD case study: expected FIB4 (outcome) under adherence to regime θ. Numbers in brackets are posterior standard deviations.

θ	IPW	Doubly Robust
0.4	1.145 (0.054)	1.116 (0.048)
1.0	1.176 (0.051)	1.133 (0.044)
1.6	1.205 (0.048)	1.159 (0.039)
2.2	1.221 (0.048)	1.183 (0.040)
2.8	1.214 (0.045)	1.184 (0.039)

the outcome of interest, as switching therapies may not have an immediate effect on liver scarring; this is likely a long-term process. The reason for the short follow-up was that after the first year, therapeutic switches were relatively rare. Also, there was a trade-off in extending the follow-up time: it would allow for more therapeutic switches but also increase artificial censoring due to going off regime. Though many confounders were included in the analysis, some may have been missed. Importantly, we did not have information on why patients switched therapy. Additionally, it would have been beneficial to study only patients co-infected with HCV and HBV, as these are at higher risk of liver complications. However, sample size limitations did not allow for this.

3.7 Discussion

In this work, we explored recently developed Bayesian semiparametric methods to infer about optimal DTRs. For this purpose, we sought to transparently develop a way to utilize Bayesian dynamic MSMs, this involved targeting experimental world causal parameters when only observational world data was available. We also inferred about optimal DTRs via posterior predictive inference and a doubly robust estimator; this approach had not been studied in a longitudinal DTR setting. Our simulations showed that the proposed methods work well, though they exhibit slightly more variability than their frequentist counterpart. The analysis of the NA-ACCORD provided a demonstration of how these methods might be used in clinical research, though we note that the results were limited by the fact that therapeutic switching was infrequent in practice. Still, this case study aimed to show that our proposed inference could be implemented meaningfully. Though our approach does not necessitate counterfactual notation, the idea of counterfactuals still permeates this work; the experimental world considered is indeed a world where, counter to fact, patients have been randomized to a specific treatment strategy of interest. Additionally, the resulting conditional posterior predictive quantities are equal to their counterfactual counterparts in this unconfounded world. Throughout, we focused on the non-parametric Bayesian bootstrap in order to draw inference in a non-informative, robust way. Indeed our choice of prior allowed us to connect our approach to the way frequentist semiparametric estimators are obtained. Though these methods may feel different, they have the same ingredients that appear in conventional Bayesian analyses. A prior leads to posterior inference in the observational world, and importance sampling allows us to infer about worlds that are of scientific interest. When we are interested about inferring about parameters in a utility, the Dirichlet process prior that we make use of implicitly induces a prior on these parameters; these ideas as explored further in Stephens et al. [2022]. We remind the reader that the proposed method is valid for any sample size. We also note that methods discussed herein are not limited to decisions taken at fixed dates; they may also be triggered by events. For example, a second-line therapy may be given only when first-line therapy lacks efficacy, as in Krakow et al. [2017].

3.8 Software

Software in the form of R code can be found on GitHub on the following link:
https://github.com/Danroduq/semiparametric-Bayesian-DTRs.

3.9 Supplementary Material

Supplementary material is available online at http://biostatistics.oxfordjournals.org.

Acknowledgments

DRD is supported by a doctoral fellowship from the National Sciences and Engineering Research Council of Canada (NSERC). EEMM and DAS acknowledge support from Discovery Grants from NSERC. EEMM is supported by a career award from the Fonds de recherche du Québec - Santé and a Canada Research Chair (CRC) in Statistical Methods for Precision Medicine. MBK is supported by a CRC Tier I and reports grants for investigator-initiated studies from ViiV Healthcare, Merck, and Gilead; and consulting fees from ViiV Healthcare, Merck, AbbVie and Gilead. The authors are grateful to the NA-ACCORD, and the full acknowledgment can be found in the Appendix. The content of this manuscript is solely the responsibility of the authors.

Chapter 4

Estimation of Optimal Dynamic Treatment Regimes using Gaussian Process Emulation

Preamble to Manuscript 2. The Bayesian dynamic MSMs considered in the previous chapter are an important contribution to Bayesian inference as they allow for a robust Bayesian inferential procedure that obviates some of the usual challenges in Bayesian inference for optimal DTRs. However, frequentist and Bayesian dynamic MSMs are not without their limitations. These require that a model for the value function of regimes in a family be specified; incorrectly specified models can lead to identifying as optimal a DTR which is in fact not. Notwithstanding, value-search methods like dynamic MSMs have some useful characteristics upon which to capitalize. When using regression-based methods, it may be necessary to posit flexible models in order to correctly identify the optimal DTR, thereby resulting in optimal DTRs that are not clinically interpretable. Contrastingly, value-search methods, specifically those that restrict themselves to a family of regimes, do not face a trade-off between model complexity and the interpretability of the optimal regime. The following manuscript takes advantage of this property.

The manuscript presented in this chapter seeks to bridge the literature gap to examine how methods for function optimization can be leveraged in the context of identifying optimal DTRs, in particular using Gaussian process emulation. The original contributions in this manuscript include i) the classification of the sources of variability that can arise in the estimated value function, particularly as it pertains to the IPW estimator, ii) the characterization of the value function maximization problem in order to understand the emulation techniques amenable to the problem, iii) the evaluation of Gaussian process methods that can be used to identify optimal DTRs, including those that can account for homoskedastic and heteroskedastic variability, iv) the creation of data-generating mechanisms for multi-dimensional, multi-modal value functions for sequential decision problems, v) the evaluation of grid-search methods for identifying optimal regimes, and vii) the exposition of these methodologies using trial data on HIV therapeutic agents to identify an individualized therapy recommendation, including examining how sampling uncertainty can be incorporated into the inferential procedure.

Note that the notation in this chapter changes slightly from previous chapters. In manuscript I, the family of regimes was indexed by r or θ; in this manuscript, the indexing variable changes to ψ, as θ parameterizes covariance matrices to be introduced. Additionally, the number of decision points is now T, as opposed to K, as K will now denote a covariance matrix.

Estimation of Optimal Dynamic Treatment Regimes using Gaussian Process Emulation

Daniel Rodriguez Duque[1], David A. Stephens[2], Erica E.M. Moodie[1].

[1]*Department of Epidemiology, Biostatistics, and Occupational Health, McGill University*

[2]*Department of Mathematics and Statistics, McGill University*

Abstract

In precision medicine, identifying optimal sequences of decision rules, termed dynamic treatment regimes (DTRs), is an important undertaking. One approach investigators may take to infer about optimal DTRs is via Bayesian dynamic Marginal Structural Models (MSMs). These models represent the expected outcome under adherence to a DTR for DTRs in a family indexed by a parameter ψ; the function mapping regimes in the family to the expected outcome under adherence to a DTR is known as the value function. Models that allow for the straightforward identification of an optimal DTR may lead to biased estimates and therefore to sub-optimal treatment recommendations. If such a model is computationally tractable, common wisdom says that a grid-search for the optimal DTR may obviate this difficulty. In a Bayesian context, computational difficulties may be compounded if a posterior mean must be calculated at each grid point. We seek to alleviate these inferential challenges by implementing Gaussian Process ($\mathcal{GP}$) optimization methods for estimators for the causal effect of adherence to a specified DTR. We examine how to identify optimal DTRs in settings where the value function is multi-modal, which are often not addressed in the DTR literature. We conclude that a $\mathcal{GP}$ modeling approach that acknowledges noise in the estimated response surface leads to improved results. Additionally, we find that a grid-search may not always yield a robust solution and that it is often less efficient than a $\mathcal{GP}$ approach. We illustrate the use of the proposed methods by analyzing a clinical dataset with the aim of quantifying the effect of different patterns of HIV therapy.

4.1 Introduction

In health research, as data capture and storage capacities improve, the questions researchers ask are becoming more complex. Ambitious questions may be posed in the quest for precision medicine where investigators seek to tailor treatment to patient-specific characteristics through stages of the clinical decision-making process. This tailoring requires sets of decision rules, termed dynamic treatment regimes (DTRs), that take patient information as inputs and that output a treatment recommendation at each stage of the treatment decision-making process. Often, researchers are interested in asking causal questions in relation to these DTRs. Most directly, such questions focus on quantifying what is the causal effect of adherence to a specific DTR and identifying what might the optimal DTR be. The search for an optimal therapy is an important one in medicine, as it aims to avoid over-treatment, all while providing sufficient care to arrive at the targeted outcome. Answering questions about DTRs is challenging, even in data-rich environments; more data may imply that we can ask more challenging questions, but the curse of dimensionality tells us that we cannot altogether escape thinking about statistical models. In this work, we examine how Gaussian Processes ($\mathcal{GP}$s) may yield a strategy that allows for the identification of optimal DTRs.

In the frequentist setting, inferential methods for DTRs have been traditionally performed via semi-parametric models. These include dynamic marginal structural models (MSMs) [Orellana et al., 2010a], g-estimation of structurally nested mean models [Robins, 1986], Q-learning [Murphy, 2005b] and outcome weighted learning [Zhao et al., 2012]. For Bayesians, where modeling the entire probabilistic dynamics is often required for inference, a variety of methods for DTRs have also been proposed, including those of Arjas and Saarela [2010], Saarela et al. [2015b], Xu et al. [2016], Murray et al. [2018], and Rodriguez Duque et al. [2022b]. Although much of the Bayesian literature in this area has focused on adapting existing frequentist estimation approaches for DTRs, the computationally intensive nature of Bayesian inference limits their usability. In this work, we focus on eliminating some of the modeling challenges with DTRs in order improve the usability of methods, be they frequentist or Bayesian.

Our work is motivated by Dynamic MSMs, where a Bayesian version was recently proposed [Rodriguez Duque et al., 2022b]. These allow for the estimation of the value function of DTRs in a family $\mathcal{G}$ indexed by a, possibly multi-dimensional, parameter $\psi \in \mathcal{I}$. In a family of DTRs of the form "treat when covariate value x exceed a threshold ψ", researchers may posit a marginal mean model such as $E[Y^{g^{\psi}}] = \beta_0 + \beta_1\psi + \beta_2\psi^2$. Unfortunately, we cannot be certain that this model is correctly specified or that it is sufficiently flexible to correctly identify the optimal regime. One way around this issue is to estimate the expected outcome under each regime in the family, if there is a finite number of them, or to estimate the expected outcome of a large set of regimes and then extrapolate the value to other regimes. This essentially amounts to a grid-search and is an appealing approach as we have access to standard estimators for the expected outcome under adherence to a DTR. Unfortunately, this may be computationally intensive, particularly in settings with many stages, complex decision rules, and a variety of confounders. Computational challenge may be compounded in Bayesian settings where sampling of a posterior distribution is often required. Even if a grid-search is feasible, it has not been established in the literature whether it reliably identifies the optimal regime or whether there are other robust approaches that use data more efficiently.

The contribution of this work is to examine how to utilize computer experiments to identify optimal DTRs in a family $\mathcal{G}$; an optimal DTR is one that maximizes the value function. In a DTR context, a "computer experiment" should sample the value function at strategically chosen points with the aim of approximating the entire value function, all while limiting the number of samples obtained. These experiments should begin by selecting an initial set of design points $\mathcal{I}$, and then using an estimator for the value of a DTR at these points to arrive at a working model for the value function. This working model can then be utilized to select new points sequentially using a criterion, known as an infill criterion (or acquisition function), that specifies where an optimum may be. We focus on the Expected Improvement criterion Jones et al. [1998] which has been well studied and is known to balance exploration of the input space with exploitation of the optimizing region. Using this approach, we focus on methods that yield models more flexible than those used with Dynamic MSMs and that allow for the sequential sampling of additional points in order to improve the estimation of optimal DTRs. The models used are obtained via a $\mathcal{GP}$ prior, with parameters fit using empirical Bayes or maximum a posteriori (MAP) inference. This is a novel approach to identifying optimal DTRs that has not previously been explored in the precision medicine literature. The computer experiment faces an additional challenge in that we do not have access to the value function, but rather to an estimator for the value of a DTR which can be evaluated point-wise. Via simulations, we find that a $\mathcal{GP}$ modeling approach that acknowledges uncertainty in the estimated regime values can successfully identify optimal DTRs. Additionally, we find that a grid-search for the optimum may not always be the best solution, especially in multi-modal settings which challenges the received wisdom that a grid-search is as reliable as other methods. We find that computer experiments via $\mathcal{GP}$s can perform better than a grid-search, all while using fewer experimental points. In addition to these contributions, we illustrate how to use the discussed methods to perform a statistical analysis on clinical data arising from HIV therapy.

4.2 Background

Traditionally, computer experiments to identify an optimum were performed using regression-based methods fit on a set of experimental points. However, these methods may not be well suited for identifying optimal responses. Huang et al. [2006] mention that regression-based approaches may be inefficient as they attempt to predict the response curve over the entire feasible domain, as opposed to the neighborhood of the optimum. Additionally, the linear regression models used are usually relatively simple, and may not fit complex systems adequately over the entire domain similar to the issues we described with Dynamic MSMs. Consequently, more recent literature on computer experiments focuses on approaches using $\mathcal{GP}$s and sequentially sampling new experimental points most relevant to the optimization.

A $\mathcal{GP}$ is a stochastic process for which all outcome vectors, regardless of the dimension, have a multivariate Normal distribution. Models arising from the $\mathcal{GP}$ assumption are often termed kriging models. In a computer experiments context, these models are widely used in two settings: 1) where researchers would like to fit a flexible model, which may be used for prediction in unobserved locations; or 2) where researchers are working with a function that is expensive to evaluate and would like to identify the optimum of this function, all while limiting the number of function evaluations. The latter is also termed Bayesian optimization [Pourmohamad and Lee, 2021] and is most relevant to our work. In addition to the $\mathcal{GP}$ assumption, much of this literature focuses on settings where the input-output relationship is known. Our setting is nuanced as we only have access to an estimated (noisy), yet deterministic, output for any given input. Consequently, we must think carefully about the problem characteristics before developing an optimization strategy.

Sacks et al. [1989] were among the first researchers to explore using $\mathcal{GP}$s for computer experiments. Later, Currin et al. [1991] used a similar methodology but in a Bayesian context. O'Hagan et al. [1999] argued that a Bayesian perspective is crucial for computer experiments with deterministic functions as, for a fixed input, the output does not change. Consequently, uncertainty about the response surface is not aleatory. Notwithstanding, a fully Bayesian treatment of this problem is often highly complex as it requires Markov Chain Monte Carlo to sample from posterior distributions, and to evaluate the infill criterion, comprised of a posterior expectation, at each point in the optimization procedure. Thus, some compromises must be made. We take this into consideration and seek practical methods that balance the benefits of the Bayesian and frequentist approach. One concern may be, as with any optimization procedure, that there exist local maxima within the operating domain of interest, making the identification of a global maximum more challenging. Jones et al. [1998] emphasize that a computer experiments methodology based on $\mathcal{GP}$s is good for modeling non-linear multi-modal functions. In addition to the $\mathcal{GP}$ model, which requires specification of a covariance function an infill criterion must be specified. There are a variety of infill criteria in the literature; we make use of the Expected Improvement criterion although some care should be taken as it encounters theoretical problems in settings with noisy outputs. We will examine these issues in what follows.

4.3 Problem Characteristics

Before characterizing our specific inferential problem, let us fix some terminology regarding the surfaces of interest. The *value surface* refers to the true relationship between a DTR, g^{ψ}, and its value $E[Y^{g^{\psi}}]$; the *estimation surface* refers to the surface obtained by point-wise evaluation of an estimator $\hat{E}[Y^{g^{\psi}}]$ for varying $\psi \in \mathcal{I}$; note that $\mathcal{I}$ is an index set that can be continuous or discrete. When we make use of an inverse probability weighted (IPW) estimator to obtain this surface, we refer to this estimation surface as the *IPW-surface*. Lastly, the *emulation surface* is the posterior mean of a given $\mathcal{GP}$ of interest that is meant to approximate the *value surface*.

Our setting is unique in that we are looking to emulate the value surface by only observing values from the estimation surface. As the estimation surface is produced by evaluating an estimator point-wise in a relevant domain, this function exhibits a non-smooth quality, and we are in a setting where the observed output is a noisy version of the true output. We will see that this non-smoothness may affect the results obtained via a grid-search. As investigators, we are likely interested in smoothing out this noisy surface, believing the true value function to be smooth; a $\mathcal{GP}$-based model allows for this possibility.

The lack of smoothness of the estimated surface is mainly a consequence of using a finite sample size to estimate the value of regimes in $\mathcal{G}$. Furthermore, we may ask whether adequately capturing this noise structure improves the resulting inference and whether this noise structure is homoskedastic or heteroskedastic. There are several components in the data analysis that may lead to a heteroskedastic structure. These include 1) measurement error, possibly including more variability in treatment arm than in the response arm; 2) relatively smaller sample-size in some regions of the regime index set than others; and 3) patient responses being more distant from the value function in some areas of the index set than others. We more precisely illustrate these considerations in the following sections and in Appendix B.1.

Kriging methods can also be used in settings with noisy observations; Picheny and Ginsbourger [2014] provide an overview of these methods in an optimization setting. Indeed, stochastic kriging is nuanced and not all methods are applicable in all settings. Stochastic kriging is often utilized when emulating a response surface where at each experimental point the output varies when re-evaluated at the same input. In settings that do not involve sequential sampling of experimental points this definition is sufficient, as a model is fit on a fixed and known set of points. However, when sequential sampling is required, more care should be taken in defining the problem. There are some settings where we observe a noisy function but where there is no uncertainty in the output when re-evaluating at already sampled points. Forrester et al. [2006] explain that in this setting there is no uncertainty in the output, even if there is noise around the true curve. In other settings, re-evaluating at the same input yields varying outputs. This detail is consequential when identifying infill criteria for stochastic kriging. In some cases, we gain information by re-sampling at the same data-point — in others we do not. Our motivating DTR setting relates most to the case where a curve exhibits a characteristic jitter but where there is no uncertainty in the output of already sampled points.

Stochastic kriging has focused on methods with homoskedastic noise; however there is a growing literature on incorporating heteroskedastic noise in the inferential procedure. For example, Ankenman et al. [2008] and Yin et al. [2011] incorporate heteroskedastic noise by estimating the noise variance at design points; these authors' approach requires that the function of interest be evaluated at the design points multiple times. Frazier et al. [2011] also discuss heteroskedastic error and propose a method for financial time series. A fully Bayesian approach is presented by Goldberg et al. [1997] who seek to place a $\mathcal{GP}$ prior on the log noise, yielding two $\mathcal{GP}$ priors. Indeed a fully Bayesian treatment is computationally intensive, but some work has been done on alleviating these issues; Wang [2014] has looked at fast MCMC procedures for $\mathcal{GP}$s with heteroskedastic noise. Thinking about practicality, Kersting et al. [2007] follow the same approach as Goldberg et al. [1997], however they focus on most likely heteroskedastic $\mathcal{GP}$s to estimate the input-dependent noise level. Zhang and Ni [2020] offer an improvement on most likely heteroskedastic $\mathcal{GP}$s by providing an approximately unbiased estimator for the input-dependent noise. In what follows we will examine the performance of the latter approach.

From the above considerations, we regard a $\mathcal{GP}$ model that acknowledges noise as possessing an important characteristic; it remains to examine what criteria may be used to sample points sequentially. Picheny et al. [2013] provide a review of infill criteria used for stochastic kriging. Frazier and Wang [2016] emphasize that the Expected Improvement criterion benefits from some optimality results in the deterministic setting but that these benefits are lost in noisy stochastic settings. In particular, in deterministic settings, the Expected Improvement criterion ensures the true optimum will be identified as the number of experimental points increases. This result hinges on the posterior variance at already sampled points being zero [Locatelli, 1997], but this property is not necessarily present in stochastic settings. Many infill criteria allow for re-evaluations at already sampled points, but this is not desirable in our setting. There are other technical issues in revisiting experimental points with the $\mathcal{GP}$, for example ill-conditioned matrices. Forrester et al. [2006] propose a solution for using the Expected Improvement in noisy settings by utilizing a re-interpolation approach for optimization. This is the approach that we explore.

4.4 Methods

We consider a sequential decision problem with T decision points and a final outcome y. Decisions taken up to stage t give rise to a sequence of treatments $\bar{z}_t = (z_1, ..., z_t)$, $z_j \in \{0, 1\}$. At each stage t, a set of covariates x_t is available for decision-making and it is assumed that these consist of all time-fixed and time-varying confounders. To denote covariate history up to time t, we write $\bar{x}_t = \{x_1, ..., x_t\}$. Subscripts are omitted when referencing history through stage T. Then, all patient information is given by $b = (\bar{x}, \bar{z}, y)$. We denote a DTR-enforced treatment history by $g(\bar{x}) = \bar{g}(\bar{x}) = (g_1(x_1), ..., g_T(\bar{x}_T))$. Throughout, we will consider a family of DTRs, indexed by $\psi \in \mathcal{I}$ to give $\mathcal{G} = \{g^{\psi}(\bar{x}); \psi \in \mathcal{I}\}$. In general, we allow ψ to be a p-dimensional column vector. The index is omitted when it is clear that our focus lies on a single DTR. Based on these definitions, we posit that values υ_i on the estimation surface are a noisy realization of the value surface $f(\psi)$ as given by the following relationship:

$$\upsilon_i = f(\psi_i) + \epsilon_i \ , \ \epsilon_i \sim N(0, \gamma^2(\psi_i)), \ i = 1, ..., m. \tag{4.1}$$

Our target of inference is the value surface f for which there is epistemic uncertainty. As equation 5.10 makes clear, this problem is further complicated as we do not observe f, but instead mereley a noisy version of it. To fix the notation about this model, suppose we have data $\mathcal{D} = \{\psi_i, \upsilon_i\}_{i=1}^{m}$. Then define the following vector quantities $\psi = (\psi_1, ..., \psi_m)^T$, $\upsilon = (\upsilon_1, ..., \upsilon_m)$ and $f = (f_1, ..., f_m)$. We also define $\bar{\gamma}^2 = (\gamma^2(\psi_1), ..., \gamma^2(\psi_m))$. Note that these are observations taken on the estimation surface. We have control of the observations that we sample from this surface, and these contrast the observations on the sample $(\bar{x}, \bar{z}, y)$ which are fixed at a sample size n.

To perform inference, we place a prior on f, which represents our belief about the value function associated with a family of DTRs indexed by $\psi \in \mathcal{I}$. We choose this to be a prior $d\pi(f)$ in a function space $f \in \mathcal{F}$. Heuristically, as in Shi and Choi [2011], updating can be done via the equation:

$$P(f \in A|\mathcal{D}, \gamma^2) = \int_A \frac{p(\upsilon|f, \gamma^2)d\pi(f)}{\int_{\mathcal{F}} p(\upsilon|f, \gamma^2)d\pi(f)}, \quad A \subset \mathcal{F}. \tag{4.2}$$

More concretely, the prior that we make use of is a $\mathcal{GP}$ prior, which has the consequence that for any finite set of observations ψ, $f|\psi \sim N(\mu_{0f}, K)$. K is a covariance matrix calculated via a covariance function $k(\psi_i, \psi_j)$ that is parameterized by parameters (θ_f, σ_f^2), with θ_f being a vector with entries θ_{fd} controlling the correlation between points in the dth dimension and σ_f^2 being a parameter that scales the correlation function to yield the covariance function. The $\mathcal{GP}$ requires specification of a set of hyperparameters $\eta_f = (\mu_{0f}, \theta_f, \sigma_f^2)$. Without further knowledge of the problem, it is challenging to specify values for these hyperparameters. Specifying priors for these hyperparameters is possible, but it may increase computational challenges to carry out a fully Bayesian treatment of this problem. More commonly, empirical Bayes is used to estimate the hyperameters via maximum likelihood, as in Shi and Choi [2011]. Alternatively, MAP estimation of the hyperparameters may be used. Conditional on fixing these hyperparameters, at their MAP or empirical Bayes estimates, standard arguments for the conditional distribution of a multivariate normal distribution yield the posterior distribution at a new point ψ_{m+1} to be:

$$\begin{aligned} f_{m+1}|\psi_{m+1}, \eta_f, \bar{\gamma}^2, \mathcal{D} &\sim N(\mu_{f_{m+1}}, \sigma^2_{f_{m+1}}) \\ \mu_{f_{m+1}} &= \mu_{0f} + k_{m+1}^T (K + S)^{-1} (\upsilon - \mu_{0f}) \\ \sigma^2_{f_{m+1}} &= k(\psi_{m+1}, \psi_{m+1}) - k_{m+1}^T (K + S)^{-1} k_{m+1}, \end{aligned} \tag{4.3}$$

where S is a diagonal matrix of noise variances with iith entry equal to $\gamma_i^2 = \gamma^2(\psi_i)$; $k_{m+1} = (k(\psi_1, \psi_{m+1}), ..., k(\psi_m, \psi_{m+1}))$ is the variance vector between already sampled points and the new point ψ_{m+1}. In the empirical Bayes setting, the covariance parameters are fixed values. Consequently, they need not be included in the conditioning, we do this however for compatibility with the MAP approach. Note that unlike the more well known $\mathcal{GP}$ model for computer experiments, this model does not necessarily interpolate the observed data. That is, $\mu_{f_{m+1}}$ does not necessarily perfectly predict the observed data points. This is desirable, as we seek a smooth response curve, but we only have access to the noisy estimation surface. To recover the interpolating model, we set $\gamma^2(\psi_i) = 0 \; \forall i$. As Forrester et al. [2006] point out, the interpolation property of a $\mathcal{GP}$ occurs when there is no measurement error in the data observation mechanism and comes from noting that the posterior variance is zero at already sampled points. In what follows, we will more closely examine non-interpolating scenarios. The remaining quantity of interest is the posterior distribution for the noisy observations:

$$\begin{aligned} \upsilon_{m+1}|\psi_{m+1}, \eta_f, \bar{\gamma}^2, \gamma^2_{m+1}, \mathcal{D} &\sim N(\mu_{\upsilon_{m+1}}, \sigma^2_{\upsilon_{m+1}}) \\ \mu_{\upsilon_{m+1}} &= \mu_{f_{m+1}} \\ \sigma^2_{\upsilon_{m+1}} &= k(\psi_{m+1}, \psi_{m+1}) - k^T_{m+1}(K+S)^{-1}k_{m+1} + \gamma^2_{m+1}. \end{aligned} \tag{4.4}$$

4.4.1 Homoskedastic Inference

If noise is a concern, an interpolating $\mathcal{GP}$ approach may not be adequate, and we may look to allow for noise around the surface. If we assume that the noise variance is homoskedastic, then we have that $\gamma^2(\psi_i) = \gamma^2 \; \forall i$. Under an empirical Bayes approach our posterior of interest is $p(\upsilon_{m+1}|\psi_{m+1}, \mathcal{D}) = p(\upsilon_{m+1}|\psi_{m+1}, \eta_f, \gamma^2, \mathcal{D})$. To compute values for the hyperparameters, we maximize $p(\upsilon|\psi, \eta_f, \gamma^2)$. Efficient computational approaches to identifying the maximizers of this marginal likelihood can be found in Park and Baek [2001] and Roustant et al. [2012]. With access to this model, we could additionally combine it with MAP estimation of θ_f in order to arrive at an approximation for $p(\upsilon_{m+1}|\psi_{m+1}, \mathcal{D})$. This requires maximizing $p(\eta_f, \gamma^2|\mathcal{D})$ with respect to η_f, γ^2 in order to obtain $\eta_f^{map}, \gamma^{2,map}$. MAP estimation then uses the approximation $p(\eta_f, \gamma^2|\mathcal{D}) \approx \mathbb{1}_{(\eta_f^{map}, \gamma^{2,map})}(\eta_f, \gamma^2) d(\eta_f, \gamma^2)$ in order to arrive at the posterior predictive distribution as:

$$p(\upsilon_{m+1}|\psi_{m+1}, \mathcal{D}) \approx \int p(\upsilon_{m+1}|\psi_{m+1}, \eta_f, \mathcal{D}) \mathbb{1}_{(\eta_f^{map}, \gamma^{2,map})}(\eta_f, \gamma^2) d(\eta_f, \gamma^2) = p(\upsilon_{n+1}|\psi_{m+1}, \eta_f^{map}, \gamma^{2,map}, \mathcal{D}).$$

Lizotte [2008] has examined MAP inference for deterministic computer experiments under a Log-Normal prior for θ_f; we also examine the consequences of this prior on MAP inference.

4.4.2 Heteroskedastic Inference

Alternatively, we may believe that the response surface exhibits heteroskedastic noise. This poses special challenges as it requires performing inference for each of the noise variances, γ_i, in the observed data. For this, we examine an approach proposed by Zhang and Ni [2020] that places a second $\mathcal{GP}$ prior on the regression residuals $e_i = |r_i|^q = |\upsilon_i - \mu_{\upsilon_i}|^q$, $q \in \mathcal{Z}^+$, with covariance function $k_e(\psi_i, \psi_j)$ and parameters $\eta_e = (\mu_{0e}, \eta_e, \sigma_e^2)$. Authors show that under these assumptions a method of moments estimator for the input-specific noise variances can be arrived at via:

$$E[|r_i|^q] = \frac{\gamma_i^q}{s(q)}, \tag{4.5}$$

where $s(q)$ is a correction factor. When $q = 1$, $s(1) = \sqrt{\pi/2}$, and the estimator for the input-dependent noise is approximately $\tilde{\gamma}_i = \sqrt{\pi/2}E[|r_i|] = \sqrt{\pi/2}\mu_{e_i}$, where μ_{e_i} is the posterior mean of the second $\mathcal{GP}$. A fully Bayesian computation that acknowledges uncertainty in γ_i would require an integral like:

$$p(\upsilon_{m+1}|\psi_{m+1}, \eta_f, \mathcal{D}) = \int\int p_1(\upsilon_{m+1}|\psi_{n+1}, \eta_f, \bar{\gamma}^2, \gamma_{m+1}^2, \eta_e, \mathcal{D}) p_2(\bar{\gamma}^2, \gamma_{m+1}^2|\psi_{n+1}, \mathcal{D}) d\bar{\gamma} d\gamma_{m+1}. \tag{4.6}$$

For known $\bar{\gamma}$, γ_{m+1}, sampling from p_1 is Normal with posterior mean and variance as described in equation (5.12). However, this computation is challenging because the γ^2 are unobserved. Goldberg et al. [1997] provide an MCMC approach to allow for sampling from p_2 which computes the integral of interest, however this is computationally intensive. Kersting et al. [2007] proposed that p_2 be approximated by the most likely noise level. The most likely noise level is calculated as the posterior mean of a $\mathcal{GP}$ that has been placed on $\log(\gamma_i)$; recall that at each function value, the $\mathcal{GP}$ is Normally distributed, therefore making the most likely value the $\mathcal{GP}$ mean. Zhang and Ni [2020] provide an improved way to estimate γ_i, as described in equation 4.5, in order to yield the approximation $\upsilon_{m+1}|\psi_{m+1}, \eta_f, \bar{\tilde{\gamma}}^2, \tilde{\gamma}^2_{m+1}, \mathcal{D} \sim N(\mu_{\upsilon_{m+1}}, \sigma^2_{\nu_{m+1}})$. As in the empirical Bayes approach, $\tilde{\gamma}^2_i$ are assumed known in the computation. Consequently, we can treat this posterior distribution as a $\mathcal{GP}$ and perform inference as before.

In the following, we examine how to pair the homoskedastic and heteroskedastic models with the expected improvement criterion in order to arrive at a sequential sampling scheme.

4.4.3 Infill Criterion

We return to the question of an appropriate infill criterion when we are interested in performing minimization. The Expected Improvement in our setting is given by: $EI(\psi) = E\left[\max(0, \upsilon(\psi) - \upsilon_{max})|\mathcal{D}\right]$. The expectation is taken with respect to the posterior distribution and $\upsilon_{max} = max(\upsilon_1, ..., \upsilon_m)$. Further computation yields:

$$EI(\psi) = (\mu_{\upsilon_{m+1}}(\psi) - \upsilon_{max})\Phi\left(\frac{\mu_{\upsilon_{m+1}}(\psi) - \upsilon_{max}}{\sigma_{\upsilon_{m+1}}(\psi)}\right) + \sigma_{\upsilon_{m+1}}(\psi)\phi\left(\frac{\mu_{\upsilon_{m+1}}(\psi) - \upsilon_{max}}{\sigma_{\upsilon_{m+1}}(\psi)}\right)$$

when $\sigma_{\upsilon_{m+1}}(\psi) > 0$ and 0 otherwise. Φ is the CDF of the Standard Normal distribution and ϕ is the corresponding pdf.

4.4.4 Re-interpolation

As discussed, using the Expected Improvement as a criterion for sequential sampling may not be theoretically justified in a deterministic computer experiment with noisy observations, in particular when a regressive model is used rather than an interpolating model. Regressive models are ones that do not interpolate the sample data, like the homoskedastic and heteroskedastic models discussed above. The challenge in using the Expected Improvement with these models arises from the fact that the error $\sigma_{\upsilon_{m+1}}(\psi)$ at sample points will be non-zero even though the output will not vary when the estimation function is re-evaluated at these sample points. Consequently, convergence toward global optimum cannot be guaranteed [Locatelli, 1997]. Forrester et al. [2006] introduce a re-interpolation method that attains zero error at the sample locations. This can be done by building an interpolating $\mathcal{GP}$ on the values predicted by the regressive model mean $\mu_{\upsilon_{m+1}}$ and sequentially sampling using the Expected Improvement based on this model. The procedure is termed re-interpolation because the interpolating model is built on the predicted mean values of the regressive model.

First, the re-interpolating procedure uses predictions at sample points obtained from the mean of $\upsilon_{m+1}|\psi_{m+1}, \mathcal{D}$ in order to create a new dataset $\mathcal{D}'$. At sample point i, we define the predicted values as $\hat{\upsilon}_i = \mu_{\upsilon_{m+1}}(\psi_i)$ to yield responses $(\hat{\upsilon}_1, ..., \hat{\upsilon}_m)$ and new data $\mathcal{D}' = \{\psi_i, \hat{\upsilon}_i\}_{i=1}^m$. Then using an interpolating $\mathcal{GP}$ assumption on these data, we obtain a similar heuristic as before: $p(\hat{\upsilon}_{m+1}|\psi_{m+1}, \mathcal{D}') = p(\hat{\upsilon}_{m+1}|\psi_{m+1}, \eta_{\hat{f}}, \mathcal{D}')$, where the posterior mean and variance are given by:

$$\mu_{\hat{\upsilon}_{m+1}} = \mu_{0\hat{\upsilon}} + k_{m+1}^T K^{-1}(\hat{\upsilon} - \mu_{0\hat{\upsilon}})$$
$$\sigma^2_{\hat{\upsilon}_{m+1}} = k(\psi_{m+1}, \psi_{m+1}) - k_{m+1}^T K^{-1} k_{m+1},$$

with $\mu_{0\hat{\upsilon}}$ being the prior mean of the interpolating process. This re-interpolating procedure leads to two essential properties: 1) the posterior mean of the υ and $\hat{\upsilon}$ processes are the same i.e. $\mu_{\upsilon_{m+1}} = \mu_{\hat{\upsilon}_{m+1}}$, and 2) the variance of the $\hat{\upsilon}$ process is zero at already sampled points. The latter is the crucial characteristic required to preserve the optimality of the Expected Improvement criterion. With this re-interpolating model, the Expected Improvement can be calculated to determine new sampling locations. In Appendix B.2, we provide additional details on the equality of the two posterior means. Forrester et al. [2006] mention that the covariance function K remains unchanged, so η_f does not need to be re-optimized.

4.4.5 Design of Experiments

One component of the design of experiments is to determine the initial number of design points. Loeppky et al. [2009] investigate this issue and conclude that ten points per dimension is a reasonable rule-of-thumb when the dimension is less than five. We simply select them in equally spaced increments. Another option, for example, is to select design points randomly, but given the nature of our experiment, we aim to eliminate variability due to the initial sampling strategy. Strategies to select design points include simple random sampling, or Latin Hypercube sampling [McKay et al., 1979]. Designs based on distance measures are also possible, like those that seek to maximize the minimal distance between experimental points [Johnson et al., 1990].

Another design element that must be considered is the covariance function. Some covariance functions in the $\mathcal{GP}$ lead to smoother surfaces than others. One common choice of covariance is the $Matern$ covariance family. Common choices in this family are the $Matern_{5/2}$ covariance which is twice differential and the $Matern_{3/2}$ covariance which is differentiable once. These are examples of isotropic covariance functions, meaning that the correlation between points depends only on the distance between them. We focus on the $Matern_{5/2}$ covariance given by:

$$k(\psi_i, \psi_j) = \sigma_f^2 \prod_{d=1}^{D} \left(1 + \frac{\sqrt{5}|\psi_{id} - \psi_{jd}|}{\theta_{fd}} + \frac{5(\psi_{id} - \psi_{jd})^2}{3\theta_{fd}^2}\right) \exp\left(\frac{-\sqrt{5}|\psi_{id} - \psi_{jd}|}{\theta_{fd}}\right), \qquad (4.7)$$

where D is the number of dimensions in the index vector.

4.4.6 Estimation Surface

As previously mentioned, the estimation surface can be produced with any estimator for the value of a DTR. In this work, we make use of the normalized IPW estimator. Then, for a family of interest, the estimator can be evaluated on a grid of ψs in order to yield the resulting estimation surface. The normalized IPW estimator is given by

$$\frac{\sum_i w_i^{\psi} y_i}{\sum_i w_i^{\psi}}, \text{where } w_i^{\psi} = \frac{\mathbb{1}_{\bar{g}^{\psi}(\bar{x}_i)}(\bar{z}_i)}{\prod_{j=1}^{T} p(z_j|\bar{z}_{j-1}, \bar{x}_j)}. \tag{4.8}$$

An additional layer of complexity is encountered if we are interested in using a Bayesian estimator to perform inference. This is because computing a posterior mean often requires sampling from the posterior distribution, which may be a computationally intensive task. In this case, a grid-search for the optimal DTR may become intractable. Rodriguez Duque et al. [2022b] provide a Bayesian estimator for the value of a DTR by making use of inverse weighting and the Bayesian bootstrap. Often, Markov Chain Monte Carlo is required to estimate posterior means; although there may be some Monte Carlo variability in the estimated mean, if we fix a seed when sampling the posterior distribution, then we can arrive at an estimation procedure that is still deterministic.

Generally, bootstrapping can allow for the quantification of sampling uncertainty. For example, in our setting, it may be that we are interested in quantifying uncertainty around the estimation surface. For a set of observations $(b_1, ..., b_n)$, the bootstrap procedure samples each observations independently with replacement and with equal probability $1/n$ in order to estimate the quantity of interest. A similar procedure can be arrived at through a Bayesian lens as first proposed by Rubin [1981]. This requires a posterior distribution that places a random probability π_i of sampling observation b_i in a bootstrapped sample; these probabilities have mean $1/n$ thereby connecting the procedure to frequentist bootstrap. This Bayesian bootstrap procedure can be arrived at by placing Dirichlet Process $\mathcal{DP}(\alpha, G)$ prior on the data-generating distribution, where α is a concentration parameter and where G is a base distribution. In particular, when α is chosen such that $|\alpha| \to 0$, we obtain the Bayesian bootstrap as the posterior predictive distribution. Under this specification, one sample drawn from the posterior $\mathcal{DP}$ is given by $p(b_{n+1}|\bar{b}, \pi) = \sum_{i=1}^{n} \pi_i \mathbb{1}_{b_i}(b_{n+1})$, where $\pi = (\pi_1, ..., \pi_n) \sim Dir(1, ..., 1)$ is a sample from the Dirichlet distribution with all concentration parameters equal to one. Under these assumptions, any distribution sampled from the posterior $\mathcal{DP}$ is uniquely determined by π. For example the Bayesian bootstrap can be operationalized to quantify posterior belief about the population mean $E[b_{n+1}|\bar{b}] = E_{\pi}[E[b_{n+1}|\bar{b}, \pi]]$ by sampling weights $(\pi_1, ..., \pi_n)$ and computing

$$E[b_{n+1}|\bar{b}, \pi] = \int_{b_{n+1}} b_{n+1} \sum_{i=1}^{n} \pi_i \mathbb{1}_{b_i}(b_{n+1}) db_{n+1} = \sum_{i=1}^{n} \pi_i b_i. \tag{4.9}$$

This quantity can be computed over many draws of the weights in order to obtain the full posterior distribution for the mean. Taking the mean across all these bootstrap samples results in an estimate for $E[b_{n+1}|\bar{b}]$.

Sources of Variation

As we have already discussed, the estimation surface exhibits non-smoothness. In this section, we examine some possible sources of heteroskedastic variation. These considerations are most consequential for finite sample sizes. In this exploration, we limit ourselves to regimes of the form "treat if $x > \psi$", as this is a common regime in the literature, and it leads to clear examples about how heteroskedasticity is manifested. Additionally, we focused on the normalized IPW estimator for the value of a regime which uses only patients observed to adhere to the regime of interest.

Note that in contrast to static treatment regimes, an individual can be simultaneously adherent with many DTRs [Cain et al., 2010]. Furthermore, for a given sample with binary treatment, there are two response curves: the treated curve and the untreated curve. For a fixed ψ, we can use the sample to estimate $E[Y^{g^{\psi}}]$. Furthermore, for an increase in ψ from ψ_1 to ψ_2, only treated patients can become non-adherent and only untreated patients can become adherent because of the form of rule under consideration. Only patients with covariate values $\psi_1 \leq x \leq \psi_2$ are eligible to become adherent/non-adherent. These properties are important in examining the variability in the estimation surface.

The first case we consider is heteroskedasticity due to distance from the value surface. This relates to how close/far the estimated treated and untreated curves are from the value surface. Recall that the value surface represents a population average; individual responses can vary substantially around this surface. For an increase in ψ from ψ_1 to ψ_2, there will be a set of patients who become non-adherent with regime ψ_2 and a set who become adherent. As the IPW estimator uses only observations on those adherent to a regime, if either the newly adherent/non-adherent patients have a response value that is far from the IPW-surface, then these observations will have a considerable influence on the estimate, especially for relatively small sample sizes. If the observations tend to have a response that is near the population average, then the IPW-surface will be less influenced by these observations.

The second case is heteroskedasticity due to the noise structure at the individual level. Consider an additive error term in the data-generating mechanism, such as: $z\epsilon_1 + (1-z)\epsilon_2$, where $\epsilon_1 = N(0, 5)$, $\epsilon_2 = N(0, 0.5)$. We might not think this is an issue, as for estimation via an estimating equation, it does not matter whether noise is heteroskedastic or homoskedastic, so long as it has zero mean. However, when estimating the value surface for the purposes of identifying a maximum, this may be consequential. As ψ increases, we lose treated patients, and we gain untreated patients. This means that we lose observations with high variability and gain observations with low variability; this noise structure at the individual level transforms into heteroskedasticity at the estimator level. Now, we may ask when this data-generating mechanism may arise. One case may be when treatment leads to relatively reliable improvements, but lack of treatment leads to disease progression taking on a variety of forms, and therefore leading to higher variability.

The third consideration that may lead to heteroskedasticity in the estimation surface is the result of differing effective samples sizes across values of ψ. It is well known that the IPW estimator for a regime ψ only uses patients who are adherent to the regime. Consequently, different regimes will use different number of patients to compute the value of the corresponding regime. This means that the estimator will exhibit differing levels of variability for a range of ψs. In Appendix B.1, we further illustrate all three cases discussed.

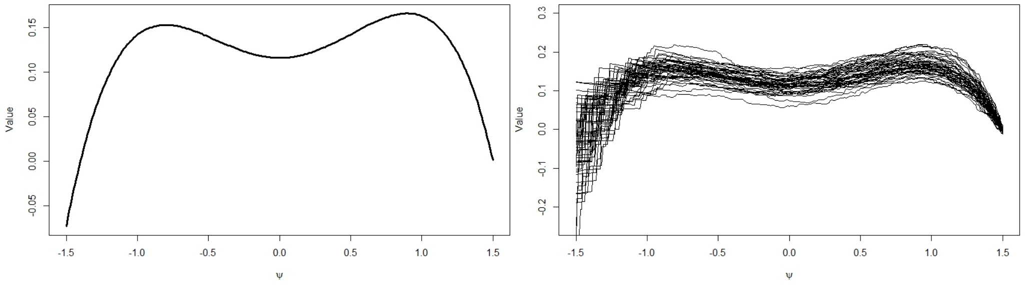
4.5 Simulations

In what follows, we examine several data-generating mechanisms and DTRs to assess whether the $\mathcal{GP}$ approaches presented do allow for the identification of optimal DTRs; we additionally compare these to a grid-search. We refer to the interpolating, homoskedastic, and heteroskedastic $\mathcal{GP}$ models as Int$\mathcal{GP}$, HM$\mathcal{GP}$, and HE$\mathcal{GP}$, respectively. We present results for a sample size of $n = 500$ with a $Matern_{5/2}$ covariance function. To produce the estimation surface, we make use of the normalized IPW estimator. To compare across modeling strategies, each analysis was performed on 500 Monte Carlo replicates. Appendices B.3 through B.5 examine scenarios with a sample size of $n = 1000$, a $Matern_{3/2}$ covariance, and a Log-Normal prior.

4.5.1 Simulation I

For this simulation, we generate covariate $x \sim U(-1.5, 1.5)$, treatment $z \sim Binom(p = expit(2x))$, error distributions $\epsilon_1 = N(0, \sigma = 0.25)$, $\epsilon_2 = N(0, \sigma = 0.05)$, and final outcome $y = -(x + .8)x(x - .9)z + z\epsilon_1 + (1 - z)\epsilon_2$. We explore the regime "treat if $x > \psi$", $\psi \in (-1.5, 1.5)$. Note that $expit(\cdot)$ refers to the inverse $logit$ function. With this data-generating mechanism, the systematic component of y varies from -2 to 2.5 and the optimal regime represents a 5 % improvement (in the range of y) over the worst regime in the class. In Figure 4.1, we observe the value function for this problem and the IPW-surface, across multiple replicates. It is visually evident that the function has two local maxima but only one global maximum at $\psi = 0.9$. There appears to be more variability for low values of ψ than for high values. Contrary to standard practice, a grid-search for the optimum may not work well, as evidenced by the large interquartile range (IQR) in Table 4.1.

For this simulation, the computer experiment was designed such that we sampled an initial set of design points in increments of 0.25, yielding an initial set of 13 points. Then, additional points were sampled sequentially using the Expected Improvement criterion, up to 25 additional points. All measures of variation correspond to Monte Carlo variation across replicates. We do not compute coverage probabilities for each $\mathcal{GP}$, as for a fixed replicate, the uncertainty represented by the $\mathcal{GP}$ is constrained to uncertainty in the IPW-surface resulting for a specific sample of size n; it does not incorporate sampling uncertainty. Incorporating sampling uncertainty requires a more computationally intensive procedure, one that we explore in the case study.



(a)

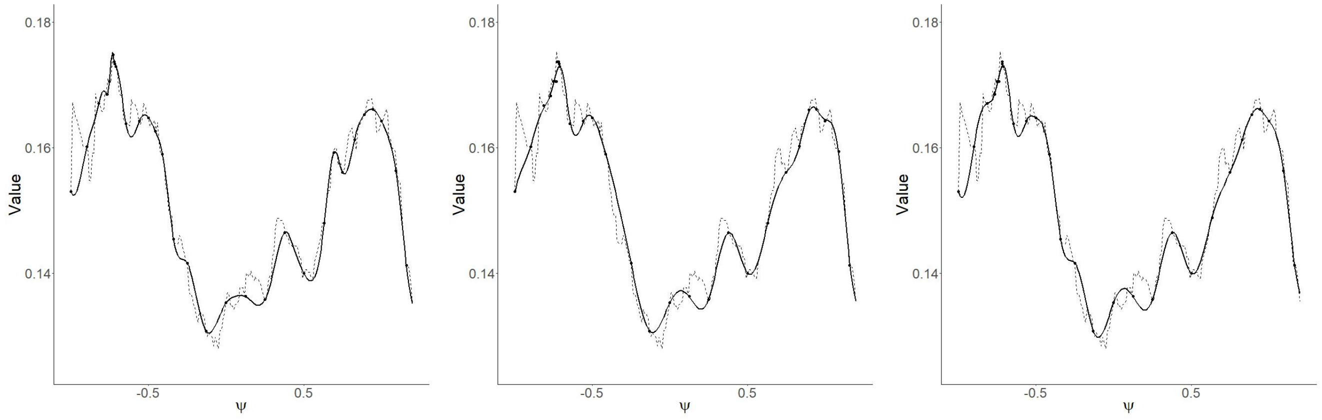
(b)

Figure 4.1: Simulation I: (a) Value function (b) IPW estimates of value function, across 50 replicates.

Table 4.1: Results for grid-search with increments of 0.01 and $n = 500$. True ψ_{opt}=0.9; true value at optimum 0.165.

Statistic	$\hat{\psi}_{opt}$	Value at $\hat{\psi}_{opt}$
Mean (SD)	0.427 (0.800)	0.172 (0.022)
Median (IQR)	0.860 (1.600)	0.171 (0.029)

From Figure 4.2, we see the results of the three modeling strategies for one replicate. These curves represent the posterior mean after sampling 25 additional points using the Expected Improvement as the infill criterion. In the figure, we restrict the domain of ψ for better visualization around the local and global optima, but in Appendix B.3 the curves can be visualized for the entire decision space. From the figure, we see why the interpolating model is likely to under-perform; occasionally, due to noise in the fit, there will be a maximizer of the IPW-surface that is not close to 0.9. In these scenarios, the Int$\mathcal{GP}$ will interpolate the data, whereas the other two methods can adjust the estimate based on the identified noise level. Careful examination of the graphs reveals that the interpolation is most consequential around the local optimizer $\psi = 0.8$. Although HE$\mathcal{GP}$ may assign higher variability to certain regions, it may also assign lower variability and become closer to interpolating. These plots contrast the differences between these methods, but they do not inform us about what will happen across many analyses. Consequently, we now look to assess their performance across multiple replications. Recall that unlike the context encountered in conventional computer experiments, here we have a target surface, the true value function, that for a given sample can be approximated by the IPW-surface. The IPW-surface is only an intermediary in the whole process, and we are interested in comparing the target surface with the emulation surface, in particular with respect to the optimizer.



(a) (b) (c)

Figure 4.2: Simulation I: Emulation surfaces at +25 points overlaid over the IPW-surface in restricted domain for $\psi \in [-1, 1.2]$ (a) Int$\mathcal{GP}$ (b) HM$\mathcal{GP}$ (c) HE$\mathcal{GP}$.

Table 4.2 shows the simulations results pertaining to the optimal threshold, for each modeling type. From this table, we note that the mean across replicates is not unbiased; this is due to the multi-modality of the problem. The variability is higher for the grid-search and for the interpolating method than for the other methods. In what follows, due to the nature of the problem, we focus mainly on medians and interquartile ranges, though additional tables relating to means can be found in Appendix B.3; as expected, due to the multi-modality of the problem, the mean appears to be a biased estimator for the optimal threshold. We note further that the performance of the interpolating model degrades slightly as more samples are added, specifically with regard to the precision. The median obtained by the HM$\mathcal{GP}$ is closest to the truth, and performance seems to increase slightly as more samples are added. For this simulation, all methods perform relatively well, even after few points are sampled. We note that at 25 additional samples, all three methods outperform the grid-search, which used 300 function evaluations, as measured by the median and IQR.

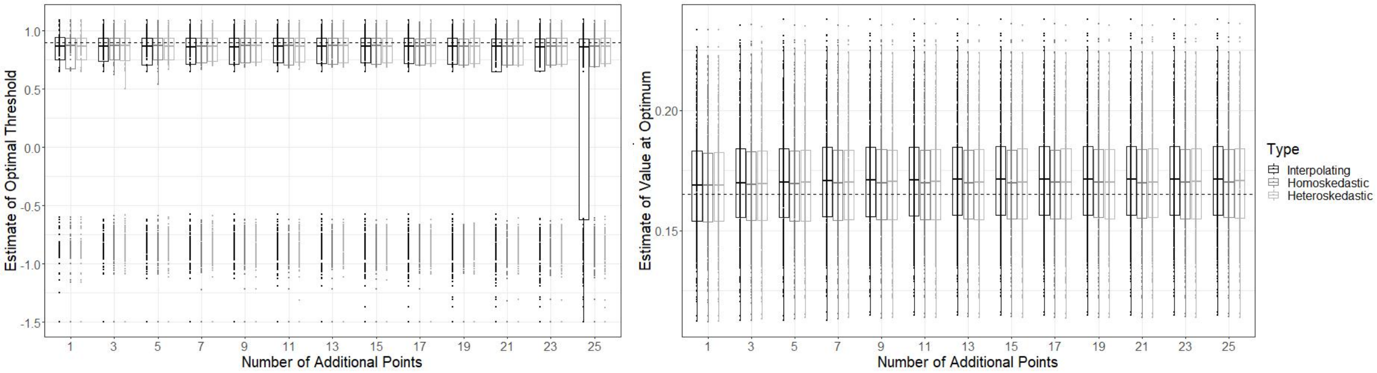
Table 4.3 shows the consequences of the estimation procedure on the value of the optimal regime. We see that, like the grid-search values in Table 4.1, these does not deviate as much as the optimizer. This is because the local and global optimizers in the value function have similar values. Figure 4.3 depicts the results for both the optimal threshold and for the value at the optimum; in panel (a) we see that the interpolating method appears to display worse performance as more points are sampled; this is an artifact of the interpolation that the method performs. From this simulation, we conclude that the HM$\mathcal{GP}$ and HE$\mathcal{GP}$, which acknowledge noise in the IPW-surface, yield results that are closest to the truth across replicates. We also conclude that any of the $\mathcal{GP}$ modeling approaches outperform the grid-search, which additionally is less computationally efficient. In Appendix B.3, we find that for a larger sample size of $n = 1000$ the performance of the grid-search improves to become comparable with the $\mathcal{GP}$ approaches.

Table 4.2: Simulation I: Estimated optimal ψ after $+m$ points; $n = 500$ with 13 design points over 500 replicates. True ψ_{opt}= 0.9.

Measure		+1	+5	+10	+15	+20	+25
Mean SD	Int$\mathcal{GP}$	0.472 (0.760)	0.474 (0.759)	0.481 (0.755)	0.475 (0.764)	0.454 (0.779)	0.440 (0.787)
Mean SD	HM$\mathcal{GP}$	0.466 (0.766)	0.501 (0.737)	0.484 (0.751)	0.477 (0.754)	0.471 (0.757)	0.469 (0.761)
Mean SD	HE$\mathcal{GP}$	0.487 (0.751)	0.504 (0.736)	0.499 (0.741)	0.479 (0.753)	0.472 (0.759)	0.476 (0.759)
Median IQR	Int$\mathcal{GP}$	0.863 (0.194)	0.867 (0.231)	0.866 (0.208)	0.867 (0.210)	0.865 (0.240)	0.861 (1.552)
Median IQR	HM$\mathcal{GP}$	0.874 (0.260)	0.873 (0.189)	0.871 (0.218)	0.869 (0.226)	0.866 (0.227)	0.868 (0.237)
Median IQR	HE$\mathcal{GP}$	0.869 (0.186)	0.868 (0.188)	0.872 (0.206)	0.866 (0.212)	0.865 (0.219)	0.865 (0.213)

Table 4.3: Simulation I: Estimated value at $\hat{\psi}_{opt}$ after $+m$ points, median (IQR); $n = 500$ with 13 design points over 500 replicates. True value at ψ_{opt}: 0.165.

	+1	+5	+10	+15	+20	+25
Int$\mathcal{GP}$	0.169 (0.029)	0.170 (0.029)	0.171 (0.029)	0.171 (0.029)	0.171 (0.029)	0.171 (0.028)
HM$\mathcal{GP}$	0.169 (0.029)	0.170 (0.029)	0.170 (0.029)	0.170 (0.029)	0.170 (0.028)	0.170 (0.028)
HE$\mathcal{GP}$	0.169 (0.029)	0.170 (0.029)	0.170 (0.029)	0.170 (0.029)	0.170 (0.029)	0.171 (0.029)



(a)

(b)

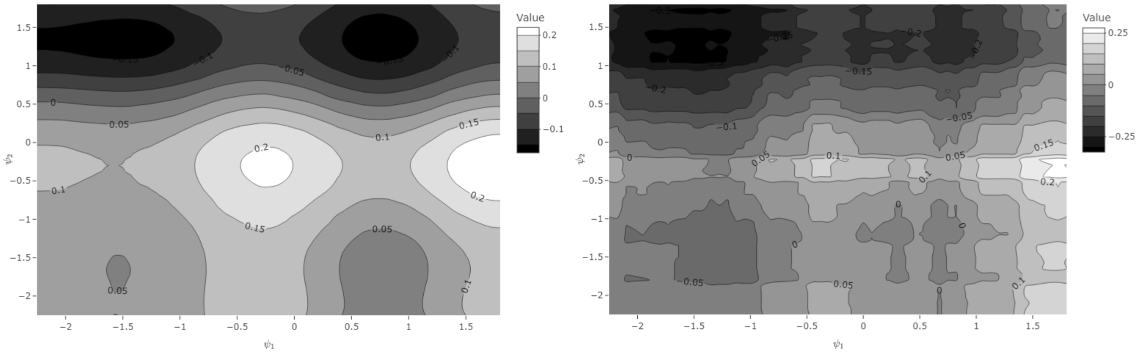
Figure 4.3: Simulation I: Boxplot at $+m$ points; $n = 500$ with 16 design points (a) Optimal ψ_1 (b) Value at optimum.

4.5.2 Simulation II

Simulation II explores a two-stage treatment rule, treat if $x_k > \psi_k$, $\psi_k \in [-2.25, 1.8]$, $k = 1, 2$. This example examines a value function that is multi-modal, with one global maximizer, and some other local maxima. The data-generating mechanism for this simulation is as follows:

$$y = 0.2x_1 - 0.2(x_1 + 2.25)(x_1 + 1.5)(x_1 + 0.3)(x_1 - 1.8)(x_1 - .75)(\mathbb{1}_{(x_1 - 1.5) > 0} - z_1)$$
$$- 0.2(x_2 + 2.1)(x_2 + 1.65)(x_2 + 0.3)(x_2 - 2.1)(x_2 - 1.35)(\mathbb{1}_{(x_2 - 0.75) > 0} - z_2) + \epsilon.$$

Intermediary variables are distributed as $x_1 \sim N(0, 1.5^2)$; $x_2 \sim 1.5z_1 + N(0, 1.5^2)$ and treatment variables as $z_1 \sim Bern(expit(-(1/1.5)x_1))$ and $z_2 \sim Bern(expit(-(1/1.5)x_2 + (1/1.5)z_1)$. Additive noise is distributed as $\epsilon \sim N(0, 0.3^2)$. In Appendix B.4, we also explore heteroskedastic additive noise. The value function is given in Figure 4.4 (a), with 3-D version found in the Interactive Supplement. As in Simulation I, this problem exhibits multi-modality, thus we focus on medians and IQRs. From Figure 4.4 (b) we observe the IPW-surface still captures the general characteristics of the value function. This can also be seen in the Interactive Supplement.



(a)

(b)

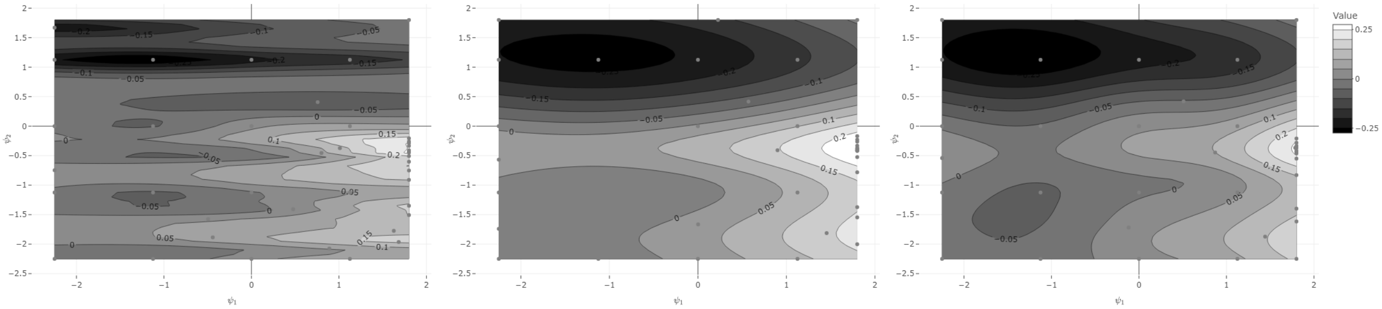
Figure 4.4: Simulation II: (a) Value function (b) IPW-surface.

An initial set of design points is taken in increments of 0.75 to yield at a total of 16 points. Before examining the results for each of these settings, we examine the results of a grid-search. From Table B.9, we see that there is a high amount of variability in the estimated optimal ψ_1 parameter, as measured by the IQR. This is similar to what was observed in Simulation I. Estimates of ψ_2 perform better, as there is no multi-modality in this axis. In what follows, we will compare the $\mathcal{GP}$ approaches to the grid-search.

Table 4.4: Simulation II: Results for grid-search with increments of 0.05 and $n = 500$. True $(\psi_{1opt}, \psi_{2opt}) = (1.8, -0.3)$; true value at optimum 0.241.

Statistic	$\hat{\psi}_{1opt}$	$\hat{\psi}_{2opt}$	Value at Optimum
Mean (SD)	1.098 (1.140)	-0.409 (0.382)	0.277 (0.094)
Median (IQR)	1.725 (1.725)	-0.375 (0.300)	0.275 (0.132)

From Figure 4.5, which shows one replicate analysis for each of the three $\mathcal{GP}$ methods, we see from the points on the plot that the cross-section at $\psi_1 = 1.8$ is explored the most; this cross-section contains the global optimizer. For this replicate, the second optimum is not well identified by any of the $\mathcal{GP}s$.



(a)

(b)

(c)

Figure 4.5: Simulation II: Contour plot of emulation surface at +25 points (a) Int$\mathcal{GP}$ (b) HM$\mathcal{GP}$ (c) HE$\mathcal{GP}$.

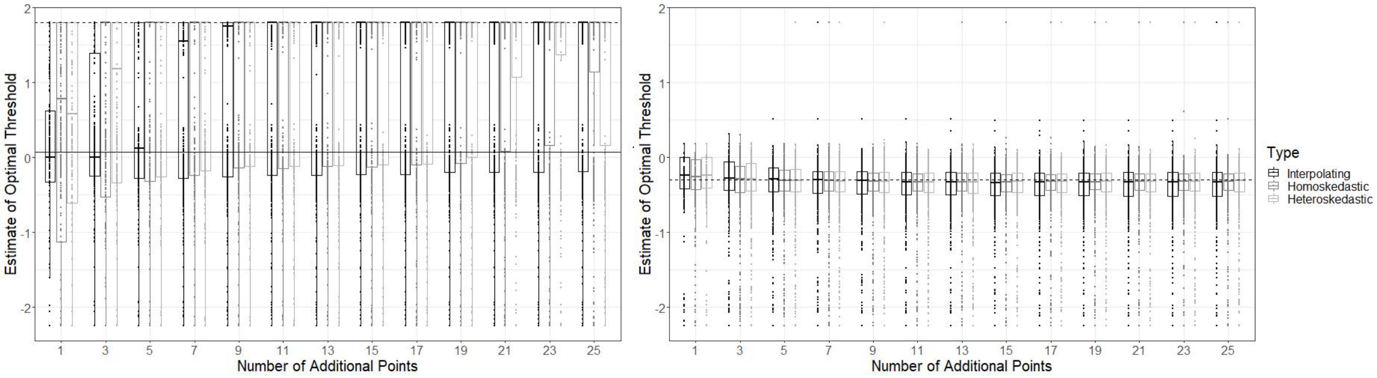
Comparing Table 4.5 with the results of the grid-search, we note that at +25 points the median optimal values resulting from the HM$\mathcal{GP}$ are closer to the truth than those arrived at via a grid-search; most notably the IQR for ψ_{1opt} is much smaller. This strengthens the observation from Simulation I that a grid-search is not always the most robust approach. We also observe that the HE$\mathcal{GP}$ outperforms the grid-search at +25 points. In the ψ_2 direction, all three methods perform similarly, with the HM$\mathcal{GP}$ and HE$\mathcal{GP}$ outperforming the grid-search at 25 additional samples. We see from Table 4.6 that all $\mathcal{GP}$s perform equally well in estimating the value at the optimum. From Figure 4.6 we can visualize how sampling additional experimental points improves the estimation of ψ_{1opt} and ψ_{2opt}. From panel (a), we see that after 11 sampled points, the first quartile and the median $\hat{\psi}_{1opt}$ are at the true value of the optimal threshold for all $\mathcal{GP}$ methods. The solid horizontal line on the plot is placed at the grid-search $IQR + \psi_{1opt}$ value. This allows us to see that after 21 sampled points the IQR for both the HM$\mathcal{GP}$ and the HE$\mathcal{GP}$ is smaller or equal to that of the grid-search IQR which uses 3721 grid points, thereby emphasizing the increased efficiency that a $\mathcal{GP}$ approach can provide. We note additionally that the HE$\mathcal{GP}$ achieves improved results slightly faster than the HM$\mathcal{GP}$, however the HM$\mathcal{GP}$ achieves comparable results after a few additionally sampled points. Panel (b) in the plot shows the results for the ψ_{2opt} parameter; we see that all $\mathcal{GP}$ methods perform consistently well, with the HM$\mathcal{GP}$ and HE$\mathcal{GP}$ performing slightly better than the Int$\mathcal{GP}$. From Figure 4.7, we see that the estimation of the optimal value is similar across all methods. Overall, this example allows us to conclude that a method that acknowledges noise in the estimation surface is important in order to more precisely estimate the optimizers. We note again that the improvement offered by the $\mathcal{GP}$ is most relevant in the direction of multi-modality. In Appendix B.4, we see that an increased sample size improves the estimation of the ψ_{1opt} parameter, but that the HM$\mathcal{GP}$ and the HE$\mathcal{GP}$ still outperform the grid-search and the interpolating approach. Although the HM$\mathcal{GP}$ seems to require slightly more data to achieve the performance of the HE$\mathcal{GP}$ for this specific setting, we must keep in mind that the HE$\mathcal{GP}$ is more computationally intensive than the HM$\mathcal{GP}$ approach as it requires fitting a $\mathcal{GP}$ on estimated residuals. We note additionally that in all simulations shown in Appendix B.4, the performance of the HM$\mathcal{GP}$ and HE$\mathcal{GP}$ is comparable

Table 4.5: Simulation II: Estimated optimal ψ_1 and ψ_2 after $+m$ points, median (IQR); $n = 500$ with 16 design points over 500 replicates. True $(\psi_{1opt}, \psi_{2opt}) = (1.8, -0.3)$.

Parameter	Method	+1	+5	+10	+15	+20	+25
ψ_{1opt}	Int$\mathcal{GP}$	0.000 (0.950)	0.123 (2.074)	1.800 (2.036)	1.800 (2.026)	1.800 (1.995)	1.800 (1.988)
ψ_{1opt}	HM$\mathcal{GP}$	0.779 (2.925)	1.800 (2.116)	1.800 (1.930)	1.800 (1.928)	1.800 (1.805)	1.800 (0.657)
ψ_{1opt}	HE$\mathcal{GP}$	0.580 (2.406)	1.800 (2.054)	1.800 (1.925)	1.800 (1.901)	1.800 (1.731)	1.800 (1.636)
ψ_{2opt}	Int$\mathcal{GP}$	-0.241 (0.420)	-0.285 (0.325)	-0.321 (0.296)	-0.334 (0.301)	-0.317 (0.311)	-0.331 (0.318)
ψ_{2opt}	HM$\mathcal{GP}$	-0.256 (0.400)	-0.306 (0.286)	-0.322 (0.242)	-0.328 (0.225)	-0.317 (0.216)	-0.318 (0.219)
ψ_{2opt}	HE$\mathcal{GP}$	-0.242 (0.411)	-0.312 (0.300)	-0.327 (0.260)	-0.319 (0.262)	-0.323 (0.249)	-0.312 (0.247)

Table 4.6: Simulation II: Estimated value at $(\hat{\psi}_{1opt}, \hat{\psi}_{2opt})$ after $+m$ points, median (IQR); $n = 500$ with 16 design points over 500 replicates. True value at $(\psi_{1opt}, \psi_{2opt})$: 0.241.

	+1	+5	+10	+15	+20	+25
Int$\mathcal{GP}$	0.196 (0.123)	0.238 (0.141)	0.258 (0.140)	0.264 (0.138)	0.264 (0.136)	0.267 (0.133)
HM$\mathcal{GP}$	0.198 (0.136)	0.234 (0.144)	0.247 (0.141)	0.259 (0.136)	0.264 (0.137)	0.265 (0.131)
HE$\mathcal{GP}$	0.200 (0.135)	0.236 (0.140)	0.255 (0.133)	0.263 (0.135)	0.264 (0.133)	0.264 (0.132)



(a)

(b)

Figure 4.6: Simulation II: Boxplot at $+m$ points; $n = 500$ with 16 design points: (a) Optimal ψ_1 (b) Optimal ψ_2. The dashed lines indicates the two optimal parameter; the solid horizontal line in panel (a) is placed at the grid-search $IQR + \psi_{1opt}$ value.

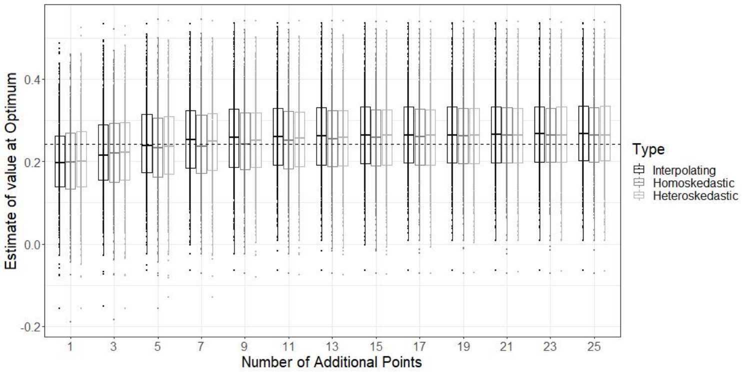
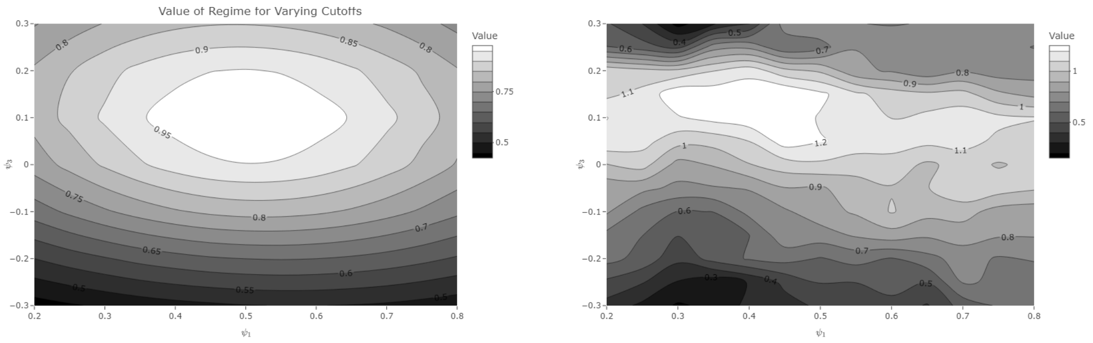


Figure 4.7: Simulation II: Boxplot of value at optimum after $+m$; n=500 with 16 design points.

4.5.3 Simulation III

For simulation III, we explore a family of regimes indexed by ψ_1, ψ_2, ψ_3 such that $\psi_1 x_{k1} + \psi_2 x_{k2} > 0.5 - 3\psi_3 u$; $k = 1, ..., 4$; x_{k1}, x_{k2} are Normally distributed intermediate covariates and u is a binary baseline covariate. Details of the data-generating mechanism can be found in Appendix B.5. In the results tables, we do not include a ψ_2 column, as we apply the following constraint: $\psi_1 + \psi_2 = 1$. Note that $\psi_1, \psi_2 \in [0.2, 0.8]$ and $\psi_3 \in [-0.3, 0.3]$. The known optimizer is $(\psi_1, \psi_3) = (0.5, 0.1)$ and the value at the optimizer is 1. A set of 20 design points is obtained by sampling in increments of 0.2 and 0.15 in ψ_1 and ψ_3 directions.

We see from Figure 4.8 (a) that this is a uni-modal setting, different from Simulation I and II. From Figure 4.8 (b) we see how the IPW-surface captures the general form of the value function. Although this is a uni-modal example, in what follows, our presentation focuses on medians and interquartile ranges, in order to maintain consistency with the other simulations. Additional tables can be found in Appendix B.5.



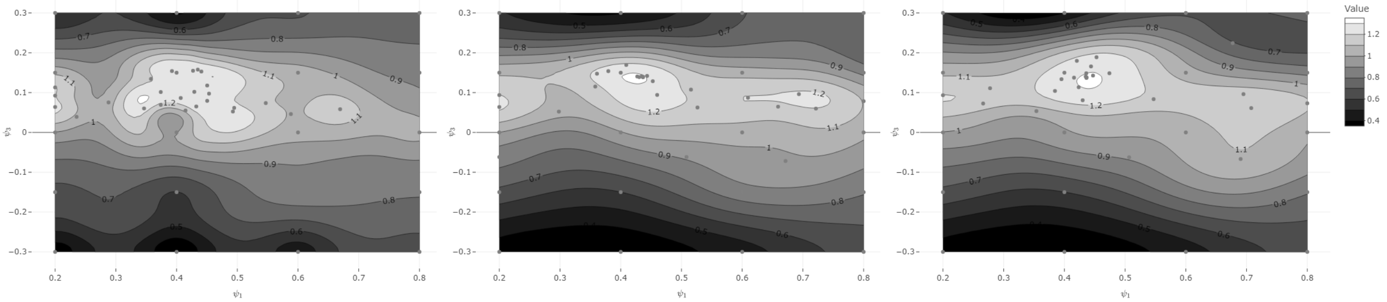
(a) (b)

Figure 4.8: Simulation III: (a) Value function (b) Estimated value function using normalized IPW.

From Table B.25, we see that the grid-search performs better than in the multi-modal examples, with variability around the optimizer decreasing. We observe from Figure 4.9 that for a fixed replicate, the HE$\mathcal{GP}$ best captures the shape of the true value function.

Table 4.7: Simulation III: Grid-search results in increments of 0.01 and $n = 500$. True $(\psi_{1opt}, \psi_{3opt}) = (0.5, 0.1)$; true value at optimum: 1.

Statistic	$\hat{\psi}_{1opt}$	$\hat{\psi}_{3opt}$	Value at Optimum
Mean (SD)	0.471 (0.153)	0.104 (0.120)	1.233 (0.147)
Median (IQR)	0.470 (0.220)	0.110 (0.150)	1.231 (0.189)



(a) (b) (c)

Figure 4.9: Simulation III: Contour plot of emulation surface at +25 points: (a) Int$\mathcal{GP}$ (b) HM$\mathcal{GP}$ (c) HE$\mathcal{GP}$.

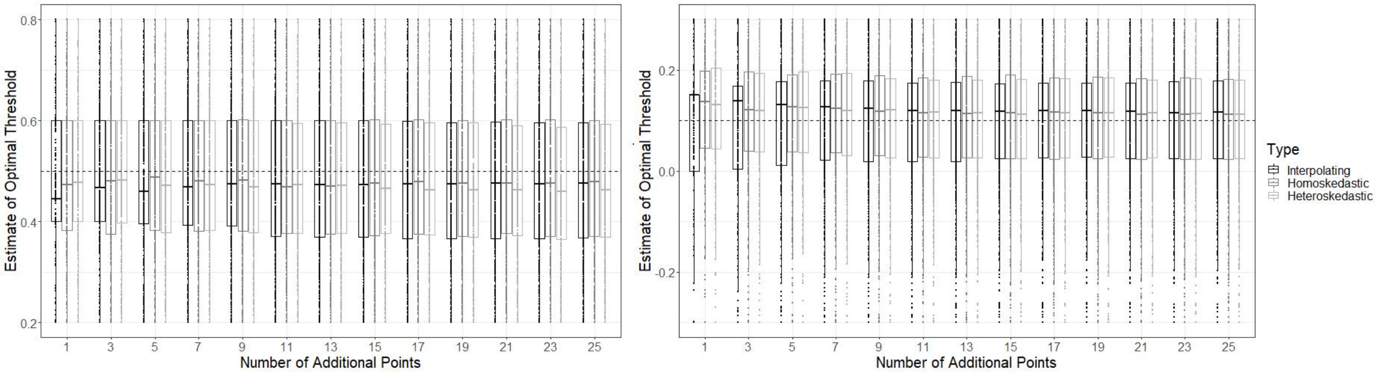
From Table 4.8, we see that all three $\mathcal{GP}$s yield good results, even for a small number of additional samples. Given the results of simulations I and II, these results suggest that the choice of $\mathcal{GP}$ modeling approach is most consequential in multi-modal settings and that there is no drawback in using a HM$\mathcal{GP}$, even if the value function is uni-modal. In a real-data analysis we do not have knowledge of whether we are in a multi-modal problem; hence, a $\mathcal{GP}$ approach that acknowledges variability in the estimation surface is advisable. We note as well that all three $\mathcal{GP}$ modeling approaches achieve good performance for a small fraction of the function evaluations required by a grid-search. Figure 4.10 shows that even at a few additional points, the optimizers are well identified. From Table 4.9, we see that as additional points are sampled, the accuracy of the estimated optimal value decreases slightly; this can also be observed in Figure 4.11. A reason for this may be that a good working model is arrived at after very few additional points in this scenario, additional sampling concentrated in one region may lead to overfitting some of the noise on the estimation surface. Three dimensional renderings related to this simulation can be found in the Interactive Supplement.

Table 4.8: Simulation III: Estimated optimal ψ_1 and ψ_3 after $+m$ points, median (IQR); $n = 500$ with 20 design points over 500 replicates. True $(\psi_{1opt}, \psi_{3opt}) = (0.5, 0.1)$.

Parameter	Method	+1	+5	+10	+15	+20	+25
ψ_{1opt}	Int$\mathcal{GP}$	0.445 (0.200)	0.460 (0.204)	0.476 (0.216)	0.473 (0.231)	0.476 (0.230)	0.476 (0.229)
ψ_{1opt}	HM$\mathcal{GP}$	0.473 (0.217)	0.488 (0.218)	0.471 (0.224)	0.475 (0.230)	0.475 (0.228)	0.479 (0.230)
ψ_{1opt}	HE$\mathcal{GP}$	0.477 (0.200)	0.471 (0.223)	0.467 (0.219)	0.466 (0.217)	0.462 (0.226)	0.462 (0.224)
ψ_{3opt}	Int$\mathcal{GP}$	0.150 (0.150)	0.131 (0.166)	0.121 (0.159)	0.118 (0.148)	0.118 (0.150)	0.116 (0.155)
ψ_{3opt}	HM$\mathcal{GP}$	0.137 (0.152)	0.127 (0.153)	0.117 (0.164)	0.115 (0.167)	0.115 (0.159)	0.112 (0.159)
ψ_{3opt}	HE$\mathcal{GP}$	0.131 (0.159)	0.125 (0.160)	0.119 (0.156)	0.113 (0.158)	0.116 (0.158)	0.112 (0.155)

Table 4.9: Simulation III: Estimated value at $(\hat{\psi}_{1opt}, \hat{\psi}_{3opt})$ after $+m$ points, median (IQR); $n = 500$ with 20 design points over 500 replicates. True value at $(\psi_{1opt}, \psi_{3opt})$: 1.

	+1	+5	+10	+15	+20	+25
Int$\mathcal{GP}$	1.118 (0.189)	1.154 (0.200)	1.174 (0.198)	1.185 (0.200)	1.187 (0.197)	1.195 (0.193)
HM$\mathcal{GP}$	1.070 (0.194)	1.108 (0.205)	1.128 (0.202)	1.147 (0.196)	1.156 (0.197)	1.160 (0.196)
HE$\mathcal{GP}$	1.074 (0.196)	1.108 (0.200)	1.129 (0.205)	1.138 (0.208)	1.148 (0.204)	1.158 (0.201)



(a) (b)

Figure 4.10: Simulation III: Boxplot after $+m$ points; $n = 500$ with 20 design points (a) Optimal ψ_1 (b) Optimal ψ_3.

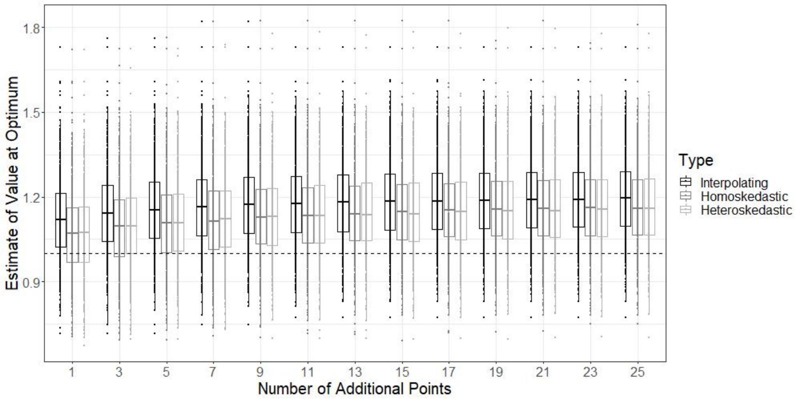


Figure 4.11: Simulation III: Boxplot of value at optimum after $+m$; $n = 500$ with 20 design points.

4.6 Case Study

In this case study, we analyze data from Hammer et al. [1996] to illustrate how the $\mathcal{GP}$ methods may be applied in a setting with real data. These data come from a double-blinded randomized trial of HIV antiretroviral therapies undertaken to compare treatments using single and double nucleosides as a means of treating HIV type 1. Patients with CD4 cell counts between 200 to 500 cells/mm^3 were enrolled in the study. A total of 2467 patients were assigned to one of four daily regimens 1) 600 mg of zidovudine, 2) 600 mg of zidovudine & 400 mg of didanosine, 3) 600mg of zidovudine & 2.5 mg zalcitabine, or 4) 400 mg didanosine. The primary end-point in the study was an observation of ≥ 50 percent decline in CD4 cell count, progression to AIDS, or death. Overall, the zidovudine regimen was found to be inferior to other regimens, with regard to the primary end-point.

Variables found in the dataset include those captured at baseline including patients' race, sex, baseline CD4, weight, age, history of antiretroviral therapy, symptoms of HIV infection, and Karnofsky score, as well as those captured later in the study such as 20 week CD4. These data may be accessed via the `LongCART` package in R [Kundu, 2021]. We restrict our analysis to the use of two dual-therapies and aim to determine which patients should be given zidovudine with zalcitabine versus zidovudine with didanosine, thereby recognizing that mono-therapy is widely considered inadequate. In particular, we examine whether tailoring on baseline CD4 cell count and baseline weight yields improved 20 week CD4 cell counts, the outcome of interest. There are no missing data in any of the variables required for this analysis. There are 524 patients in the zidovudine & zalcitabine arm and 522 in the zidovudine & didanozine arm. The known treatment probability is 0.5 by design, however we estimate these probabilities, as this has been shown to improve efficiency when using IPW estimators [Henmi and Eguchi, 2004]. Now, the specific family of regimes that we consider is: receive zidovudine with didanosine if baseline weight $> \psi_W$ and baseline CD4 $> \psi_{CD4}$, where $\psi_W \in [50, 100]$ and $\psi_{CD4} \in [200, 600]$. For every regime index (ψ_W, ψ_{CD4}), a value is estimated, and this is used to inform the resulting $\mathcal{GP}$, regardless of whether it is a one-stage decision rule or a multi-stage decision rule. In this analysis, we make use of the normalized IPW estimator for the value of a regime, and pair it with the proposed $\mathcal{GP}$ approaches.

Using the normalized IPW estimator on a fine grid of points yields the value function displayed in Figure 4.12. We see that there appears to be a trough for combinations of low ψ_W and low ψ_{CD4}; there is also a high value region for large ψ_W, across a wide range of ψ_{CD4}. From the 3-D rendering in the Interactive Supplement, we see that the IPW-surface is rather smooth; this, in part, is brought about by the use of the normalized IPW. We have also examined the resulting surface when using the standard IPW estimator, and it was characteristically more noisy, leading to the possibility of more modeling challenges.

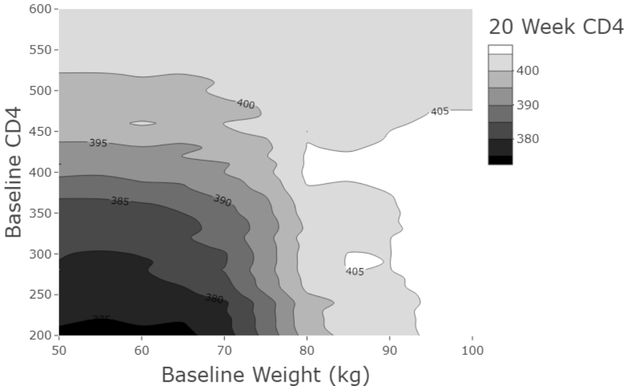


Figure 4.12: HIV Study: Contour plot of normalized IPW-surface. Axes labels show the variable associated with the threshold on the plot.

A standard approach that one may take via value-search estimation is to perform a grid-search for the optimal regime $(\psi_{opt}^{W}, \psi_{opt}^{CD4})$. It is noteworthy that for a single sample, as in this analysis, a grid-search for the optimal regime will not provide a measure of sampling uncertainty around the identified optimum. To arrive at a complete statistical analysis of these data, we seek to quantify this uncertainty by using the Bayesian bootstrap over 500 samples, where at each sample a Dirichlet vector is observed, with all concentration parameters equal to one and dimension equal to the number of patients in the study. If, for each of these bootstrap samples, we compute the optimal regime, then what results is a distribution for the optimum. We report the median optimal index and optimal value, with 95% credible intervals. We do this for both a coarse and fine grid to examine whether the grid choice impacts the results, which are shown in Table 4.10. The coarse grid has increments of 15 kg and 35 cells/mm^3 in the weight and CD4 axes, respectively; the fine grid has increments of 10 kg and 20 cells/mm^3. We see from the table that as expected, the choice of grid impacts the resulting inference. Table 4.10 also shows the results of fitting a quadratic MSM with mean $(\beta_0 + \beta_1\psi_W + \beta_2\psi_W^2 + \beta_3\psi_{CD4} + \beta_4\psi_{CD4}^2 + \beta_5\psi_W\psi_{CD4})$, using the same bootstrapping approach. The fitted model appears to fit the data relatively well, as shown in Interactive Supplement. However, note that there is no variability in the optimal ψ_W, revealing some deficiencies in the model.

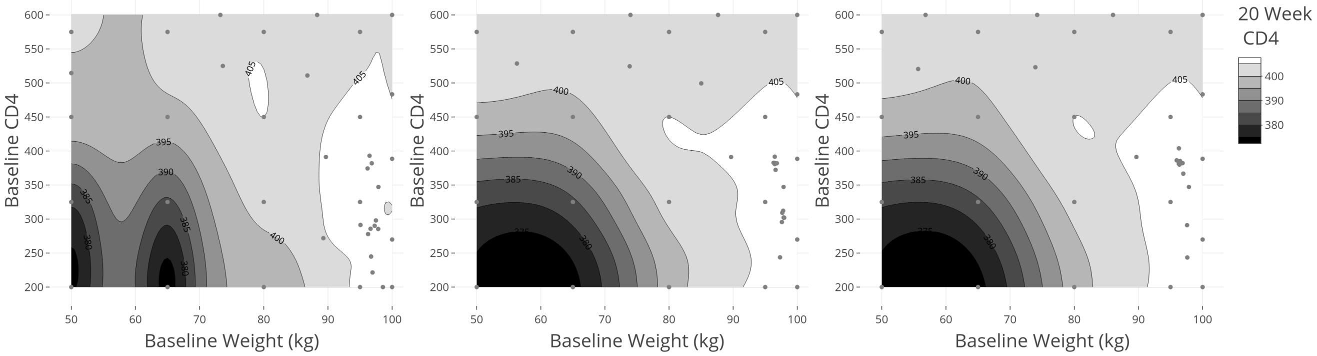
In addition to sampling uncertainty, there is another type of uncertainty that can be important to quantify in the grid-search approach. This is uncertainty reflecting the coarseness of the grid chosen. For a fixed grid, the identified optimum has uncertainty relative to the optimum that would be found were we to use a grid with increments approaching zero. However, there is no clear way to incorporate this uncertainty using a grid-search. As we will discuss, the $\mathcal{GP}$ approach can attribute more uncertainty to regions that have not been well explored and combine this with the sampling uncertainty.

Table 4.10: HIV Study: Estimated optimal value and optimal index via a coarse and fine grid-search, with 95% credible intervals.

Type	Coarse Grid	Fine Grid	MSM
$\hat{\psi}_{opt}^{CD4}$	305 (200-533)	280 (200-460)	343 (200-440)
$\hat{\psi}_{opt}^{W}$	95 (80-95)	100 (80-100)	100 (100-100)
Week 20 CD4	408 (396-421)	408 (396-421)	409 (396-423)

As before, we compare the performance of the Int$\mathcal{GP}$, HM$\mathcal{GP}$, and HE$\mathcal{GP}$. The analysis presented makes use of the $Matern_{5/2}$ covariance, and in Appendix B.6 we present the result for a $Matern_{3/2}$ covariance. A natural initial number of design points comes from creating a grid in increments of 15 kg and 125 cells/mm^3, from 50 kg to 100 kg, and from 200 cells/mm^3 to 600 cells/mm^3, respectively. This yields a total of 16 design points which is of the order explored by [Loeppky et al., 2009]. We investigated sampling up to an additional 25 points. By this point, the Expected Improvement relating to the HM$\mathcal{GP}$ and HE$\mathcal{GP}$ had reached a plateau and ψ_{opt}^{W}, ψ_{opt}^{CD4} had converged around a point; the Int$\mathcal{GP}$ did not yet show signs of complete convergence at 25 additional points, but this is not surprising, as we know in a noisy optimization setting, the interpolating method is not the most appropriate approach. Figure 4.13 shows the estimated value function for each of the modeling approaches . All three yield approximately the same conclusion regarding where the optimal regime may be, and all three methods have focused on choosing additional points in the high ψ_w region. As we noted previously, the IPW-surface is only moderately noisy. Consequently, it is not surprising that the resulting curves appear to yield similar inference, even with the interpolating $\mathcal{GP}$ model.

When a noisier estimator is used, like the un-normalized IPW estimator, we have observed the interpolating model to yield an estimated response surface that is flat everywhere except for regions very near already sampled points. This emphasizes the fact that although the proposed methodology can be used with any off-the-shelf estimator, the resulting inference can be impacted by the properties of the chosen estimator.



(a) (b) (c)

Figure 4.13: HIV Study: Contour plot of emulation surface at +25 points (a) Int$\mathcal{GP}$ (b) HM$\mathcal{GP}$ (c) HE$\mathcal{GP}$.

As with the grid-search approach, in addition to estimating the optimal regime, we are interested in providing a measure of uncertainty about this optimum. Again, there are two sources of uncertainty to consider. The first is the sampling uncertainty: how will the estimated optimum change across samples. The second relates to the uncertainty represented in the posterior $\mathcal{GP}$. This posterior informs us about uncertainty in the functional relationship between inputs and outputs, having explored a finite number of points in the index set. Consequently this can also inform us about the uncertainty in the maximizer of the functional relationship between inputs and output. To further characterize what this uncertainty represents, we should consider that if we were to sample the index space very densely; the posterior uncertainty around the curve and the consequent maximizer would be minimized. However, densely sampling the index space does not mean that the uncertainty in the maximizer has gone to zero, as there still remains sampling uncertainty.

We first examine how to quantify the posterior uncertainty and we then explore how to incorporate sampling uncertainty. To obtain a measure of the posterior uncertainty in the maximizer, after having explored an additional $+m$ points, we first compute the model parameters at $+m$ observations, and we then draw N sample paths from the posterior. For each sample path, we compute the optimizer to obtain a distribution for the optimal regime. Then, to incorporate the sampling uncertainty, we can use the Bayesian bootstrap for this procedure over 500 replicates. Ultimately, this yields a distribution of optimal regimes that represent both sources of uncertainty. We have implemented this for all three $\mathcal{GP}$ modeling approaches and the results are presented in Table 4.11. We note that the HM$\mathcal{GP}$ and HE$\mathcal{GP}$ arrive at very similar conclusions, after having explored +25 points, and that the median optimal regime in the interpolating approach is in the credible interval of the other two methods. We note additionally that the credible intervals of the Int$\mathcal{GP}$ are much wider than those of the other two methods; the HE$\mathcal{GP}$ approach results in slightly tighter credible intervals for the ψ_{opt}^{CD4} parameter. Inference at +15 points is very similar to that which results at +25 points. The estimated optimal regime is at thresholds of 98 kg and 290 cells/mm^3, yielding an expected CD4 cell count of 408 cells/mm^3. There is considerable uncertainty regarding the optimal threshold in the ψ_{CD4} direction, but this can be understood from the relatively flat relationship that appears for high values of ψ_W, as can be well explored in the Interactive Supplement. Producing Table 4.11 is a computationally intensive procedure. In Appendix B.6, we discuss some efficiency considerations for performing this analysis.

Table 4.11: HIV Study: Estimates and 95 % credible intervals for each $\mathcal{GP}$ modeling strategy; 250 sample paths; 500 Bayesian bootstrapped samples.

Model	Parameter	+1	+5	+15	+25
Int$\mathcal{GP}$	$\hat{\psi}_{opt}^{W}$	98 (66-98)	98 (58-98)	94 (54-98)	94 (54-98)
Int$\mathcal{GP}$	$\hat{\psi}_{opt}^{CD4}$	365 (200-597.5)	305 (200-597.5)	327.5 (200-597.5)	357.5 (200-597.5)
Int$\mathcal{GP}$	20 Week CD4	409.2 (397.1-421.7)	408.7 (397.2-421.09)	409.4 (398.0-425.5)	410.1 (398.3-426.3)
HM$\mathcal{GP}$	$\hat{\psi}_{opt}^{W}$	98 (66-98)	98 (66-98)	98 (78-98)	98 (78-98)
HM$\mathcal{GP}$	$\hat{\psi}_{opt}^{CD4}$	357.5 (200-597.5)	305 (200-597.5)	290 (200-522.5)	290 (200-492.5)
HM$\mathcal{GP}$	20 Week CD4	409.0 (397.0-421.4)	408.3 (396.8-420.5)	408.3 (397.15-420.2)	408.2 (397.3-420.5)
HE$\mathcal{GP}$	$\hat{\psi}_{opt}^{W}$	98 (70-98)	98 (66-98)	98 (74-98)	98 (78-98)
HE$\mathcal{GP}$	$\hat{\psi}_{opt}^{CD4}$	350 (200-597.5)	305 (200-597.5)	290 (200-515)	290 (200-462.5)
HE$\mathcal{GP}$	20 Week CD4	408.9 (397.0-421.4)	408.2 (396.6-420.4)	408.1 (396.8-420.3)	408.4 (396.9-420.5)

Increments for the sample paths were by $4kg$ in the ψ_W axis and by 7.5 cells/mm^3 in the ψ_{CD4} axis

The purpose of this case study was to show how an off-the-shelf estimator could be combined with $\mathcal{GP}$ techniques in order to arrive at a conclusion about the optimal regime. We saw that the homoskedastic and heteroskedastic analyses produce similar inference. Overall, we can conclude that there are regions of higher and lower value in the value function and that based on the HM$\mathcal{GP}$ and HE$\mathcal{GP}$ there is an optimal threshold of 98 kg and 290 cells/mm^3 in the weight and CD4 direction, respectively. There is relatively low uncertainty around ψ_{opt}^{W}, but there still remains high levels of uncertainty around ψ_{opt}^{CD4}.

4.7 Discussion

We have been motivated by the possibility that some value-search estimators may not be robust in identifying optimal DTRs, in particular Dynamic MSMs or a grid-search. We explored whether a Bayesian optimization approach via $\mathcal{GP}s$ could allow for the inference of optimal DTRs. We determined that the estimation surface resulting from the use of an estimator for the value of a DTR tends to exhibit a non-smooth quality resulting from the point-wise variation of the estimator. This led us to examine possible sources of variability in the estimation surface and to consider approaches that allow for varying noise structures. Via simulation studies, we examined the performance of three $\mathcal{GP}$ methods and found that out of these methods the HM$\mathcal{GP}$ and HE$\mathcal{GP}$ consistently yielded comparable or more accurate and precise inference than the Int$\mathcal{GP}$. Simulations also showed that a grid-search is not always the most accurate approach with $\mathcal{GP}$ methods tending to provide more accurate and precise results. These methods also required significantly less estimator evaluations to arrive at an estimate for the optimum, thereby making them more efficient than a grid-search. We conclude that there can be much to gain in using an HM$\mathcal{GP}$ or HE$\mathcal{GP}$. The performance of the HM$\mathcal{GP}$ and HE$\mathcal{GP}$ was similar across all twelve simulations, except in simulation II, where for a sample size of $n = 500$, the HE$\mathcal{GP}$ yielded more precise inference slightly faster than the HM$\mathcal{GP}$. After a few extra sampled points, the HM$\mathcal{GP}$ achieved comparable inference to the HE$\mathcal{GP}$ and that the HM$\mathcal{GP}$ is much more computationally efficient to fit, and therefore we would recommend utilizing the HM$\mathcal{GP}$ in general applications. The comparable inference that these two methods yield was confirmed in the case study, which additionally served to emphasized that this methodology can be applied meaningfully in order to identify an optimal decision rule. Additionally, the case study allowed us to examine how both sampling and posterior uncertainty in the value function can be well represented. Future work should look to examine whether a fully Bayesian treatment benefits the inferential process. Additionally, examining whether the $\mathcal{GP}$ can jointly represent sampling variability in addition to uncertainty about the value function is an important area of investigation. Studying the use of other infill criteria and examining the consequences of different stopping rules is also of methodological interest.

Chapter 5

Bayesian Inference for Optimal Dynamic Treatment Regimes in Practice

Preamble to Manuscript 3. The previous chapters proposed new methods for identifying optimal DTRs, which contrast with current approaches to the estimation of optimal DTRs. The third manuscript, contained in this chapter, seeks to advance the use of the methods proposed in the previous two chapter by showcasing a standard analysis using these methods.

The original contributions of this manuscript include i) the design of a plasmode simulation based on a clinical trial for HIV therapy to illustrate the use of Bayesian dynamic MSMs, Bayesian singly and doubly robust estimators paired with a grid-search, and Gaussian process emulation to identify optimal DTRs and ii) the creation of an `R` package that performs the analyses using methods considered in this thesis, which requires a thorough classification of the components needed to perform the proposed analyses. At the time of thesis submission, this manuscript was under review in a statistical journal.

Bayesian Inference for Optimal Dynamic Treatment Regimes in Practice

Daniel Rodriguez Duque[1], Erica E.M. Moodie[1], David A. Stephens[2].

[1] *Department of Epidemiology, Biostatistics, and Occupational Health, McGill University*

[2] *Department of Mathematics and Statistics, McGill University*

Abstract

In this work, we examine recently developed methods for Bayesian inference of optimal dynamic treatment regimes (DTRs). DTRs are a set of treatment decision rules aimed at tailoring patient care to patient-specific characteristics, thereby falling within the realm of precision medicine. In this field, researchers seek to tailor therapy with the intention of improving health outcomes; therefore, they are most interested in identifying *optimal* DTRs. Recent work has developed Bayesian methods for identifying optimal DTRs in a family indexed by ψ via Bayesian dynamic marginal structural models (MSMs) [Rodriguez Duque et al., 2022b]; we review the proposed estimation procedure and illustrate its use via the new `BayesDTR` `R` package. Although methods in [Rodriguez Duque et al., 2022b] can estimate optimal DTRs well, they may lead to biased estimators when the model for the expected outcome if everyone in a population were to follow a given treatment strategy, known as a value function, is misspecified or when a grid-search for the optimum is employed. We describe recent work that uses a Gaussian process ($\mathcal{GP}$) prior on the value function as a means to robustly identify optimal DTRs [Rodriguez Duque et al., 2022a]. We demonstrate how a $\mathcal{GP}$ approach may be implemented with the `BayesDTR` package and contrast it with other value-search approaches to identifying optimal DTRs. We use data from an HIV therapeutic trial in order to illustrate a standard analysis with these methods, using both the original observed trial data and an additional simulated component to showcase a longitudinal (two-stage DTR) analysis.

5.1 Introduction

Precision medicine builds on the concept of evidence-based medicine to determine not just the average efficacy of therapeutic or surgical interventions, but which intervention is right for whom. With this aim in mind, statisticians have sought to develop methods that allow for the discovery of tailored interventions. This has been done via statistical methods for dynamic treatment regimes (DTRs). DTRs are a set of decision rules that take patient information as inputs and that output a decision [Murphy et al., 2001]. Most importantly, researchers in this realm have proposed methods that can determine the causal effect of being assigned to a specific DTR and to identify optimal DTRs, that is, DTRs with the highest expected outcome, or value. Frequentist inference has been given much attention to this field while Bayesian methods have received significantly less heed. In this work, we examine methods that allow for Bayesian causal inference of optimal DTRs, in particular methods that can robustly identify the optimal strategy.

There are many frequentist methods for identifying optimal DTRs. These include though are not limited to g-computation [Robins, 1986], g-estimation of structural nested models [Robins, 1993], Q-learning Murphy [2005b], dynamic marginal stuctural models (MSMs) [Orellana et al., 2010a], and outcome weighted learning [Zhao et al., 2012]. Bayesian methods have also been proposed, including by Saarela et al. [2015a] who use a predictive Bayesian approach that requires the specification of parametric distributions for outcomes and intermediate covariates, Murray et al. [2018] who propose a Bayesian adaptation to Q-learning, Arjas and Saarela [2010] who use Bayesian non-parametric regression and backward induction, and Xu et al. [2016] who use Bayesian non-parametrics in a survival context, where patients can transition between disease states. Recently, a Bayesian method for inferring optimal DTRs via dynamic MSMs was developed [Rodriguez Duque et al., 2022b]. In addition to allowing for population-level inference, this approach also allows for individualized inference by enabling a decision-maker to determine whether a patient with a specific set of characteristics is receiving optimal therapy.

Although Bayesian inference via dynamic MSMs provides a means for identifying optimal DTRs, limitations remain; for example, inference hinges on the correct specification of a marginal model. A Gaussian Process ($\mathcal{GP}$) prior has recently been proposed to model the value function and consequently identify optimal DTRs via a sequential sampling scheme [Rodriguez Duque et al., 2022a]. In principle, $\mathcal{GP}$-based methods can utilize any estimator for the expected outcome under adherence to a DTR, known as the value for the regime, and avoid some of the drawbacks associated with some value-search approaches. For example using a dynamic MSM to directly model the value surface may not perform well if the model is incorrectly specified. Alternatively, if a grid-search is used to obviate the issues of directly modeling the value surface, an inefficient procedure results which may not correctly identify the optimal DTR when the value function is multi-modal and which may be computationally intractable when Bayesian estimators are utilized. Although identifying a local optimum may result in improved treatment over standard practice, identifying the global optimum is of importance, as only then can we claim a patient is receiving optimal therapy, especially when the difference in value between the optima is of clinical significance. A Bayesian approach that makes use of $\mathcal{GP}$s to represent uncertainty in the value function has the potential to more efficiently utilize information by strategically selecting experimental points that are expected to be optimizers and by providing a very flexible model for the value function.

In this paper, we aim to review how a Bayesian approach may capitalize on semiparametric inference as presented in Saarela et al. [2015b] and Rodriguez Duque et al. [2022b] in order to identify optimal DTRs. This is important as the ideas required for this inferential approach are nuanced and therefore challenging for practitioners to implement. We introduce a new package `BayesDTR` to illustrate how to utilize these methods in practice. With these foundations in place, we further study how $\mathcal{GP}$ optimization can be used to identify optimal DTRs, and we examine how the `BayesDTR` package provides functionalities to perform an analysis reliant on these methods. There are several packages available in the Comprehensive `R` Archive Network (CRAN) that performs estimation or inference about DTRs. These include `DTRreg` which implements dynamic weighted least squares, g-estimation, and Q-learning [Wallace et al., 2020], `DTRlearn2` which performs outcome weighted learning [Chen et al., 2020], `DynTxRegime` which permits several methods including inverse probability weighting (IPW) and augmented IPW [Holloway et al., 2020], and `SMARTbayesR` which allows for Bayesian inference of optimal DTRs with data arising from SMART designs with binary outcomes [Artman, 2021]. Currently, there are no packages that allow for Bayesian semiparametric inference of optimal DTRs, nor any that directly use $\mathcal{GP}$ optimization with estimators for the value of a DTR.

This article is organized as follows: section 2 introduces recently developed Bayesian methods for identifying optimal DTRs, section 3 describes the functions in the `BayesDTR` that allow for the use of these methods. For illustrative purposes, we adapt data from a clinical trial on HIV therapy, available in the `LongCART` `R` package [Kundu, 2021], to perform a plasmode simulation depicting a standard analysis with these methods and package. Section 4 demonstrates how to use this package to perform a standard analysis with these methods. We summarize and conclude in section 5.

5.2 Methods

5.2.1 Bayesian Dynamic MSMs for Optimal Dynamic Regimes

In this section, we examine how to perform inference for optimal DTRs via dynamic MSMs, using the methods developed by Rodriguez Duque et al. [2022b]. To do this, the inferential setting must first be formalized and notation defined. Consider a multi-stage decision problem with K decision points and final continuous-valued outcome y. At every decision point k, a set of covariates x_k is observed. It is assumed that these consist of all time-fixed and time-varying confounders, if there are any. Covariate history up to time k is denoted by

$\bar{x}_k = \{x_1, ..., x_k\}$, and observed treatment history up to stage k is given by $\bar{z}_k = (z_1, ..., z_k)$, $z_j \in \{0, 1\}$. Subscripts are omitted when referencing history through stage K. All patient information is grouped into $b = (\bar{x}, \bar{z}, y)$. As interest is centered around examining the effect of adherence to specific DTRs, the DTR-enforced treatment history can be considered by $g(\bar{x}) = (g_1(x_1), ..., g_K(\bar{x}_K))$, $g_i(\bar{x}_i) \in \{0, 1\}$. This is the sequence of treatments that would be observed if a patient followed a treatment strategy g throughout the entire follow-up period; it contrasts the treatment history $\bar{z}$ that is observed in patients in an analytic dataset. The observed treatment history $\bar{z}$ and the DTR-enforced treatment history $g(\bar{x})$ only coincide in patients who have treatments consistent with those suggested by a DTR g. The DTR-enforced treatment history up to stage k is given by $\bar{g}_k(\bar{x}_k) = (g_1(x_1), ..., g_k(\bar{x}_k))$. Attention is restricted to a family of DTRs indexed by $\psi \in \mathcal{I}$ to give $\mathcal{G} = \{g^{\psi}(\bar{x}); \psi \in \mathcal{I}\}$. Note that we only consider deterministic DTRs that assign treatment deterministically using patient information. An example of a DTR indexed by a parameter ψ is one of the form "treat at stage k when $x_k > \psi_k$". Interest lies in two treatment and covariate distributions: the *observational* world distribution $P_{\mathcal{O}}$ which denotes the law giving rise to the data in the study population, and the *experimental* world distribution $P_{\mathcal{E}}$, which is problem specific, and should be defined such that causality can be inferred. Under these two worlds, the marginal distribution of x_1 is identical, and the dependence of x_k on previous treatment and covariates is also unchanged for $k = 2, ..., K$, however the component of the joint distribution governing treatment allocation differs. For example, Saarela et al. consider a world in which treatments are sequentially randomized so that stage-specific treatment effects can be estimated [Saarela et al., 2015b]. Rodriguez Duque et al. focus on an experimental world where patients are randomly assigned to a DTR in $\mathcal{I}$ at study start [Rodriguez Duque et al., 2022b]. Lastly, variables sampled from a posterior distributions are shown with *.

Inference for Bayesian dynamic MSMs begins by considering a utility $U(b, g^{\psi}, \beta,)$; focus is on the negative squared error loss utility, $U(b, g^{\psi}, \beta,) = -(y - h(\beta, \psi))^2$, where $h(\beta, \psi)$ models $E_{\mathcal{E}}[Y|G = g^{\psi}]$, indexed by an unknown parameter β and where the expectation is taken with respect to the true data-generating distribution in the experimental world. This utility is of interest because it allows for an explicit model of the quantity of interest, the expected outcome under assignment to a regime g^{ψ}, in a world where regime assignment is unconfounded. For a Bayesian decision-maker, interest lies in the value of β that maximizes the posterior expected utility $E_{\mathcal{E}}[U(B^*, G, \beta)|\bar{b}]$. This is an expectation taken with respect to the experimental measure in which patients are randomized to regimes in $\mathcal{G}$ at study start, with probability $p(G = g)$. The basis for this decision theoretic approach is well laid out in Walker [2010]. When a finite set of regimes is considered, with patients having equal probability of randomization, $p(g)$ can be replaced with $1/C_G$, where $C_G = |\mathcal{I}|$. With a chosen utility, the next step in this approach lies in linking $\mathcal{E}$ and $\mathcal{O}$ with respect to a posterior predictive distribution. The required linkage is given by the following equation:

$$E_{\mathcal{E}}[U(B^*, G, \beta)|\bar{b}] = E_{\mathcal{O}}\left[\frac{1}{C_G}\sum_{\{\psi\in\mathcal{I}\}} w^{\psi*}U(B^*, g^{\psi}, \beta)\middle|\bar{b}\right], \tag{5.1}$$

with weight $w^{\psi*}$ given by

$$w^{\psi*} = \frac{\mathbb{1}_{g^{\psi}(\bar{X}^*)}(\bar{Z}^*)}{\prod_{j=1}^{K} p_{\mathcal{O}}(Z_j^*|\bar{Z}_{j-1}^*, \bar{X}_j^*, \bar{b})}. \tag{5.2}$$

The denominator in the weight is the treatment probability in the observational world. The numerator is the probability of a sequence of treatments conditional on regime assignment; as only deterministic DTRs are considered, these probabilities are either 0 or 1, thereby yielding the indicator function. Randomization to regime g^{ψ} is equiprobable for all regimes in the experimental world, and this is captured by the constant C_G. The * notation clarifies that the expectation in equation (5.1) is taken with respect to a posterior predictive distribution. For equation (5.1) to hold, a patient following regime g^{ψ} with recorded history $(\bar{x}_K, \bar{z}_K)$ should have a positive probability of being observed in the observational world; effectively this is the positivity condition encountered in the causal inference literature [Murphy et al., 2001]. Additionally, as is frequently found in the causal inference literature, the sequential no unmeasured confounders assumption is also required [Murphy et al., 2001]. Note that the weight formula is *not* stabilized and that having a stabilization term involving the marginal treatment probabilities in $\mathcal{O}$ would change $\mathcal{E}$ to one where the marginal treatment probabilities are as in $\mathcal{O}$. It may be that the resulting probability law in $\mathcal{E}$ is not well defined, given that treatments in $\mathcal{E}$ are dictated by DTRs.

Having linked the experimental world with the observational world, focus becomes centered on how to infer about the parameters of interest β. Equation (5.1) now allows for the use of observed world data to perform posterior inference in the experimental world. To perform Bayesian inference in this setting, a prior must be specified. Unlike parametric Bayesian inference, where a prior for β is specified directly, a prior is placed on the family of data generating distributions in the observational world $P_{\mathcal{O}}$, denoted by $P_{\mathcal{F}}$. Effectively, this prior induces a prior on β as $P_B(\beta \in \Omega) = P_{\mathcal{F}}(\{P_{\mathcal{O}} : \beta(P_{\mathcal{O}}) \in \Omega\})$. The prior of choice is the non-parametric Dirichlet process $\mathcal{DP}(\alpha, G_x)$ prior with scaling parameter $|\alpha| \to 0$. This prior has the benefit of converging asymptotically to the true data-generating distribution [Ghosal and van der Vaart, 2017]. Under this specification, the Bayesian bootstrap yields the posterior predictive distribution [Rubin, 1981]. A sample drawn from the posterior $\mathcal{DP}$ is given by $p_{\mathcal{O}}(b^*|\bar{b}, \pi) = \sum_{i=1}^{n} \pi_i \mathbb{1}_{b_i}(b^*)$, where $\pi = (\pi_1, ..., \pi_n)$ is a sample from $\pi \sim Dir(1, ..., 1)$, a Dirichlet distributed random variable with all concentration parameters equal to one. Under the $\mathcal{DP}$ prior that yields the Bayesian bootstrap, any distribution sampled from the posterior $\mathcal{DP}$ is uniquely determined by π. Stephens et al. [2022] provide further details on the $\mathcal{DP}$ model and its consequences on Bayesian causal inference. Incorporating these prior assumptions allows for the expected posterior experimental world utility to be computed as:

$$E_{\mathcal{E}}[U(B^*, G, \beta)|\bar{b}] = E_{\pi}[E_{\mathcal{E}}[U(B^*, G, \beta)|\bar{b}, \pi]] = E_{\pi}\left[\frac{1}{C_G}\sum_{i=1}^{n}\sum_{\psi \in \mathcal{I}} \pi_i w_i^{\psi} U(b_i, g^{\psi}, \beta)\right]. \quad (5.3)$$

Note that the right-most expression depends only on observed data, hence the * notation is dropped; this includes dropping the * from the weight w_i^{ψ}, as it is no longer a random variable but rather an instantiation of that random variable. With this expression for the posterior expected utility, focus turns to maximization. The maximizer of the experimental world expected posterior utility can be obtained by solving: $\beta_{opt} = \arg\max_\beta \; E_\pi \left[\sum_{i=1}^n \pi_i \sum_{\psi \in \mathcal{I}} w_i^\psi U(b_i, g^\psi, \beta)\right]$.

Uncertainty in β_{opt} may be characterized by noting that β_{opt} is a deterministic function of π, in arguments similar to those in Walker [2010]. Thus, draws from the posterior distribution of β_{opt} can be done via:

$$\beta^*_{opt}(\pi) = \arg\max_\beta \sum_{i=1}^n \pi_i \sum_{\psi \in \mathcal{I}} w_i^\psi U(b_i, g^\psi, \beta).$$

This relationship emphasizes that uncertainty in the posterior distribution reflects uncertainty in β_{opt}. C_G may be disregarded for the purposes of predictive inference. This is an exact Bayesian procedure, modulo Monte Carlo error. Under the specified negative squared error loss utility inference is arrived at by solving:

$$\beta^*_{opt}(\pi) = \arg\max_\beta \left[-\sum_{i=1}^n \pi_i \sum_{\psi \in \mathcal{I}} w_i^\psi (y_i - h(\beta, \psi))^2 \right]. \tag{5.4}$$

From equation (5.3), it is evident that to draw inference in the experimental world, the weight w^ψ needs to be computed; this leads to modeling the treatment assignment probabilities. For each draw of π a model $p_{\mathcal{O}}(z_j|\bar{z}_{j-1}, \bar{x}_j, \gamma_j(\pi))$, $k = 1, ..., K$ can be considered. The parameters γ_j may be regarded as coming from a posterior utility maximization framework with the same $\mathcal{DP}$ prior. When the utility is the log-likelihood, the following maximization is required:

$$\gamma^*_{k,opt}(\pi) = \arg\max_{\gamma_k} \sum_i^n \pi_i \log p_{\mathcal{O}}(z_{i,k}|\bar{z}_{i,k-1}, \bar{x}_{i,k}, \gamma_k).$$

Then, for every draw π, the weighted treatment propensity model can be fit, and the resulting weight, w^ψ, in equation (5.2) is now dependent on π. Effectively, for each draw π, computing $\beta_{opt}(\pi)$ is coupled with computing $\gamma_{k,opt}(\pi)$. Thus across draws of π, uncertainty in γ_j is being incorporated into the estimation procedure. From a practical perspective, the `glm` function in R can be used to fit these models, making use of the weights argument to supply the relevant information; the optimizer for the negative squared error loss utility is the same as that which maximizes the Gaussian likelihood. There is some flexibility in the specification of $h(\beta, \psi)$. One example is $h(\beta, \psi) = \beta_0 + \beta_1\psi + \beta_2\psi^2$, which can be maximized analytically to identify an optimal DTR. To fit this model requires plugging in $h(\beta, \psi)$ into equation (5.4) and solving it. To solve equation (5.4) requires a data augmentation procedure that duplicates patient data rows for as many regimes as to which they are adherent. This procedure is detailed in Cain et al. [2010] with further considerations for the specification of these models found in Rodriguez Duque et al. [2022b]. A description for the estimation procedure can be found in Algorithm 1.

```
Data: $DATA_{\mathcal{O}}$                              // One row per patient; $n$ patients
for $\psi \in \mathcal{I}$ do                  // Create $AUGDATA_{\mathcal{O}}$ based on regime adherence
    Duplicate rows of $DATA_{\mathcal{O}}$ for patients adherent to regime $g^{\psi}$
    Add column specifying regime index $\psi$
end
Posit model for $h(\beta, \psi)$
for $i \leftarrow 1$ to $B$ do                             // B is number of posterior draws
    Draw $\pi = (\pi_1, ..., \pi_n)$ from $Dir(1, ..., 1)$
    Using weighted logistic regression, estimate $p_{\mathcal{O}}(z_k|\bar{z}_{k-1}, \bar{x}_k, \gamma_k, \pi)\ \forall k$
    Compute weights $w_i^{\psi}$, $i = 1, ..., n$, using probabilities in the previous step
    Add weights to $AUGDATA_{\mathcal{O}}$
    Perform regression with mean $h(\beta, \psi)$ and with weights $\pi_i w_i^{\psi}$ to obtain $\beta^*_{opt}(\pi)$
    Maximize $h(\beta(\pi), \psi)$ to obtain a sample from $\psi^*_{opt}(\pi)$
end
Output: Posterior distribution of $\psi^*_{opt}$
```

$DATA_{\mathcal{O}}$ is an input dataset with one row per patient and is used to fit treatment models. $AUGDATA_{\mathcal{O}}$ is an augmented dataset, where patients are duplicated for each DTR to which they adhere. This dataset is used to run regression for $h(\beta, \psi)$.

Algorithm 1: Fitting procedure for Bayesian dynamic MSM and for identifying ψ_{opt}.

In the remainder of this section, we introduce three additional Bayesian methods of estimation and inference for optimal DTRs (via inverse weighting with a grid-search in section 2.2, via a doubly robust grid-search approach in 2.3, and using Gaussian processes to emulate the value function in 2.4, with additional considerations for individualized inference in section

2.5 and normalization of IPW weights relevant to the methods of sections 2.2-2.3 in section 2.6). As with dynamic MSMs, the target of inference of the methods in sections 2.2 and 2.3 is a marginal mean, namely the expected outcome under adherence to a regime g^{ψ}. Importantly, one difference in the terminology used in this paper is that dynamic MSMs allow for parametric models of the value given g^{ψ} for a continuum of ψs, whereas methods in sections 2.2 and 2.3 target the value of a regime, one regime at a time.

5.2.2 Optimal DTRs via Bayesian IPW Inference and a Grid-Search

It may be that we want to avoid using the methods in the previous section, as we do not want to model the value function directly with $h(\beta, \psi)$. This can be because an incorrectly specified model may lead to incorrectly identifying the optimal DTR. One way to avoid this is to use an estimator for the value of a DTR and to perform a grid-search for the optimum over the indices $\mathcal{I}$ in a family. This requires estimating the value under adherence to a regime for a discrete grid of indices, $\mathcal{I}_{grid} \subseteq \mathcal{I}$. One way to estimate the value of each regime in the grid is to use the Bayesian IPW which uses a similar framework as in the previous section, with a few differences. In particular, posterior predictive inference is paired with IPW to yield an inferential procedure that uses weighting to create an importance sampling projection of $\mathcal{O}$ into a regime-enforced world were everyone in the study population follows a fixed regime g^{ψ}. This allows us to use data from $\mathcal{O}$, where not all patients follow the regime of interest, to infer about a regime-enforced world. This regime-enforced world contrasts the previously considered experimental world where patients are randomized to DTRs in a family at baseline. If we use this method of estimation to compute $E_{g^{\psi}}[Y^{*}|\bar{b}]$ for each regime in the grid, we can then identify the regime yielding the highest value.

As with the Bayesian MSM, an importance sampling argument and a $\mathcal{DP}$ prior on the observational world data-generating distribution, leads us to the following:

$$E_{g^\psi}[Y^*|\bar{b}] = E_\pi[E_{g^\psi}[Y^*|\bar{b},\pi]] = E_\pi\left[\sum_{i=1}^{n}\pi_i w_i^\psi y_i\right]. \tag{5.5}$$

The weights w_i^ψ are computed as in the previous section. Over repeated draws of π, we can compute an estimate for $E_{g^\psi}[Y^*|\bar{b}]$ and its associated variability, relying again on the Bayesian bootstrap to provide an appropriate posterior predictive distribution. Defining $\tilde{y}^\psi(\pi) = E_{g^\psi}[Y^*|\bar{b},\pi]$, for conciseness, the optimal regime and its associated variability can be obtained by computing $\psi_{opt}(\pi) = \arg\max_{\psi\in\mathcal{I}_{grid}}\{y^{\psi_1}(\pi), ..., y^{\psi_p}(\pi)\}$, where $p = |\mathcal{I}_{grid}|$, for each draw of π. The treatment models can be incorporated into the estimation procedure in the same way as in the previous section.

In practice, for each draw of π, treatment models are fit using the entire observed data and the probability that patients received the treatment they were observed to receive, $p_{ik}(\pi) = p_{\mathcal{O}}(z_{i,k}|\bar{z}_{i,k-1}, \bar{x}_{i,k}, \gamma^*_{k,opt}(\pi))$, is computed for each decision point. Then, for each regime in $\mathcal{I}_{grid}$, weights w^ψ are computed. Patients who do not follow regime g^ψ will have a weight of zero, meaning they do not contribute directly to the IPW expression. Patients who do follow regime g^ψ have weights that depend on $p_{ik}(\pi),\ k = 1, ..., K$. Although patients who do not adhere to regime g^ψ do not contribute directly to the IPW expression, they do contribute to the analysis as they inform the treatment assignment models. Once the value of each regime in $\mathcal{I}_{grid}$ has been estimated, the regime that optimizes the value can be identified in order to identify $\psi^*_{opt}(\pi)$. This procedure can be repeated over draws of π in order to obtain the posterior distribution of the optimal regime.

5.2.3 Optimal DTRs via Bayesian Doubly Robust Inference and a Grid-Search

Another related approach, which has been explored by Rodriguez Duque et al. [2022b] to identify optimal DTRs, is to perform a grid-search using Bayesian posterior predictive inference

ence and the doubly robust (DR) estimator proposed by Orellana et al. [2010a]. Bayesian predictive inference was first paired with doubly robust estimators by Saarela et al. [2016] to estimate the effect of static treatment regimes. Attention is first given to the characteristics of this estimation approach in order to arrive at a Bayesian estimate of the expected outcome under adherence to a regime g^{ψ}. In the context of identifying an optimal DTR, the doubly robust estimator can then be used to estimate the value of a discrete set of regimes in a family indexed by $\mathcal{I}$ and the optimal regime in the family identified via a grid-search, as presented in the previous section. This means that, like in the previous section, a model $h(\beta, \psi)$ does not need to be specified. In particular, the DR estimator used yields consistent inference when either a set of treatment models is correctly specified or when a set of outcome models is correctly specified. Thus, in addition to fitting a sequence of treatments models, as is needed with the IPW estimator, the doubly robust estimator requires that a sequence of conditional outcomes $\phi_k^{\psi *}$, $k = 1, ..., K$ be estimated. These are defined as

$$\phi_K^{\psi *}(\bar{x}_K) = E_{\mathcal{O}}[Y^* | \bar{X}_K^* = \bar{x}_K, \bar{Z}_K^* = \bar{g}_K^{\psi}(\bar{x}_K), \bar{b}] \text{ for } k = K \text{ and as}$$
$$\phi_k^{\psi *}(\bar{x}_k) = E_{\mathcal{O}}[\phi_{k+1}^{\psi *}(\bar{x}_{k+1}) | \bar{X}_k^* = \bar{x}_k, \bar{Z}_k^* = \bar{g}_k^{\psi}(\bar{x}_k), \bar{b}] \text{ for } k = K-1, ..., 1.$$

Note that these expectations are taken with respect to the probability distribution form the observational world, conditional on subjects who have covariate history $\bar{x}_k$ and who followed the regime g^{ψ} up to time k. These $\phi_k^{\psi *}$ can be interpreted as the posterior expected outcome conditional on covariates $\bar{x}_k$ and treatments $\bar{z}_k = \bar{g}_k(\bar{x}_k)$ in a world where regime g^{ψ} is followed from stage $k+1$ to K. We use the * notation on the ϕs to emphasize that they are expectations taken with respect to a posterior distribution. Further details on these quantities can be found in Orellana et al. [2010a]. It can be shown via a conditional expectation argument that $E_{g^{\psi}}[Y^* | \bar{b}] = E_{\mathcal{O}}[\phi_1^{\psi *}(X_1^*) | \bar{b}]$, the estimand of interest.

The next section describes how models for $\phi_k^{\psi *}$ may be fit using regression by parameterizing them with τ such that $\phi_k^{\psi *}(\bar{x}_k) = \phi_k^{\psi *}(\bar{x}_k; \tau)$. With these models fit, uncertainty in the parameters can be treated analogously to how uncertainty in γ is treated: it is made dependent on π via the Bayesian bootstrap. Rather than positing a likelihood model as was done for the treatment assignment mechanism, a negative squared error loss utility can be maximized instead. The result is that for every draw of π, $\phi_k^*(\bar{x}_k, \tau(\pi))$ can be estimated.

Now that these outcome models have been specified, it remains to provide an expression that exhibits the double robustness property when the expectation is taken with respect to the true data generating mechanism. Such an expression is obtained from the following equality:

$$E_{g^\psi}[Y^*|\bar{b}] = E_{\mathcal{O}}\left[\phi_1^{\psi*}(X_1^*) + \sum_{k=2}^{K} w_{k-1}^{\psi*}(\phi_k^{\psi*}(\bar{X}_k^*) - \phi_{k-1}^{\psi*}(\bar{X}_{k-1}^*)) + w_K^{\psi*}(Y^* - \phi_K^{\psi*}(\bar{X}_K^*))\middle|\, \bar{b}\right]. \tag{5.6}$$

Then, the expression inside the expectation on the right hand side exhibits the double robustness property if the outcome models $\phi^{\psi*}$ are correctly specified or if the treatment models in $w_k^{\psi*}$ are correctly specified. Note that parameters γ and τ in the models have been suppressed for brevity. We note that for this expression to possess the desired property, the positivity condition and the no unmeasured confounders assumption in Orellana et al. [2010a] must be met. To incorporate the sampling scheme, a single sample from the posterior distribution of the estimand of interest can be obtained by conditioning on a single draw π in order to obtain:

$$E_{g^\psi}[Y^*|\bar{b},\pi] = \sum_{i=1}^{n}\pi_i\left[\phi_{i1}^{\psi^*}(x_{i1}) + \sum_{k=2}^{K} w_{ik-1}^{\psi}(\phi_{ik}^{\psi*}(\bar{x}_{ik}) - \phi_{ik-1}^{\psi*}(\bar{x}_{ik-1})) + w_{iK}^{\psi}(y_i - \phi_{iK}^{\psi*}(\bar{x}_{iK}))\right]. \tag{5.7}$$

By resampling Dirichlet weights, $E_{g^\psi}[Y^*|\bar{b}] = E_\pi\left[E_{g^\psi}[Y^*|\bar{b},\pi]\right]$ and its associated uncertainty can be computed. Models for the ϕs and ws are coupled with π and may be incorporated into the inferential process as was done with the IPW estimators of the previous two sections. To arrive at an optimal regime, this DR estimator can be used to perform a grid-search for the optimum.

Fitting Outcome Models

To fit the outcome models, some additional definitions are required. First, define the function

$$Q_K^{\psi*}(\bar{x}_K, \bar{z}_K) = E_{\mathcal{O}}[Y^*|\bar{X}_K^* = \bar{x}_K, \bar{Z}_K^* = \bar{z}_K, \bar{b}] \text{ and the stage } K \text{ pseudo-outcome as}$$
$$\Delta_K^{\psi*}(\bar{x}_K, \bar{z}_{K-1}) = Q_K^{\psi*}(\bar{x}_K, \bar{z}_{K-1}, z_K = g^{\psi}(\bar{x}_K)).$$

This is the expected outcome under observed treatment and covariate values, except for at stage K where treatment is assigned according to regime g^{ψ}. For the remaining stages $k = K-1, ..., 1$ define

$$Q_k^{\psi*}(\bar{x}_k, \bar{z}_k) = E_{\mathcal{O}}[\Delta_{k+1}^{\psi*}|\bar{X}_k^* = \bar{x}_k, \bar{Z}_k^* = \bar{z}_k, \bar{b}], \text{ with stage } k \text{ pseudo-outcome}$$
$$\Delta_k^{\psi*}(\bar{x}_k, \bar{z}_{k-1}) = Q_k^{\psi*}(\bar{x}_k, \bar{z}_{k-1}, z_k = g^{\psi}(\bar{x}_k)).$$

Then, as elaborated on in Tsiatis et al. [2019], we can compute the quantities of interest through $\phi_k^{\psi*}(\bar{x}_k) = Q_k^{\psi*}(\bar{x}_k, \bar{g}_k(\bar{x}_k))$ for $k = 1, ..., K$. Of course, in practice $Q_k^{\psi*}$ and $\Delta_k^{\psi*}$ are unknown, consequently regression models for $Q_k^{\psi*}$ should be fit and $\Delta_k^{\psi*}$ predicted based on these models. Once all models for $Q_k^{\psi*}$ have been fit, then $\phi_k^{\psi*}(\bar{x}_k)$ can be estimated. The functions in the `BayesDTR` package render the estimation of these outcome models straightforward, as all that is required is that users specify the stage-specific models for the pseudo-outcomes (or outcome if at the final stage); the package will perform the required computations in order to arrive at a fit for the $\phi_k^{\psi*}$s. This regression approach is one of several ways to fitting the required outcome models, with Tsiatis et al. [2019] expanding on other methods that can be used.

With these definitions, we now provide a two-stage example of how to obtain estimates for the $\phi_k^{\psi*}$s. For illustrative purposes, we omit notation pertaining to posterior inference, and then comment on how to incorporate this. The estimation procedure begins by specifying the following two models:

$$Q_2^\psi(\bar{x}_2, \bar{z}_2) = E[y|\bar{x}_2, \bar{z}_2] = \beta_{21}x_1 + (\beta_{22} + \beta_{23}x_1)z_1 + \beta_{24}x_2 + (\beta_{25} + \beta_{23}x_2)z_2, \tag{5.8}$$

$$Q_1^\psi(x_1, z_1) = E[\Delta_2^\psi|x_1, z_1] = \beta_{11}x_1 + (\beta_{12} + \beta_{13})z_1, \tag{5.9}$$

where $\Delta_2^\psi = E[Y|\bar{X}_2 = \bar{x}_2, Z_1 = z_1, Z_2 = g_2^\psi(\bar{x}_2)]$. We can use, for example, the `lm` function in `R` to fit these models. Note that Δ_2^ψ is not observed and so it must be predicted using the stage two model. Once these models have been fit, we may compute the outcomes for the doubly robust estimator by using the data and the estimated models to predict:

$$\phi_2^\psi(\bar{x}_2) = Q_2^\psi(\bar{x}_2, \bar{g}^\psi(\bar{x}_2)) = \beta_{21}x_1 + (\beta_{22} + \beta_{23}x_1)g_1^\psi(x_1) + \beta_{24}x_2 + (\beta_{25} + \beta_{23}x_2)g_2^\psi(\bar{x}_2),$$
$$\phi_1^\psi(x_1) = Q_1^\psi(x_1, g^\psi(x_1)) = \beta_{11}x_1 + (\beta_{12} + x_1)g_1^\psi(x_1).$$

To incorporate the posterior sampling component, it is necessary to additionally weight by π when fitting models in equations (5.8) and (5.9) so that the estimated βs are dependant on π. This can be done through the `weights` argument in the `lm` function. These outcomes may then be used in equation (5.7) to obtain an estimate of the value under adherence to a DTR g^ψ. Over repeated draws of π, $E_{g^\psi}[Y^*|\bar{b}]$ and its associated uncertainty can be computed. Having computed these estimates of the value for all candidate regimes in $\mathcal{I}_{grid}$, the value-maximizing regime is selected as optimal.

5.2.4 Identifying Optimal DTRs via Gaussian Process Emulation

As discussed in the preceding sections, there are several value-search approaches to identifying optimal DTRs. The value surface can be modeled directly via a dynamic MSM and consequently maximized or a grid-search can be employed in order to identify the optimal regime. Directly modeling the value surface with a dynamic MSM can yield accurate, interpretable results, but this is only guaranteed when the value surface is correctly specified;

for example, incorrectly specifying a quadratic MSM can lead to inadequate inference about optimal regimes if the relationship is not in fact quadratic or poorly approximated by such a function over the range of ψs considered. A grid-search also has limitations in that it may not robustly identify the optimal regime, especially when the estimator used exhibits higher variability in some regions of the decision space than in others or when the value surface is multi-modal. In addition, the grid-search is not a particularly efficient approach as it requires many estimator evaluations, which may be computationally burdensome, especially in Bayesian settings where posterior predictive quantities must be computed. An important question that arises from these considerations is whether these limitations can be avoided by alternate methods.

One approach recently explored by Rodriguez Duque et al. [2022a] is to make use of computer experiments to identify optimal DTRs. The term "computer experiment" refers to the idea of sampling function values at strategically chosen points in order to approximate the function, with a limited number of samples. In a DTR context, this involves considering a DTR family, indexed by $\psi \in \mathcal{I}$, selecting an initial set of design points in $\mathcal{I}$, and using an estimator for the value of a DTR at these points. With a working model for the value surface, more points can be selected sequentially using a criterion that specifies where an optimum may be. Traditional approaches for computer experiments use regression-based methods to approximate a response surface of interest, like the value surface. However, these approaches have been critiqued, for example, by Huang et al. [2006] who emphasize that regression models are often too simple and unlikely to well-represent complex systems over the entire domain. This critique is analogous to the concerns that arise when using smoothly modeled MSMs to identify optimal DTRs.

Contemporary literature on computer experiments focuses on using $\mathcal{GP}$s to approximate complex functions and to identify optimizing points [Santner et al., 2018]. A $\mathcal{GP}$ is a stochastic process where any finite collection of variables in the process has multivariate Normal distribution. Much of the computer experiments literature has centered around settings in which the function to be maximized is known. However, it should be apparent that this is not the scenario under consideration here in the DTR context. In particular, an analyst wishing to perform a DTR analysis does not have access to direct observations of the value function; they have access only to a noisy, estimated version of the value function. With this nuanced difference in mind, some further considerations are required.

In order to better understand the problem characteristics, some terminology regarding the functional relationships in the problem should be set. The target of inference is the *value surface* which represents the relationship between a DTR g^{ψ} idexed by ψ and its value $E_{g^{\psi}}[Y]$. As the value surface is not accessible, it must be approximated via the *estimation surface*, a surface that results from point-wise evaluation of an estimator to obtain $\hat{E}_{g^{\psi}}[Y]$ for varying $\psi \in \mathcal{I}$. Evaluating the estimation surface on a fine grid is not desirable as not all points on the grid provide the same information about the optimal DTR's location. It would be beneficial to have a sample where each data point provides a high level of information toward identifying the optimizing point. Consequently, the aim is to use a restricted number of points from the estimation surface to produce an *emulation surface* which represents posterior belief about the value surface based on the information gathered from the estimation surface, with the goal of performing fewer evaluations than would be needed for the grid-search approaches of sections 2.2 and 2.3. As will be clarified in what follows, this posterior belief will be represented by a $\mathcal{GP}$.

Another consideration is that belief about the value surface should emphasize some smoothness, however the estimation surface used to infer about the value surface is not smooth. This is because it is the result of point-wise evaluations of an estimator which utilizes a finite sample to generate an estimate. Recent work concludes that this non-smooth or noisy quality may be heteroskedastic and consequently an inferential approach that accounts for this characteristic may be desirable [Rodriguez Duque et al., 2022a]. Authors in Rodriguez Duque et al. [2022a] examine some methods that allow for optimization via $\mathcal{GP}$, while accounting for the noise structure. They find that a homoskedastic treatment of the problem yields improved results over a $\mathcal{GP}$ method that does not account for noise or a grid-search, while providing comparable results to an approach that allows for heteroskedastic noise and that is more computationally intensive. The implementation in the `BayesDTR` package focuses on a homoskedastic treatment of the noise structure in order to perform optimization; we review this here. The estimation process begins by positing that the estimation surface is a noisy version of the value surface which is denoted by $f(\psi)$:

$$\upsilon_i = f(\psi_i) + \epsilon_i \ , \ \epsilon_i \sim N(0, \gamma^2), \ i = 1, ..., m, \tag{5.10}$$

with m being the number of observed points on the estimation surface. Using a Bayesian non-parametric framework, a $\mathcal{GP}$ prior is placed on f; this prior allows for f to belong to a broad class of continuous functions. Practically, this means that for any ψ , $f|\psi$ is $N(\mu_0, K)$ with covariance matrix K computed via a covariance function $k(\psi_i, \psi_j)$ and parameterized by $\eta_f = (\theta_f, \sigma_f^2)$. θ_f is a vector, where entries θ_{fd} control the correlation between points in the dth dimension; σ_f^2 scales the correlation function to yield the covariance. Bayesian formulations of this problem have been advocated for by O'Hagan et al. [1999], who emphasize that uncertainty in f is not solely aleatory. For example, in a setting where $\gamma^2 = 0$, f is a "knowable" function in the sense that it can be evaluated at different values of ψ. However, as it has not been evaluated at all values, there is uncertainty about the function's values in the locations where it has not yet been observed. This uncertainty is not sampling uncertainty arising from the variability in output under a sequence of identical experiments. Prior to continuing, some further notation should be defined, recalling that in this problem the units of observation are now sample points from the estimation surface, not sample points $(\bar{x}, \bar{z}, y)$ relating to patient information which are fixed at a sample size n. In this problem, data are observed as $\mathcal{D} = \{\psi_i, \upsilon_i\}_{i=1}^m$, and the following vectors are defined

$\psi = (\psi_1, ..., \psi_m)^T$, $\upsilon = (\upsilon_1, ..., \upsilon_m)^T$ and $f = (f_1, ..., f_m)^T$. Recall that ψ_i is the regime index for the ith regime (sample point) in the sample and that it could be a vector quantity.

Assuming known hyperparameters, the posterior distribution for the value of a new observation ψ_{m+1} is given by:

$$\begin{aligned} f^*_{m+1}|\psi_{m+1}, \eta_f, \gamma^2, \mathcal{D} &\sim N(\mu_{f^*_{m+1}}, \sigma^2_{f^*_{m+1}}) \\ \mu_{f^*_{m+1}} &= \mu_0 + k^T(K + \gamma^2 I_m)^{-1}(\upsilon - \mu_{0f}) \\ \sigma^2_{f^*_{m+1}} &= k(\psi_{m+1}, \psi_{m+1}) - k^T_{m+1}(K + \gamma^2 I_m)^{-1} k_{m+1}, \end{aligned} \tag{5.11}$$

with k_{m+1} being the covariance vector between observed points ψ and the new point ψ_{m+1}. The posterior distribution for value of an observation on the noisy estimation surface is given by:

$$\begin{aligned} \upsilon^*_{m+1}|\psi_{m+1}, \eta_f, \gamma^2, \gamma^2_{m+1}, \mathcal{D} &\sim N(\mu_{\upsilon^*_{m+1}}, \sigma^2_{\upsilon^*_{m+1}}) \\ \mu_{\upsilon^*_{m+1}} &= \mu_{f^*_{m+1}} \\ \sigma^2_{\upsilon^*_{m+1}} &= k(\psi_{m+1}, \psi_{m+1}) - k^T_{m+1}(K + S)^{-1} k_{m+1} + \gamma^2. \end{aligned} \tag{5.12}$$

In an empirical Bayes framework the posterior predictive distribution is given by $p(\upsilon^*_{m+1}|\psi_{m+1}, \mathcal{D}) = p(\upsilon^*_{m+1}|\psi_{m+1}, \eta_f, \gamma^2, \mathcal{D})$, meaning the parameters are assumed known even though they must be estimated in practice. These are estimated by maximizing the likelihood $p(\upsilon|\psi, \eta_f, \gamma^2)$. This maximization is performed in the `BayesDTR` package, using the concentrated likelihood discussed in Roustant et al. [2012] and Park and Baek [2001]. The concentrated likelihood is obtained by plugging-in estimated parameters that have maximum likelihood estimates with analytic expressions. We clarify this in what follows, but first it must be noted that these likelihoods are not always easy to maximize, even with gradient methods, so it is advisable to perform the maximization with several random starting locations as is made possible with the `DesignFit` function to be discussed in later sections.

In order to maximize the likelihood, the covariance function must first be specified. Common choices for the covariance functions, which yield smooth sample paths, are the $Mat\acute{e}rn_{3/2}$ and $Mat\acute{e}rn_{5/2}$ covariances. The $Mat\acute{e}rn_{3/2}$ covariance function between two regime indices ψ_i, ψ_j is given by:

$$k(\psi_i, \psi_j) = \sigma_f^2 \prod_{d=1}^{D} \left(1 + \frac{\sqrt{3}|\psi_{id} - \psi_{jd}|}{\theta_{fd}}\right) \exp\left(\frac{-\sqrt{3}|\psi_{id} - \psi_{jd}|}{\theta_{fd}}\right),$$

where D is the dimension of ψ and ψ_{id} and θ_{fd} are the dth entries in the ψ_i and θ_f vectors, respectively. This product emphasizes the point that different candidate rules in $\mathcal{G}$ should have the same dimension, and each entry in the index should represent the same rule element.

Although empirical Bayes requires maximizing a likelihood dependent on parameters μ_0, η_f, γ^2, the maximization is more efficiently performed by changing the parameterization. We now provide this new parameterization; full details of this parameterization can be found in Roustant et al. [2012]. By defining $\alpha = \sigma_f^2/(\sigma_f^2 + \gamma^2)$ and considering the correlation matrix R defined by $K = \sigma_f^2 R$, $(K + \gamma^2 I_m)$ can be re-expressed as $v(\alpha R + (1-\alpha)I_m)$, where $v = (\sigma_f^2 + \gamma^2)$. This re-parameterization results in a likelihood dependent on $\mu_{0f}, \theta_f, v, \alpha$, whereas the likelihood in the original parameterization dependended on $\mu_{0f}, \theta_f, \sigma_f^2$, and γ^2. As there are analytic expressions for the optimal μ_{0f} and v, the user only needs to concentrate on the maximization in the θ_f and α directions.

Priors for θ_f can be incorporated independently for each dimension d, for example, via a Log-Normal prior distribution with parameters μ_d, σ_d^2, which can be used to express belief about the size of θ_{fd}s and consequently the correlation between points. This prior is independent for each θ_{fd} and can be added into the log concentrated likelihood with the following term

$$\sum_{d=1}^{D} -\frac{(\log(\theta_{fd}) - \mu_d)^2}{2\sigma_d^2} - \log(\theta_{fd}\sigma_d\sqrt{2\pi}). \tag{5.13}$$

Maximizing the concentrated log likelihood with the added term above amounts to maximum a posteriori inference, where parameter estimates are fixed at the maximizers of the posterior distribution. One approach to setting the prior hyperparametrs is to identify what a 10% change is in the direction of interest. Then, the hyperparameters should be chosen such that the 5th and 95th percentiles of the Log-Normal distribution yield a correlation between 0.05 and 0.95. This posits that in the direction of interest, a unit change of 10% of the range of values will have function outputs that can either be very different from each other or very similar. This is similar to Lizotte [2008], who set hyperparameter values that prevent the θ_{fd}s from getting very small or very large, thereby preventing that function's value from being nearly exactly correlated or uncorrelated.

Sequential Sampling and Stopping Considerations

Recall the desired experimental setup: an initial set of design points are obtained and an initial model is fit on these data. This model, together with a rule for sampling additional points is used to identify new points that are most informative about the optimization process. The rule used to identify new points to sample is generally termed an infill criterion, and a review of possible criteria for stochastic computer experiments can be found in Picheny et al. [2013]. The focus here lies in using the well-known expected improvement criterion [Jones et al., 1998] as the infill criterion. In the deterministic setting, Frazier and Wang [2016] mention that this criterion benefits from a result that states that the true optimum will be identified as the number of experimental points increases, as shown by Locatelli [1997]; this is not guaranteed in the stochastic setting, as uncertainty remains in already observed points, thus requiring for some adaptations.

One solution for this, proposed by Forrester et al. [2006], is to use a re-interpolation approach. To perform the re-interpolation, the mean $\hat{\upsilon}_m = E[\upsilon^*_m|\psi_m, \mathcal{D}]$ is computed for each observed data point; this results in a new dataset $\mathcal{D}' = \{\psi_i, \hat{\upsilon}_i\}_{i=1}^m$. A $\mathcal{GP}$ can then be fit on these new data, assuming there is no noise, ϵ, in the process. The resulting $\mathcal{GP}$ has the property that there is zero uncertainty at already sampled points, thereby allowing for the use of the expected improvement criterion as a basis for sequential sampling. When the objective is maximization, this criterion is given by $EI(\psi) = E\left[\max(0, \upsilon(\psi) - \upsilon_{max})|\mathcal{D}'\right]$, with $\upsilon_{max} = max(\upsilon_1, ..., \upsilon_m)$. It is important to understand what this criterion means, in order to understand why it should be maximized to identify new points to add to the sample. At a point ψ^{new} where υ_{max} is expected to be greater than $\upsilon(\psi^{new})$, this criterion is zero. At a point ψ^{new} where $\upsilon(\psi^{new})$ is expected to be greater than υ_{max}, this criterion is large, with magnitude increasing with the difference in values. Therefore, maximizing this criterion adds points to the sample that are believed to have a higher value than the currently observed maximizer. Importantly, the expectation is taken with respect to the posterior distribution and it can be further developed to yield the well-known formula:

$$EI(\psi) = (\mu_{\upsilon^*_{m+1}}(\psi) - \upsilon_{max})\Phi\left(\frac{\mu_{\upsilon^*_{m+1}}(\psi) - \upsilon_{max}}{\sigma_{\upsilon^*_{m+1}}(\psi)}\right) + \sigma_{\upsilon^*_{m+1}}(\psi)\dot{\Phi}\left(\frac{\mu_{\upsilon^*_{m+1}}(\psi) - \upsilon_{max}}{\sigma_{\upsilon^*_{m+1}}(\psi)}\right) \quad (5.14)$$

when $\sigma_{\upsilon_{m+1}}(\psi) > 0$ and 0 otherwise. Φ is the CDF of the Standard Normal distribution and $\dot{\Phi}$ is the pdf.

As the expected improvement is zero at each visited point, it is clear that the function exhibits multi-modality. Maximization of this function can be performed via a genetic algorithm, as is done in Roustant et al. [2012], and implemented by Mebane, Jr. and Sekhon [2011] with the `rgenoud` package in `R`.

Finally, one natural question that arises is when to stop sampling. One approach may be to stop sampling when the expected improvement at newly sampled points plateaus near zero. Another approach, which we utilize in the illustrative example in section 4, is to plot the newly sampled points in order of sampling, to determine if sampling has converged around a specific region, which may suggest that the algorithm is sampling in a region where it believes the optimum to be.

Uncertainty Quantification and Fitting Procedure

One important element that should be addressed is the quantification of uncertainty in the estimated optimal DTR. In a grid-search, uncertainty in the optimum can be estimated via the Bayesian bootstrap. However this can be computationally intractable if the estimator employed arises from a posterior distribution with no analytic expression for the mean, as this requires complex computation for each bootstrap sample. Furthermore, bootstrapping the grid-search does not quantify uncertainty arising from the grid size selected. Certainly a coarse grid should have a different level of uncertainty about the optimizer than a fine grid, however, it is not clear how to quantify this.

With the $\mathcal{GP}$ approach presented, a Bayesian bootstrapping scheme can also be used to quantify sampling uncertainty. It can be further combined with the posterior uncertainty which represents uncertainty in the value function after having sampled m points form the estimation surface. For example, for each bootstrapped sample, N sample paths can be obtained from the posterior distribution and the optimum identified for each of these sample paths. Over bootstrapped samples, the resulting distribution of optima is reflective of both uncertainties. In what follows, we will examine how to quantify uncertainty in this manner with the `BayesDTR` package. This is of course a computationally intensive procedure.

In Algorithm 2, we provide a full description of how to identify optimal DTRs with the discussed $\mathcal{GP}$ methodology.

5.2.5 Individualized Inference

The Bayesian methods discussed so far permit individualized inference. This is best understood via an example. Consider the regime "treat if $x > \psi$" and suppose that a new patient is observed with covariate value x^{new}. Interest lies in deciding whether this patient should receive treatment, based on what is known about the optimal threshold, ψ_{opt}. This involves computing $P(x^{new} > \psi^*_{opt}|\bar{b})$ by taking a sample of size m from the posterior distribution of

```
/* First obtain point estimates for $\psi_{opt}$ */
Estimate value, $\tilde{y}^{\psi} := E_{g^{\psi}}[Y]$, at experimental points $\mathcal{P} = \{\psi_1, ..., \psi_m\}$
Estimate $\mathcal{GP}$ parameters
Perform re-interpolation as in Forrester et al. [2006]
do
    Sample new point by solving $\psi^{new} = \arg\max_{\psi}\{EI(\psi); \psi \in \mathcal{I}\}$
    Estimate value at $\psi^{new}$
    Add $\psi^{new}$ to experimental points: $\mathcal{P} = \{\psi^{new}\} \cup \mathcal{P}$
    Identify $\psi^{opt} = \arg\max_{\psi}\{\tilde{y}^{\psi}; \psi \in \mathcal{P}\}$
    Estimate $\mathcal{GP}$ parameters and perform re-interpolation;
while Not converged                    // Assess convergence as in section 2.4.1
Set $m_+ = |\mathcal{D}|$                              // Now have point estimate for $\psi_{opt}$

/* Now computing variability around optimal thresholds */
for $i \leftarrow 1$ to $B$ do                 // B is number of Bayesian bootstrap draws
    Draw $\pi = (\pi_1, ..., \pi_n)$ from $Dir(1, ..., 1)$
    /* Estimates, $\tilde{y}^{\psi}(\pi)$, now depend on $\pi$ as in section 2.2 and 2.3 */
    As above, sequentially sample points by maximizing $EI(\psi)$ and updating $\mathcal{GP}$
      parameters
    Stop sampling when total of $m_+$ experimental points are in $\mathcal{P}$
    Draw $N$ sample paths from posterior $\mathcal{GP}$
    Compute optimizer for each sampled path
    Store vector of length $N$, containing $N$ optimizers
end
Output: Vector of length $N \cdot B$ containing posterior distribution of $\psi^{opt}$
```

Algorithm 2: Finding optimal DTRs using $\mathcal{GP}$ emulation.

ψ^*_{opt} and computing $p = (1/m)\sum_{\psi} \mathbb{1}(x^{new} > \psi^*_i)$. Given uncertainty in ψ_{opt}, this measure informs a decision maker about the probability that x^{new} is above the true optimal threshold. Effectively, then, it provides evidence for whether the patient should receive treatment if the optimal regime is to be followed. This approach is relevant to all types of decision rules. We will see in the illustrative example how to implement this individualized inference about the treatment decision.

5.2.6 Frequentist and Normalized Estimators

The Bayesian approaches discussed in sections 2.1-2.3 all have frequentist counterparts. Point estimates for the quantities of interest can be arrived at in a straightforward manner. For the dynamic MSMs in section 2.1, it is necessary that π_i for $i = 1, ..., n$ be removed from equation (5.4). Solving this new equation will yield the frequentist point estimates. For the IPW method, it is required that the expectation in equation (5.5) be computed to yield $\sum_{i=1}^{n} \frac{1}{n} w_i^{\psi} y_i$, as $E_{\pi}[\pi_i] = 1/n$. For the doubly robust approach, it is required that the π_i in equation (5.7) be replaced with $1/n$. Treatment models are now fit without any dependence on π.

In practice, using estimators with less variability can improve the resulting inference. In the case of the IPW and DR estimators, it is clear that reducing the variability in the weights will reduce the variability in the estimator. This may be achieved via normalized weights. Weight normalization is discussed in Hernán and Robins [2020], and has been explored in Xiao et al. [2010], as a means of reducing variability in weighted estimators. In this Bayesian setting, there is a contribution to the weights from the importance sampling weights and from the Dirichlet weights. For each sample of Dirichlet weights $\pi = (\pi_1, ..., \pi_n)$, the normalized weights can be defined as:

$$\bar{w}_{ik}^{\psi} = \frac{\dfrac{\pi_i \mathbb{1}_{\bar{g}_k^{\psi}(\bar{x}_{ik})}(\bar{z}_{ik}) y_i}{\prod_{j=1}^{k} p_{\mathcal{O}}(z_{ij}|\bar{z}_{ij-1}, \bar{x}_{ij})}}{\displaystyle\sum_{i=1}^{n} \dfrac{\pi_i \mathbb{1}_{\bar{g}_k^{\psi}(\bar{x}_{ik})}(\bar{z}_{ik})}{\prod_{j=1}^{k} p_{\mathcal{O}}(z_{ij}|\bar{z}_{ij-1}, \bar{x}_{ij})}}, k = 1, ..., K. \tag{5.15}$$

Taking the expectation in the numerator and the denominator across π yields the normalized weights that could be used in a frequentist analysis as in Hernán and Robins [2020]. Replacing $\pi_i \bar{w}_{ik}^{\psi}$ in the IPW or DR estimator by the weight in equation (5.15) yields the normalized estimators.

5.3 Implementation

In this section, we examine the functions in the `BayesDTR` package that can be used to carry out inference with the methods described previously. We first examine the `BayesMSM` function, which permits identification of optimal DTRs using Bayesian dynamic MSMs, IPW, and doubly robust estimators. We then focus our attention on the DesignFit and SequenceFit functions which perform estimation using the $\mathcal{GP}$ methodology of section 2.4. We also examine the FitInfer function which allows for the quantification of uncertainty in the optimal DTR when using $\mathcal{GP}$s.

5.3.1 Functions to Identify Optimal DTRs using Bayesian Dynamic MSMs, IPW, and Doubly Robust Estimators

The following code provides the syntax required to use the BayesMSM function. The BayesMSM function has three distinct functionalities: I) to infer about the parameters of a dynamic MSM via IPW, II) to estimate the value of a grid of regimes via IPW, and III) to estimate the value of a discrete set of regimes via the doubly robust estimator.

```r
#loading BayesDTR package
library(BayesDTR)
#Basic parameters in the BayesMSM function
BayesMSM(PatID,Data,Outcome_Var,Treat_Vars,Treat_M_List,Outcome_M_List,MSM_Model,
        G_List,Psi,Bayes=TRUE,DR=FALSE,Normalized=FALSE,B=100,Bayes_Seed=1)
```

PatID and Data allow users to supply an analysis dataset and to indicate the patient identifier. Note that the analytic dataset should contain only one row per patient. Outcome_Var is a character variable specifying the final-stage outcome, and Treat_Vars is a character vector specifying the stage-specific treatment variables. Treatment variables should be coded as {0, 1}. Treat_M_List and Outcome_M_List are lists containing the formulas for the treatment and outcome models, depending on which estimator is being used. In each list, there should be as many formulas as treatment decision points and they should be ordered chronologically. A formula for the MSM of interest can be supplied via the MSM_Model parameter, if the aim is to make use of functionality I.

The next set of parameters are those relevant to the family of dynamic regime of interest.

The G_List variable allows the user to define the stage-specific decision rules of interest. The Psi parameter is a matrix specifying the DTR index grid that will be used to create an augmented dataset if using a dynamic MSM or to perform a grid-search if directly using an estimator for the value. In Psi, there should be one column per regime index coordinate and it is necessary that the column names match the names in the regime indices provided to G_List. For example if at stage one, the regime of interest is "treat when psi_1>x", then Psi should contain a column named psi_1. The rows of Psi corresponds to a single point in the grid. The function can handle decision rules that involve one of five comparison operators per stage ==,>,<,>=,<= and these can be put together with logical operators &,|. On either side of the comparison operator, there can be parameters that index the family of regimes, or there can be tailoring covariates. The parameters and tailoring covariates can appear in the same expression with the usual mathematical operators for example as given by the rule "treat when psi_1*x_1+psi_2*x_2>0". Lastly, the Bayes, DR, and Normalized parameters indicate whether a Bayesian or frequentist analysis should be carried out, whether or not the DR estimator should be used, and whether weights should be normalized or not. B allows the user to indicate the number of Bayesian bootstrapped samples to perform when Bayes=TRUE. The default fit for this function is to use a grid-search with the IPW estimator.

The Bayesian analysis returns a matrix containing the posterior distribution of interest. For functionality I, there are as many columns as terms in the MSM_Model formula, and the number of rows is equal to B. Each matrix column represents a sample from the posterior distribution for a regression coefficient. For functionalities II and III, columns in the matrix represent points in the grid of Psi and the rows, like in functionality I, represent distinct posterior draws. The frequentist analysis returns point estimates, with columns representing the same parameters as in the Bayesian analysis. If the user is interested in providing a measure of variability for the frequentist estimates, the non-parametric bootstrap can be used by calling the BayesMSM function within the bootstrapping function in the boot package. The illustrative example in Section 4 will provide clarity as to the required format that variables should be provided in and how to perform each of the analyses of interest. Additionally, a systematic description of required and optional parameters for each functionality is provided in Appendix C.1.

5.3.2 Functions to Identify Optimal DTRs using Gaussian Processes Emulation

We now examine the syntax for functions in the BayesDTR package used to identify optimal DTRs using $\mathcal{GP}s$. The first function we examine is the DesignFit function which allows us to fit a $\mathcal{GP}$ model on an initial set of design points.

```
DesignFit(PatID,Data,Outcome_Var,Treat_Vars,Treat_M_List,Outcome_M_List,
        G_List,Psi,Normalized=TRUE,DR=FALSE,
        Numbr_Samp,IthetasU,IthetasL,Covtype,
        Likelihood_Limits,Prior_List=NA, Prior_Der_List=NA)
```

The parameters used on the first two lines of the DesignFit function above are those already introduced with the BayesMSM function. In particular, these allow the user to utilize the frequentist IPW or DR estimator to produce the estimation surface. Note that the MSM_Model parameter should not be used in this application, as the $\mathcal{GP}$ only makes use of the IPW or DR estimator for the value of a single regime at a time. If a value for this parameter is passed to the function, it will be ignored and the default normalized IPW estimator will be used. Other required parameters used by the function include Numbr_Samp which specifies the number of random starts when optimizing the Gaussian likelihood and Covtype which specified whether the $Mat\acute{e}rn_{3/2}$ (Covtype=1) or $Mat\acute{e}rn_{5/2}$ (Covtype=2) covariance functions will be used. The user should also provide the limits for the parameter coordinates in θ_f in the likelihood via IthetasL, IthetasU, where L stands for the lower bound of the parameter and U stands for the upper bound. Placing bounds in the optimization is useful, otherwise the gradient optimizers may explore a region of the parameter space that yields a non-invertible covariance matrix, thereby interrupting the optimization procedure. As we are interested in assessing whether the model is being fit correctly, the Likelihood_Limits parameter is a list of vectors allowing the user to set the limits for plotting the likelihood for the θ_f coordinates. Each list element is a vector containing the lower and upper bound for each covariance parameter.

Independent priors can also be placed on these parameters by defining the optional parameters Prior_List and Prior_Der_List. These are lists containing the formula for the log prior distributions and for the derivatives of the prior. Importantly, a specific naming convention for elements of these lists should be maintained. For example, in a two-stage setting, the θ_{f1} parameter should be represented by theta1 and the θ_{f2} parameter should be represented by theta2. Adding more dimensions to the problem simply requires adding more elements to the list and maintaining the naming convention. Appendix C.2 provides a systematic description of which parameters are required and which are optional. Note that by default, these optimization functions identify a DTR that maximizes the value function. If the objective is to minimize the value function, users should supply the negative of the outcome variable to Outcome_Var and allow the function to maximize the value.

The DesignFit function returns a list of several important parameters. In particular, it returns an Update list containing information about the updated $\mathcal{GP}$ fit as well as a ReInter list containing the x_max_ri and Y_max_ri values corresponding to the optimal regime index and value identified with the currently available experimental points. The function also returns the parameter values related to the estimated hyperparameters, these can be found in the thetas and alpha parameters.

Now we explore how to sequentially sample additional points using the SequenceFit function which identifies new points to sample by maximizing the expected improvement and then re-estimates the $\mathcal{GP}$ parameters based on the new information.

```
SequenceFit(Previous_Fit,Additional_Samp,
        Control_Genoud=list(Domain=matrix(c(200,200,500,500),ncol=2)))
```

All parameters in the SequenceFit function are required with the first being the Previous_Fit parameter which stores an object returned by either the DesignFit function or the SequenceFit function. Being able to supply an object returned by the SequenceFit function is important as we may want to continue sampling sequentially even after we have called this function once. Effectively, all options in the object passed to the Previous_Fit parameter are inherited in the SequenceFit function. The additional_samp parameter allows the user to tell the function how many additional samples to take sequentially. The last parameter in the function is relevant to the optimization of the expected improvement via the genetic algorithm as implemented by the genoud function. This is the Control_Genoud parameter which is a list of parameters to be passed to the genoud function. Importantly, the only required parameter to be passed to the genoud function is the Domain parameter which carries information about the domain in which optimization will be performed; it should be a matrix with number of columns equal to the dimension of the regime index, and with columns indicating the lower and upper boundary in each dimension. The SequenceFit function returns the same object as the DesignFit function, with the addition of the EI_hist parameter. This parameter contains the expected improvement value at each of the sequentially sampled points.

One option that may be of interest to a user is to compute the posterior mean after arriving at a $\mathcal{GP}$ fit. This can be done via the PostMean function.

```
PostMean(X,GP_Object)
```

This function only requires that an object returned from the DesignFit or SequenceFit functions be supplied to GP_Object, in addition to a parameter X specifying a coordinate at which to evaluate the mean.

Lastly, as discussed in section 2.4.2, it may be important to provide a measure of uncertainty when identifying the optimal DTR. The function FitInfer allows the user to do this.

```
FitInfer(Design_Object,Boot_Start,Boot_End,N,Psi_new,Location,Additional_Samp)
```

This function requires a Design_Object parameter, which is the object returned by the DesignFit function. The Boot_Start and Boot_Stop parameters allow the user to specify the number of Bayesian bootstrapped samples, for example from Boot_Start=1 to Boot_End=100, all while allowing for reproducibility as each bootstrapped sample is linked to a specific seed for random number generation. If we were interested in reproducing only bootstrap number 50, we could set Boot_Start=50 and Boot_End=50, and run the function. Furthermore, the N parameter tells the function how many sample paths to obtain from the posterior $\mathcal{GP}$ at each bootstrapped sample. The Additional_Samp parameter is the same as that in SequenceFit function, and the Psi_new parameter is the grid of points for which to search for an optimum in each sampled path drawn from the posterior $\mathcal{GP}$. It also determines the dimension of the covariance matrix used to generate the sampled paths, so a very fine grid may be computationally intractable. The only optional parameter in this function is the Location parameter which allows the user to specify where the output of the function should be saved.

This function returns a matrix with number of columns equal to the number of regime index elements plus one. The last column corresponding to the optimal value, and the prior columns correspond to the estimated optimal index. The number of rows in the matrix is N(Boot_Start-Boot_End+1), as for each bootstrapped sample there are N posterior paths sampled and an optimum identified for each of these paths.

As we are dealing with a $\mathcal{GP}$, which depends on a covariance matrix that needs to be inverted, numerical issues may arise. For example, when using the SeqFit function to sequentially sample points, users should take care to check that the sampling has not focused on a very specific region. If it has, this is evidence for convergence of the algorithm and may lead to a non-invertible covariance matrix if too many points are sampled in the same region. This non-invertability arises as nearby points can exhibit nearly perfect correlation. This issue can be carried downstream to the quantification of uncertainty if convergence for some bootstrapped samples is achieved faster thereby possibly yielding non-invertible matrices. Issues with non-invertibility are mainly numerical; conceivably, given enough precision in the matrix computation, matrices would be invertible.

5.4 Illustrative Example with the BayesDTR Package

For illustrative purposes, we adapt data from Hammer et al. [1996] to demonstrate how the discussed methods may be applied with the `BayesDTR` package. These data originate from a double-blinded randomized trial performed to compare treatments using single and double nucleosides as a means of treating HIV type 1. Focus is given to patients' CD4 cell count which provides a measure of the health of patients' immune system, with higher values indicating better health. Study enrollment required patients to have CD4 cell counts between 200 and 500 cells/mm^3. A total of 2467 patients were assigned to daily doses of one of four treatments 1) 600 mg of zidovudine, or 2) 600 mg of zidovudine & 400 mg of didanosine, or 3) 600 mg of zidovudine & 2.5 mg zalcitabine, or 4) 400 mg didanosine. Variables found in the dataset include patients' race, sex, baseline CD4, 20 week CD4, weight, age, history of antiretroviral therapy, symptoms of HIV infection, and Karnofsky score. These data may be accessed via the `LongCART` package in `R` [Kundu, 2021].

We restrict our analysis to the use of two dual-therapies, in order to determine which patients should be given zidovudine with zalcitabine versus zidovudine with didanosine, coded as 1 and 0, respectively. In particular, we examine whether tailoring therapy on baseline and 20 week CD4 cell counts yields improved 96 week CD4. As the original trial involved treatment assignment only once, we perform a plasmode simulation that randomly assigns an additional treatment decision point at 20 weeks. Variables for this analysis did not exhibit any missing values. There were 524 patients in the zidovudine & zalcitabine arm and 522 in the zidovudine & didanozine arm. The known stage-specific treatment probability was 0.5 by design, however we estimate these probabilities, as this can improve efficiency when using IPW estimators [Henmi and Eguchi, 2004]. As we added an additional treatment variable, we also simulate the final outcome which depends on $cd4.0$ and $cd4.20$ variables representing baseline and 20 week CD4 cell count, respectively, a sex variable that equals 1 for males and 0 for females, and treatment variables z_1 and z_2. This outcome is deterministically generated by:

$$y = max(0, 0.2(5cd4.0 + 6sex + (-3000 + 9cd4.0)z_1 + (-3000 + 9cd4.20)z_2)). \qquad (5.16)$$

For illustrative purposes, we allow y to represent the final outcome which we take to be 96 week CD4 cell count and the aim is to maximize this value. Without the $max()$ function in this data generating mechanism, a small proportion of values would be negative, which is not meaningful given that the outcomes represent a cell count. These adapted data can be found in the `BayesDTR` package via the `BayesDat` dataset. The specific regime that we explore is "at each stage, assign to zidovudine with zalcitabine if CD4 cell count is greater than ψ_k, for $k = 1, 2$". ψ_1 and ψ_2 are restricted to vary between 200 and 500 cells/mm^3. A regime like this may be of interest if a patient requires one therapy when CD4 cell counts are low and another therapy when CD4 cell counts are closer to stable levels. As we know the data-generating mechanism given in equation (5.16), we can compute the mean outcome under adherence to a specific regime g by setting $z_1 = g_1(x_1)$ and $z_2 = g_2(x_2)$ and plugging these values into the equation. Doing so for a fine grid of regime indices allows us to produce Figure 5.1, an approximation for the value function. Employing a grid-search with a grid of increments of five yields an optimal regime at $(\psi_1, \psi_2) = (335, 335)$ with an optimal value of 610 cells/mm^3. This matches very closely to the regime obtained theoretically, if we assume that the effect of truncating a small set of negative values at zero is small. With this assumption, the theoretical optimum can be approximated to be at $(333.3, 333.3)$. In what follows, we will compare this optimum, which is a very good approximation for the true optimal regime, with optima estimated via other methods.

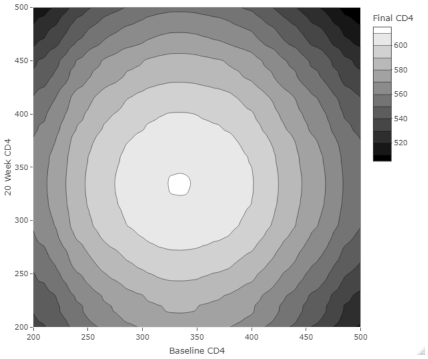


Figure 5.1: Value function for the rule "treat with zidovudine and zalcitabine when CD4 cell count is greater than ψ_k for $k = 1, 2$" found via Monte Carlo methods using the known data generating mechanism.

5.4.1 Bayesian MSM, IPW, and Doubly Robust Inference

We now examine how to define the parameters in the `BayesMSM` function in order to analyze these data using each of the three estimation approaches available in the function. In the treatment models, we include variables that may not have achieved balance by chance, even though in these data treatment was randomized. In the outcome models, we include these variables as well and additionally include the variables that interact with treatment and that therefore allow for tailoring. Of course, the outcome models are very slightly misspecified, due to the truncation at zero in the data-generating mechanism. However, we will see that this does not appear to have a serious impact on the results.

```
#identifying variables in dataset
Outcome_Var="cd4.outcome"
Treat_Vars=c("z1","z2")
PatID="pidnum"
#defining treatment models
Treat_M_List=list(tformula1="z1~karnof+race+gender+symptom+str2+cd4.0+wtkg",
        tformula2="z2~karnof+race+gender+symptom+str2+cd4.20+wtkg+z1")
#defining outcome models
Outcome_M_List=list(
  oformula1="Pseudo_Outcome~karnof+race+gender+symptom+str2+cd4.0+z1+cd4.0:z1",
  oformula2="cd4.outcome~karnof+race+gender+symptom+str2+cd4.0+cd4.20+z1+
     cd4.0:z1+z2+cd4.20:z2")
#defining stage specific decision rules
G_List=list(g1=expression(cd4.0>=psi1),
        g2=expression(cd4.20>=psi2))
#defining MSM model for when directly modeling the value function
MSM_Model="cd4.outcome~1+psi1+I(psi1^2)+psi2+I(psi2^2)"
```

It is necessary that in the lists defined above, the naming convention of the list elements be maintained (e.g., treatment models being named `tformula1`, `tformula2`, etc.). Furthermore, the formulas provided should follow the general conventions for formulas supplied to the `glm` function. Models in `Treat_M_List` are each fit using logistic regression with the `glm` function and with parameter `family="binomial"`; models in `Outcome_M_List` are each fit using the `lm` function as the outcomes are assumed to be continuous. The expressions in the `G_List` parameter should contain the conditions for receiving the treatment coded as 1. For functionality I, the target of inference is the coefficients associated with the terms in the model supplied to `MSM_Model`. Based on the model supplied above, there are five coefficients of interest, each corresponding to one of the terms supplied to `MSM_Model`. We refer to these coefficients as $\beta_0, ..., \beta_4$. We now provide code for calling the function when estimating optimal DTRs using each of the three estimation methods from section 2.1-2.3. First, we examine code relevant to directly modeling the value function by fitting a Bayesian dynamic MSM with IPW.

```
#defining grid for augmented dataset
Psi=as.matrix(expand.grid(seq(200,500,50),seq(200,500,50)))
colnames(Psi)=c("psi1","psi2")
#fitting quadratic MSM
QuadMSM=BayesMSM(Data=BayesDat,PatID=PatID,Outcome_Var=Outcome_Var,
        Treat_Vars=Treat_Vars,Treat_M_List=Treat_M_List,
        G_List=G_List,Psi=Psi,MSM_Model=MSM_Model,Bayes=TRUE,B=100)
```

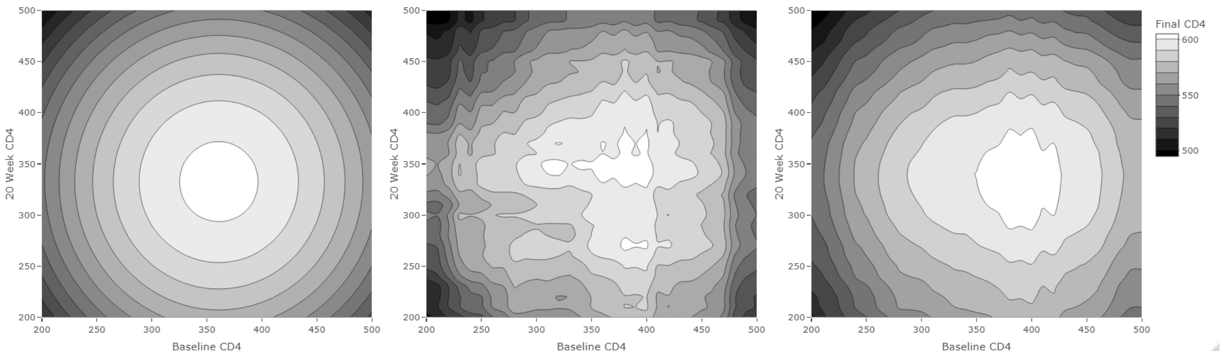
The code above defines the grid upon which to create the augmented dataset required for fitting a dynamic MSM, as outlined by Cain et al. [2010]. In this example, `psi1` and `psi2` index the family of regimes of interest and they match with the variable names defined in `G_List`. It is important that the column names of `Psi`, representing coordinates of the regime index, be named exactly as they appear in `G_List`; the function checks for this match in labels, and it will produce a warning if the names do not match. This function call also requires supplying the `MSM_Model` parameter and setting the `Bayes` parameter to `TRUE` in order to obtain the Bayesian estimator. Additionally, setting `B=100` returns 100 posterior draws of the parameters associated with the MSM supplied by `MSM_Model`. Note that the `Normalized` parameter cannot be used when the `MSM_Model` parameter is supplied. This function call returns a matrix with columns representing $\beta_0, ..., \beta_4$ corresponding to the MSM specified by `MSM_Model`. Rows in this matrix correspond to posterior draws of the βs. To identify the optimal regime, the quadratic function can be maximized for each posterior draw; this yields the posterior distribution of the optimum. We now examine how the `BayesMSM` function can be used to estimate the value of a grid of DTRs via the IPW or DR estimator, which are the second and third functionalities available with the `BayesMSM` function.

```
#defining grid for grid-search
Psi=as.matrix(expand.grid(seq(200,500,15),seq(200,500,15)))
colnames(Psi)=c("psi1","psi2")
#fitting IPW estimator to a discrete set of regimes
Grid_IPW=BayesMSM(Data=BayesDat,PatID=PatID,Outcome_Var=Outcome_Var,
        Treat_Vars=Treat_Vars,Treat_M_List=Treat_M_List,G_List=G_List,
        Psi=Psi,Bayes=TRUE,Normalized=TRUE,B=100)
#fitting DR estimator to a discrete set of regimes
Grid_DR=BayesMSM(Data=BayesDat,PatID=PatID,Outcome_Var=Outcome_Var,
  Treat_Vars=Treat_Vars,Treat_M_List=Treat_M_List,Outcome_M_List=Outcome_M_List,
  G_List=G_List,Psi=Psi,Bayes=TRUE,Normalized=TRUE,DR=TRUE,B=100)
```

In the code above, we first define the grid used for the grid-search. We then call the `BayesMSM` function to return the posterior samples using the IPW and DR estimators. The main difference between the two calls is that the DR approach needs the added parameter `Outcome_M_List` and setting `DR=TRUE`. For both of these estimation procedures, the `BayesMSM` function returns a matrix where each column represents a regime index in the same order as provided by the `Psi` parameter and where each row represents a single draw from the posterior distribution. `Normalized=TRUE` indicates that we are using normalized weights.

Based on the function calls above, we can estimate the mean value for each of the regimes in the grid; the result is given in Figure 5.2. We see that all methods agree about the general shape of the value function and that in this case both the doubly robust and MSM yield relatively smooth, interpretable surfaces. As with any posterior distribution, summary statistics can be provided. To obtain the posterior distribution for the optimal DTR, the regime index that yields the highest value should be identified for each posterior sample. Doing this across all posterior samples yields the posterior for the optimum. This is shown in the following code for the doubly robust analysis:

```r
#obtaining index of regime that maximizes value for each posterior sample
max_index=apply(Grid_DR,1,FUN=function(X){which(X==max(X))})
#obtaining posteriro distribution of value at optimum
max_val=apply(Grid_DR,1,max)
#obtaining posterior distribution for stage 1 and stage 2 optimal thresholds
Psi[max_index,]
```



(a) (b) (c)

Figure 5.2: Estimation surface for the rule "treat with zidovudine and zalcitabine when CD4 cell count is greater than ψ_k for $k = 1, 2$" using (a) quadratic MSM (b) normalized IPW grid-search (c) normalized DR grid-search.

Table 5.1 shows the posterior median and the 95% credible intervals for the optimal stage-specific threshold using each of the three described methods. We see that broadly all three methods agree regarding the location of the optimal thresholds. The doubly robust estimator is best at identifying the stage two optimal parameter whereas the quadratic MSM is best in identifying the first stage parameter. All methods seem to perform better at identifying the second stage parameter than the first stage parameter. We also see that the optimal DTR estimated by the DR estimator exhibits less variability than the IPW estimator, which is known to possess the most variability.

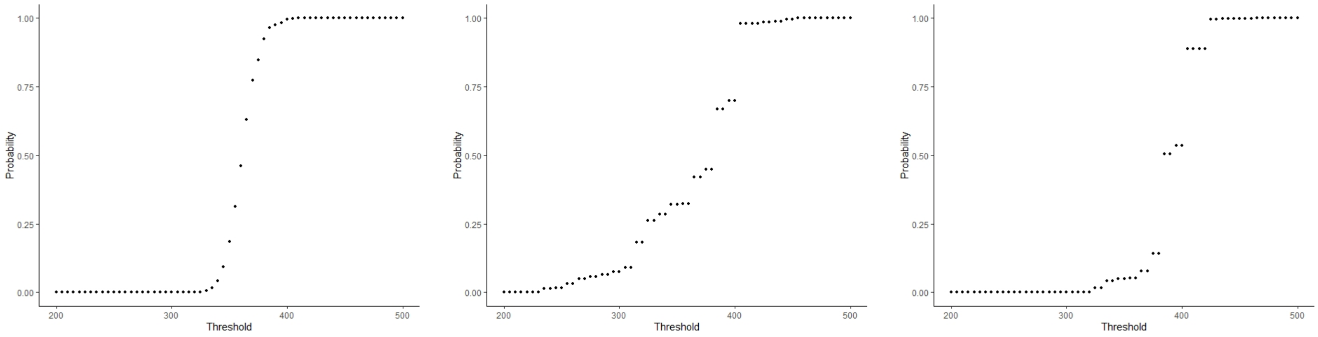
Table 5.1: Estimated optimal thresholds with 95% credible intervals for the rule "treat with zidovudine and zalcitabine when CD4 cell count is greater than ψ_k for $k = 1, 2$".

Method	$\hat{\psi}_{1opt}$	$\hat{\psi}_{2opt}$	Value at Optimum
Quadratic MSM	361.4 (337.1,389.2)	332.0 (306.7,357.4)	603.7 (569.3,642.7)
Normalized IPW grid-search	380 (250,400)	350 (260,400)	613.4 (578.8,652.1)
Normalized DR grid-search	380 (330,420)	340 (330,340)	607.5 (582.5,632.6)

Individualized inference can also be implemented. The code below illustrates how this is done for the first stage. First, the posterior distribution for ψ_{1opt} is computed, in this case we use the `Grid_DR` matrix returned from the `BayesMSM` function. Then, the probability that a patient's baseline CD4 cell count is greater than the optimal threshold is obtained. Below, we compute this probability for a range of CD4 values, `Psi1_Grid`,that correspond to new patients in order to obtain the `probas1` variable.

```
#computing posterior distribution for optimal index
max_index=apply(Grid_DR,1,FUN=function(X){which(X==max(X))})
#defining range of new cd4 cell counts
#this range is not for one patient but for a set of new patients with varied CD4s
Psi1_Grid=seq(200,500,5)
#computing the probability that a patient's baseline CD4 cell count is greater
#than the optimal stage 1 threshold
probas1=sapply(Psi1_Grid, FUN=function(X, max_index){mean(X>Psi[max_index,1])},
        max_index=max_index)
```

Having computed these probabilities, we can plot them to better visualize the uncertainty. Figure 5.3 shows the first stage probabilities associated with each of the estimation methods discussed and for a range of CD4 cell values that can correspond to newly seen patients. We see that the plot associated with the quadratic MSM is smoothest and displays the most certainty about the optimal treatment allocation, as evidenced by the narrow window of the threshold over which the probabilities are farther from 0 or 1.



(a)

(b)

(c)

Figure 5.3: Individualized optimal treatment allocation probabilities for the rule "treat with zidovudine and zalcitabine when CD4 cell count is greater than ψ_k for $k = 1, 2$" (a) quadratic MSM (b) normalized IPW grid-search (c) normalized DR grid-search.

5.4.2 Illustrative Example using Gaussian Processes

In this section, we continue using the case study presented to examine how to use the BayesDTR package to implement an analysis that uses $\mathcal{GP}$ emulation to identify an optimal DTR. We begin by fitting a $\mathcal{GP}$ model on an initial set of design points; the required code is given below. The initial set of design points in this setting is in Psi and is limited to 16.

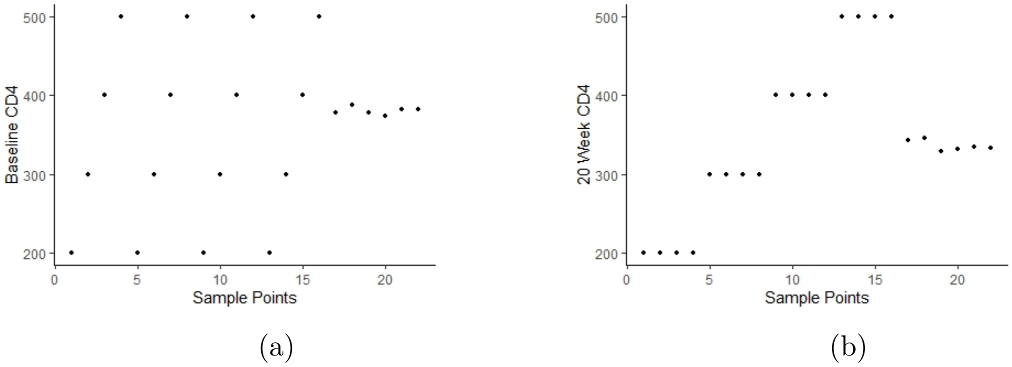
```r
#creating grid of initial design points
Psi=as.matrix(expand.grid(seq(200,500,100),seq(200,500,100)))
colnames(Psi)=c("psi1","psi2")

#fitting GP model on initial set of design points
start_fit=DesignFit(PatID=PatID,Data=BayesDat,Treat_M_List=Treat_M_List,
        Outcome_Var=Outcome_Var,Treat_Vars=Treat_Vars,G_List=G_List,Psi=Psi,
        Numbr_Samp=5,IthetasU=c(600,600),IthetasL=c(0.01,0.01),Covtype=2,
        Likelihood_Limits=list(seq(250,500,2), seq(250,500,2),
        Prior_List=NA, Prior_Der_List=NA))
```

With a $\mathcal{GP}$ process model being fit on an initial set of design points, the next step is to sample an additional set of experimental points by maximizing the expected improvement. This is done with the SequenceFit function. In this case we select an additional six points.

```r
#Updating model with newly sequentially sampled points
second_fit=SequenceFit(Previous_Fit=start_fit,Additional_Samp=6,
        Control_Genoud=list(Domain=matrix(c(200,200,500,500),ncol=2)))
```

Once additional samples have been obtained, the sample points can be plotted to examine whether the algorithm has focused sampling in a specific region, thereby providing evidence that an adequate maximizer has been identified. Figure 5.4 shows these plots. The first 16 points correspond to the design points, the remaining six points correspond to those sequentially sampled; we see that the sequentially sampled points have remained very much in the same area thereby providing evidence for convergence. Note that it does not matter what order the first 16 points are plotted in, as they were sampled simultaneously. Users should be cautious about how many additional points to sample in settings like this, as sampling points that are proximal to each other can result in non-invertible covariance matrices. Using the model fit of six additional points, we can determine the optimal thresholds and the value at the optimal thresholds. `Y_max_ri` gives the value at the optimum to be 601.8 cells/mm^3; the optimizer can be found with `x_max_ri` which is determined to be 381.6 cells/mm^3 and 334.4 cells/mm^3 for ψ_{1opt} and ψ_{2opt}, respectively. These estimates are similar to those obtained with other methods (See Table 1). We will see in what follows how uncertainty can be quantified.



(a) (b)

Figure 5.4: Design points with an additional six sequentially sampled points from estimation surface corresponding to the rule "treat with zidovudine and zalcitabine when CD4 cell count is greater than ψ_k for $k = 1, 2$" (a) stage 1 threshold (b) stage 2 threshold.

With convergence attained, the resulting posterior mean can be visualized. This can be done via the `PostMean` function provided below.

```
#creating grid
Psi=as.matrix(expand.grid(seq(200,500,10),seq(200,500,10)))
colnames(Psi)=c("psi1","psi2")
#computing posterior mean on grid of points
estimated_y=apply(Psi,1,FUN=PostMean,GP_Object=second_fit)
```

Evaluating the posterior mean on a grid of points yields Figure 5.5. We see that it broadly resembles the other value surfaces (see Figure 2).

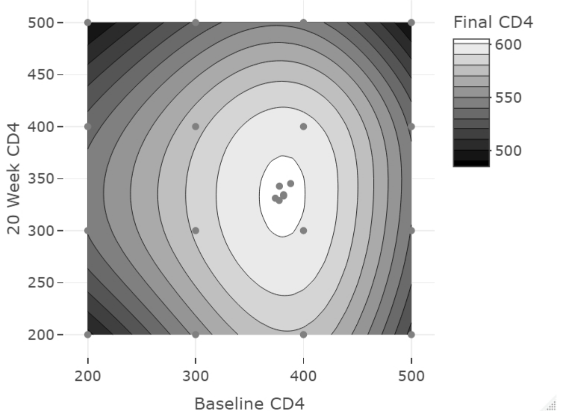


Figure 5.5: Emulated surface after six sequentially sampled points corresponding to the rule "treat with zidovudine and zalcitabine when CD4 cell count is greater than ψ_k for $k = 1, 2$".

The last step in the inferential process is to use the `FitInfer` function to quantify the uncertainty around the optimizers. Based on the convergence plots above, we chose to perform inference at 6 additionally sampled points. We perform 100 bootstraps, with 100 sampled paths in each bootstrap.

```
#defining additional parameter for this function
Location="posterior_sample.csv"
#computing uncertainty around optimal DTR
Variability_Matrix=FitInfer(Design_Object=start_fit,Boot_Start=1,Boot_End=100,
                   N=100,Psi_new=Psi_new,Location=Location,Additional_Samp=6)
```

Using the output of the `FitInfer` function, we compute the median and 95% credible interval of the quantities of interest, and we obtain that ψ_{1opt} is 380 cells/mm^3 (335,410), ψ_{2opt} 335 cells/mm^3 (245,395), and the value at the optimum is 607 cells/mm^3 (569,646). These estimates broadly match the results obtained with the grid-search and direct modeling approaches. The credible intervals exhibit slightly more variability, but they reflect more sources of uncertainty than those that result from other methods. With these parameter settings, the `FitInfer` function takes roughly two hours to run on an *Intel Core i7* processor with 16 GB of *RAM*.

Although we did not make use of priors on the covariance parameters for the example analysis above, Log-Normal priors could have been used. The code needed to specify such priors is given below, where prior parameters are set using the strategy described in section 2.4.

```
#defining independent logged log-normal prior distributions
Prior_List=list(
theta1_prior="-(log(theta1)-3.64)^2/(2*0.76^2) -log(theta1*0.76*sqrt(2*pi))",
theta2_prior="-(log(theta2)-3.64)^2/(2*0.76^2) -log(theta2*0.76*sqrt(2*pi))")
#defining derivative of logged log-normal prior distributions
Prior_Der_List=list(
theta1_der_prior="-(log(theta1)-3.64)/(0.76^2*theta1)-1/theta1",
theta2_der_prior="-(log(theta2)-3.64)/(0.76^2*theta2)-1/theta2")
```

5.5 Discussion

Herein, we have examined recent Bayesian methodologies for identifying optimal dynamic treatment regimes and have used data adapted from an HIV trial to illustrate how to perform a standard DTR analysis with these methods. The `BayesDTR` package contains the `BayesMSM` function which allows users to smoothly model the value surface of regimes in a family via Bayesian dynamic MSMs with IPW estimation. These methods allow the user to take an approach that is similar in flavour to frequentist semiparametric methods but that results in estimators that are entirely Bayesian. The function additionally allows users to perform a grid-search for the optimal value, and thereby estimate the oprimal treatment strategy, using a Bayesian IPW or doubly robust estimator. Given the limitations of these methods which include a potentially high computation burden (grid-search methods) or vulnerability to model misspecification (parametric MSM), the package also incorporates functions that perform Gaussian process optimization to allow for the identification of optimal DTRs in conjunction with IPW and DR estimators. The `DesignFit` function in the package fits a $\mathcal{GP}$ on an initial set of design points, and the `SequenceFit` function allows users to sequentially sample more points based on belief about where the optimum lies. Lastly, the `FitInfer` function allows users to quantify uncertainty around the optimal regime. Although this optimization takes an empirical Bayes approach, it is important that the inherent Bayesian perspective is acknowledged in this setting, as a frequentist approach, although it may work well in practice, does not acknowledge that uncertainty in the problem extends beyond aleatory uncertainty. More precisely, this means that a computer experiment as described in this paper does not have an outcome that depends on chance; the outcome will always be the same if the experiment is performed multiple times. Consequently, a frequentist framework does not accommodate the characteristics of this problem. In contrast, although the computer experiment may have a deterministic nature, there is still uncertainty about the optimal regime once the experiment is complete; it is only Bayesian methods that allow for the quantification of this uncertainty.

There are still several improvements that can be made in future versions of the package, for example introducing a function that allows the user to use an estimator for the regime value of their choosing, so as not to be limited to the ones implemented in the `BayesDTR` package. Although the illustrative analysis considered a two-stage problem, this package places no restrictions on the number of stages in the decision-problem. Additionally, as discussed, the treatment rules that can be considered with the package may involve multiple covariates per stage. Additionally, adding other methods to stabilize covariance matrices when they are near non-invertability could be beneficial, for example by adding a nugget effect.

Funding: DRD is supported by a doctoral fellowship from the Fonds de recherche du Québec, Nature et technologie. EEMM and DAS acknowledge support from Discovery Grants from the Natural Sciences and Engineering Research Council of Canada (NSERC). EEMM is a Canada Research Chair (Tier 1) in Statistical Methods for Precision Medicine and acknowledges the support of a chercheur de mérite career award from the Fonds de recherche du Québec, Santé.

Chapter 6

Conclusion

6.1 Summary

This thesis focused on the development of methods for inference of optimal dynamic treatment regimes using a Bayesian lens. Chapter 3 examined the conditions under which dynamic marginal structural models could be regarded as arising from a Bayesian semiparametric inferential framework. This is significant because semiparametric inference is generally a robust inferential framework that can lead to interpretable results but that is generally employed in frequentist settings. Indeed, the resulting inferential procedure proposed in chapter 3 is one that is accessible to the wider research community, all while being entirely Bayesian. This contrasts with other Bayesian methods in this area. The use of posterior predictive inference with the Bayesian bootstrap was examined for both the doubly and singly robust estimators. These estimators' performance was evaluated using simulations which showed that accurate and precise inference can be obtained, with much to be gained in efficiency when outcome models in the doubly robust estimator are correctly specified. The proposed Bayesian methods also benefit from the fact that direct probability statements can be made about quantities of interest; this property is demonstrated by computing the probability that a specific patient is receiving an optimal treatment allocation, thereby allowing for individualized inference.

Chapter 4 of the thesis sought to improve issues that might arise with model misspecification of dynamic MSMs. An improvement was proposed using Gaussian processes to flexibly model the value function and through the use of the expected improvement as an acquisition criterion to identify new points where an optimum may exist. The bridging of this optimization methodology to the DTR realm is consequential, as it has a much sought after duality: it allows for the value function to be flexibly modeled, a desirable property of regression-based approaches, all while allowing for interpretable optimal DTRs to be identified, a property more tractable to value-search approaches when clinically significant families of DTRs are chosen. This chapter further characterized the unique elements involved in optimizing the value function when using point-wise evaluations of an estimator to gain information. Particular attention was given to the sources of variability that can result in the use of such point-wise evaluation of the IPW estimator; this characterization can be found nowhere else in the literature. Simulations further examined how using models that acknowledge that only a noisy version of the value function is observed can lead to improved performance, particularly in settings where the value function is multi-model. This is one of very few works in the DTR literature that creates data-generating mechanisms for multi-modal value functions. Additionally, an examination of these multi-modal value functions showed that contrary to common belief, a grid-search is not always be the best alternative to identifying optimal DTRs, and that it is an inefficient manner by which to use information about the value function. The analysis of an HIV therapeutic trial examined how this precision medicine methodology can be used to obtain a flexible model for the value function and to consider how uncertainty may be quantified.

Chapter 5 explored the use of the proposed estimation methods to infer a two-stage optimal DTR, coming from HIV trial data with an additional simulated component. In particular it showed how therapy could be tailored based on stage-specific CD4 cell counts. The resulting analyses contrasted using a quadratic dynamic MSM, a grid search with Bayesian IPW and augmented IPW estimators, as well as with Gaussian process emulation. The fitting procedures for these methods are provided in detail, an important element due to the subtle constituents in some of these approaches. The development and use of the `BayesDTR` package to implement all methods proposed in this thesis ensures these methodologies to becomes more accessible to the wider research community.

6.2 Future Work

Although this thesis has advanced the methods for DTRs, there are still a variety of interesting, consequential research avenues.

One interesting area of future work is to determine if the utility maximization framework discussed in chapter 3 can be extended to other settings. For example, examining how doubly robust inference can be cast in the proposed utility maximization framework is of importance. Furthermore, work can be done in determining if these methods can be successfully adapted to a survival setting by considering the utility as the negative Cox partial likelihood. This utility framework also has potential to be adopted in the weighted learning approaches discussed in chapter 2, where the focus is to minimize a weighted missclassification error, which can be considered a negative utility function.

Some Bayesian causal inference methods adapt frequentist semiparametrics methods in order to arrive at an estimation procedure; the work in chapter 3 is a fruitful example of this. It would be interesting to identify a formal framework, whereby frequentist semiparamteric methods could be adapted to a Bayesian setting, with all its benefits. For the work presented in chapter 3, the proposed method used a robust non-informative prior; however further study to examine how informative priors can be used, for example to code information about where an optimal regime may be, is warranted.

The emulation methodology presented in chapter 4 also leads to several avenues of future work. For example, evaluating the consequences that a fully Bayesian treatment, both on inference itself and on practical feasibility, would be important in order to understand the trade-offs between computational feasibility and inference. Further comparison of other sequential sampling and stopping criteria would also be beneficial. Examining ways by which to incorporate sampling uncertainty that do not require such a computationally intensive procedures is also important.

As with many statistical methods, the resulting inference for methods used in chapters 3-5 relies on the correct specification of models, in this context of treatment assignment and/or outcome models. It would be of interest to specify a more formal framework whereby flexible models can be specified and sensitivity analyses conducted to determine how sensitive the results are to model specification. This is particularly consequential in longitudinal settings where the number of models that must be posited grows with the number of time points. These sensitivity analyses would also be beneficial in examining how sensitive results are to the specification of the marginal structural models in dynamic MSMs, which are often taken to be a parsimonious function. Integrating the possibility of these sensitivity analyses into the `BayesDTR` package is also of future interest. Similarly, designing methods that allow for sensitivity analyses in cases where there are unmeasured confounders would be of tremendous practical use.

6.3 Concluding Remarks

This thesis contributes to the methodological body of work for DTRs, specifically using a Bayesian lens. New methods were developed, and, importantly, they were contextualized into the contemporary literature on DTRs and precision medicine methods. In particular, these methods have a very different mechanics to many other methods in the current DTR literature. The illustrative examples on HIV therapy support the utility of these methods in clinical research, a field that will continue to increase its focus on precision health in the coming years.

APPENDIX A

Appendix to Manuscript 1

A.1 Technical Details

No Unmeasured Confounders Assumption: Consider the *unobserved history* up to time k, $\mathcal{F}_k = \{(y, z_t, x_t, u_t), t = 1, ..., k\}$, where u_t are unobserved covariates. Furthermore, consider *observed history* up to time k is given by $\mathcal{H}_k = \{(y, z_t, x_t), t = 1, ..., k\}$. Then, the sequence of treatments $\{z_t\}$ is unconfounded relative to latent variables $\{u_t\}$ if for each k, z_k and $\{u_t, t = 1, .., k\}$ are conditionally independent given $(\mathcal{H}_{k-1}, x_k)$. Mathematically, this may be written as $p_{\mathcal{O}}(z_k|\mathcal{F}_{k-1}, u_k, x_k) = p_{\mathcal{O}}(z_k|\mathcal{H}_{k-1}, x_k)$, $k = 1, ..., K$.

De Finetti Representation: Below we consider a more general form of the De Finetti representation presented in the main paper. We do this by considering the vector $(y_i, \bar{x}_i, \bar{z}_i, u_i)$, where u_i are determinants of the outcome and intermediate variables. We assume that these vectors are infinitely exchangeable in order to deduce the de Finetti representation in the observational world:

$$\begin{aligned} p_{\mathcal{O}}(b_1, ..., b_n) = \int_{\phi,\gamma,\tau} \prod_{i=1}^{n} \Bigg[\int_u & p_{\mathcal{O}}(y_i|\bar{x}_i, \bar{z}_i, u_i, \tau) \\ & \prod_{j=1}^{K} p_{\mathcal{O}}(x_{ij}|\bar{z}_{i(j-1)}, \bar{x}_{i(j-1)}, u_i, \phi_{1j}) p_{\mathcal{O}}(u_i|\phi_2) du_i \\ & \prod_{j=1}^{K} p_{\mathcal{O}}(z_{ij}|\bar{z}_{i(j-1)}\bar{x}_{ij}, \gamma_j) \Bigg] p(\phi, \gamma) d\tau d\phi d\gamma. \end{aligned}$$

The absence of u_i in the treatment assignment probability is due to the no unmeasured confounders assumption. We can also look at the representation in the experimental measure by considering: $v_i = (b_i, g_i) \equiv (y_i, \bar{x}_i, \bar{z}_i, g_i)$, and assuming infinite exchangeability in order to obtain

$$\begin{aligned} p_{\mathcal{E}}(v_1, ..., v_n) = \int \prod_{i=1}^{n} \Bigg[\int_u & p_{\mathcal{E}}(y_i|\bar{x}_i, \bar{z}_i, g_i, u_i, \tau) \\ & \prod_{j=1}^{K} p_{\mathcal{E}}(x_{ij}|z_{i(j-1)}, x_{i(j-1)}, u_i, g_i, \phi_{1j}) p_{\mathcal{E}}(u_i|\phi_2) du_i \\ & \prod_{j=1}^{K} p_{\mathcal{E}}(z_{ij}|z_{i(j-1)}, x_{ij}, g_i, \alpha_j) p(g_i) \Bigg] p(\phi, \alpha) d\tau d\phi d\alpha. \end{aligned}$$

Change of Measure Details Corresponding to Equation (2.3):

Let us first see how to fully develop the importance sampling argument, and then how to obtain the form of the weights. We connect the experimental world with the observational world as follows:

$$\begin{aligned}
E_{\mathcal{E}}[U(b^*,g,\beta)|\bar{b}] =& E_{G_{\mathcal{E}}}\left[E_{b^*_{\mathcal{E}}|g}[U(b^*,g,\beta)|g,\bar{b}]\middle|\bar{b}\right] \\
=& E_{G_{\mathcal{E}}}\left[\int_{b^*} U(b^*,g,\beta)p_{\mathcal{E}}(b^*|g,\bar{b})\frac{p_{\mathcal{O}}(b^*|\bar{b})}{p_{\mathcal{O}}(b^*|\bar{b})}\middle|\bar{b}\right] \\
=& E_{G_{\mathcal{E}}}\left[E_{\mathcal{O}}\left[U(b^*,g,\beta)\frac{\mathbb{1}_{g(\bar{x}^*)}(\bar{z}^*)}{\prod_{k=1}^{K} p_{\mathcal{O}}(z^*_k|\bar{z}^*_{k-1},\bar{x}^*_k,\bar{b})}\middle|\bar{b}\right]\middle|\bar{b}\right] \\
=& E_{\mathcal{O}}\left[\frac{\frac{1}{C_G}\sum_{\{r\in\mathcal{I}\}} U(b^*,g^r,\beta)\mathbb{1}_{g^r(\bar{x}^*)}(\bar{z}^*)}{\prod_{k=1}^{K} p_{\mathcal{O}}(z^*_k|\bar{z}^*_{k-1},\bar{x}^*_k,\bar{b})}\middle|\bar{b}\right]. \\
=& E_{\mathcal{O}}\left[\frac{1}{C_G}\sum_{\{r\in\mathcal{I}\}} w^{*r}U(b^*,g^r,\beta)\middle|\bar{b}\right].
\end{aligned}$$

Now let us examine how we may obtain the weights w^* for DTR-MSMs. Note that we need only consider the single-stage problem, as the multi-stage case follows directly.

$$\begin{aligned}
&p_{\mathcal{E}}(Y=y,Z=z,X=x|G=g) \\
&\quad=\frac{p_{\mathcal{E}}(Y=y,Z=z,X=x|G=g)}{p_{\mathcal{O}}(Y=y,Z=z,X=x)}p_{\mathcal{O}}(Y=y,Z=z,X=x) \\
&\quad=\frac{p_g(Y=y,Z=z,X=x)}{p_{\mathcal{O}}(Y=y,Z=z,X=x)}p_{\mathcal{O}}(Y=y,Z=z,X=x) \\
&\quad=\frac{p_g(Y=y,g(X)=z,X=x)}{p_{\mathcal{O}}(Y=y|Z=z,X=x)p_{\mathcal{O}}(Z=z|X=x)p_{\mathcal{O}}(X=x)}p_{\mathcal{O}}(Y=y,Z=z,X=x) \\
&\quad=\frac{p_g(Y=y|g(X)=z,X=x)p_g(g(X)=z|X=x)p_g(X=x)}{p_{\mathcal{O}}(Y=y|Z=z,X=x)p_{\mathcal{O}}(Z=z|X=x)p_{\mathcal{O}}(X=x)}p_{\mathcal{O}}(Y=y,Z=z,X=x)
\end{aligned}$$

Note that in the above argument, when we condition on $g(X)=z, X=x$, we may run into issues if $g(x)$ does not equal z. However, in practice this is not a concern as the joint probability $p_{\mathcal{E}}(Y=y,g(X)=z,X=x)$ would take the value zero in such a situation, and so this term would not contribute to the calculation. Continuing, we find:

$$
\begin{aligned}
&p_{\mathcal{E}}(Y=y, Z=z, X=x|G=g)\\
=&\frac{p_g(Y=y|g(X)=z, X=x)p_g(g(X)=z|X=x)}{p_{\mathcal{O}}(Y=y|Z=z, X=x)p_{\mathcal{O}}(Z=z|X=x)}p_{\mathcal{O}}(Y=y, Z=z, X=x)\\
=&\frac{p_g(Y=y|g(X)=z, X=x)\mathbb{1}_{g(x)}(z)}{p_{\mathcal{O}}(Y=y|Z=z, X=x)p_{\mathcal{O}}(Z=z|X=x)}p_{\mathcal{O}}(Y=y, Z=z, X=x).
\end{aligned}
$$

Now, we are looking for cancellation between the outcome probabilities. We have already established that when $g(x) \neq z$, the numerator is equal to zero. When $g(x) = z$, we have that $p_g(Y = y|g(X) = z, X = x) = p_{\mathcal{O}}(Y = y|Z = z, X = x)$. Thus we may finish by writing:

$$
\begin{aligned}
&p_{\mathcal{E}}(Y=y, Z=z, X=x|G=g)\\
=&\frac{\mathbb{1}_{g(x)}(z)}{p_{\mathcal{O}}(Z=z|X=x)}p_{\mathcal{O}}(Y=y, Z=z, X=x),
\end{aligned}
$$

yielding the weights that we were seeking.

A.2 Discussion on Non-Regularity in DTRs

We note that the arguments presented in this paper are Bayesian. Thus, conditional on the posited model, the resulting inference is valid for any sample size. We emphasize that the premise of the Bayesian bootstrap is not related to attaining asymptotic consistency, but simply it is about proposing a specific model for the data, and carrying out inference conditional on this model. That being said, we may still ask how well we would expect these methods to perform as more data are observed. As noted in the main paper, the parameters of dynamic MSMs can be shown to be consistent [Orellana et al., 2010a, van der Laan and Petersen, 2007]. We mainly make use of the estimator for the value of a specific DTR, and this is also asymptotically normal and regular as laid out by Murphy et al. [2001].

In what follows, we emphasize that estimation of dynamic MSMs do not suffer from non-regularity as is the case with other methods, like Q-Learning, G-estimation of structural nested mean models, and dynamic weighted ordinary least squares (dWOLS) [Wallace and Moodie, 2015].

We proceed by discussing the relevant literature on non-regularity in order to understand why it does not play a role in the estimation of parameters in dynamic MSMs. Additionally, we present a simulation that illustrates our point. Our simulation is similar in spirit to that of Chakraborty et al. [2010], where we draw 1000 bootstrapped samples and evaluate whether the obtained coverage differs significantly from the nominal 95%. As we expect, the parameters in the MSM do not exhibit issues with non-regularity.

It was Robins [2004] who first raised the issue of non-regularity in methods aimed at estimating parameters relevant to identifying optimal DTRs. The key issue is illustrated in the context of estimating the absolute value of a population mean, $|\mu|$, from n $i.i.d$ observations. A maximum likelihood approach may first estimate the mean $\hat{\mu}$, and then this may be plugged into $|\cdot|$ to obtain an estimator for $|\mu|$. What Robins [2004] emphasizes is that $|\hat{\mu}|$ has different asymptotic distributions depending on the value of μ (when $\mu = 0$ $vs.$ $\mu \neq 0$). This is what yields a non-regular estimator, and the crux of this issue is in the fact that the absolute value function is discontinuous at zero. Consequently, Wald-type confidence intervals do not perform well. Chakraborty et al. [2010] examine whether bootstrap confidence intervals yield appropriate inference in non-regular settings, but they point out that the success of the bootstrap relies on the smoothness of the estimator. Accordingly, one should not expect the bootstrap to provide adequate inference at or near the point of non-regularity.

For Q-learning, it is clear where non-regularity arises. Consider a two-stage setting where the stage II model is $y_i = \gamma_{20} + \gamma_{21} z_1 + \gamma_{22} x_1 z_1 + \gamma_{23} z_2 + \gamma_{24} x_2 z_2$. The stage I pseudo-outcome becomes $\tilde{y}_i = \gamma_{20} + \gamma_{21} z_1 + \gamma_{22} x_1 z_1 + \mathbb{1}(\gamma_{23} z_2 + \gamma_{24} x_2 z_2 > 0)$. This pseudo-outcome is discontinuous at $\gamma_{23} z_2 + \gamma_{24} x_2 z_2 = 0$. Therefore, we should expect that plugging-in $\hat{\gamma}_{20}, \hat{\gamma}_{21}, \hat{\gamma}_{22}, \hat{\gamma}_{22}, \hat{\gamma}_{22}$

to compute $\hat{\tilde{y}}_i$ will cause issues with the estimation of stage I parameters, as these will depend on a discontinuous function of other parameters. Non-regularity is not only an issue at $\gamma_{23}z_2 + \gamma_{24}x_2z_2 = 0$ but also near it; Chakraborty et al. [2010] explored this via simulation and found non-regularity to impact inference. Earlier works also noted non-regularity to arise in G-estimation [Moodie and Richardson, 2010].

The parameters in dynamic MSMs do not suffer from the above-mentioned issues. Unlike G-estimation, dWOLS, and Q-learning, dynamic MSMs do not require recursively solving estimating equations, where the stage I equation has plug-in estimators obtained by solving a stage II estimating equation. Therefore, for dynamic MSMs, the estimators are not functions of discontinuous functions of other estimators. Ultimately, this means that the parameters in dynamic MSMs do not suffer from the same types of difficulties with non-regularity. Let us now examine an example in which non-regularity impacts inference in Q-learning but plays no role in the inference of parameters in dynamic MSMs. We consider a family of regimes that says treat if $x_k > \theta_k$ for $k = 1, 2$. The proposed data-generating mechanism is one that allows for straightforward marginalization so that we can posit a correct model for $E[Y^{\theta_1\theta_2}]$. The outcome is given by:

$$Y = \gamma_0 + \gamma_1 z_1 + \gamma_2 x_1 z_1 + \gamma_3 z_2 + \gamma_4 x_2 z_2 + \epsilon \tag{A.1}$$

Variables are distributed as: $x_1 \sim N(0, 9), x_2 \sim N(0, 4)z_1,\ z_2 \sim bern(0.5)$. Then,

$$E[Y^{\theta_1\theta_2}] = \gamma_0 + \gamma_1 C_{11}(\theta_1) + \gamma_2 C_{12}(\theta_1) + \gamma_3 C_{21}(\theta_2) + \gamma_4 C_{22}(\theta_2) \tag{A.2}$$

where,

$$\begin{aligned} C_{21}(\theta_2) &= E[\mathbb{1}(x_2 > \theta_2)|x_1, z_1] = p(x_2 > \theta_2), \\ C_{22}(\theta_2) &= E[x_2\mathbb{1}(x_2 > \theta_2)|x_1, z_1] = \frac{4}{\sqrt{2\pi}} exp(-\theta_2^2/(2 \cdot 4^2)). \end{aligned}$$

C_{11}, C_{12} have an analogous form. Then, we have an analytic form for the marginal model. We assume further that $\gamma_3, \gamma_4 > 0$ and consider the following scenarios:

- Scenario I: $\gamma_0 = 1, \gamma_1 = 1, \gamma_2 = 1, \gamma_3 = 0, \gamma_4 = 0$.
- Scenario II: $\gamma_0 = 1, \gamma_1 = 1, \gamma_2 = 1, \gamma_3 = 0.001, \gamma_4 = 0.001$.
- Scenario III: $\gamma_0 = 1, \gamma_1 = 1, \gamma_2 = 1, \gamma_3 = 1, \gamma_4 = 1$.

Scenario I explores inference in a non-regular setting; scenario II explores a near non-regular setting, and scenario III explores a regular setting. We make use of $B = 1000$ bootstrap samples, a sample size of $n = 1000$, and $R = 500$ replications. We first examine these scenarios in the context of Q-learning. The correctly specified models that we fit are as follows:

$$\begin{aligned}
&\textit{Stage I}: \gamma_{10} + \gamma_{11} z_1 + \gamma_{12} x_1 z_1 \\
&\textit{Stage II}: \gamma_{20} + \gamma_{21} z_1 + \gamma_{22} x_1 z_1 + \gamma_{23} z_2 + \gamma_{24} x_2 z_2
\end{aligned}$$

The pseudo-outcome in stage I is: $\gamma_{20} + \gamma_{21} z_1 + \gamma_{22} x_1 z_1 + (\gamma_{23} + \gamma_{24} x_2) \mathbb{1}(\gamma_{23} + \gamma_{24} x_2 > 0)$. Note that because of the specific data-generating mechanism, these models are correctly specified.

Table A.1 shows that, as expected, the parameters for the stage II model present no evidence of non-regularity as measured by coverage or bias. We note that apart from the point estimates, stage II inference is the same for all scenarios, hence the shorter table. From Table A.2, we see where the non-regularity becomes present. The stage I intercept exhibits coverage that is significantly different from nominal in the non-regular case. This persists even in the close-to-non-regular setting. Furthermore, as is shown in Table A.5, evidence of non-regularity disappears in a gradient, as the data-generating mechanism gets further from the completely non-regular setting.

Table A.1: Scenario I Coverage of 95% CI for Q-learning stage II parameters. $B = 1000; n = 1000; R = 500$.

Parameter	Coverage	Mean	Bias	SD
γ_{20}	0.946	0.9997	-0.0003	0.0112
γ_{21}	0.958	1.0004	0.0004	0.0124
γ_{22}	0.952	1.0001	0.0001	0.0029
γ_{23}	0.940	0.0000	0.0000	0.0131
γ_{24}	0.938	0.0000	0.0000	0.0045

*indicates significant difference from 0.95

Table A.2: Coverage of 95% CI for Q-learning stage I parameters $\gamma_{10}, \gamma_{11}, \gamma_{12}$. $B = 1000; n = 1000; R = 500$.

Parameter	$\gamma_3 = \gamma_4$	Coverage	Estimate	Bias	SD
γ_{10}	0	0.884*	1.0059	0.0059	0.0098
γ_{11}		0.958	1.0004	0.0004	0.0124
γ_{12}		0.952	1.0001	0.0001	0.0030
γ_{10}	0.001	0.898*	1.0065	0.0051	0.0098
γ_{11}		0.958	1.0004	0.0004	0.0124
γ_{12}		0.952	1.0001	0.0001	0.0030
γ_{10}	1	0.944	2.3997	0.0041	0.0653
γ_{11}		0.954	0.9948	-0.0052	0.0927
γ_{12}		0.954	0.9998	-0.0002	0.0213

*indicates significant difference from 0.95

The Q-learning results are only presented for the frequentist bootstrap, as the use of the Bayesian bootstrap has not been studied in this literature. In the following, we look at the resulting inference for the Frequentist and Bayesian dynamic MSMs. The θ used to create the augmented data required for these methods are $\{-4, -2.5, -1, 0.5, 2, 3.5\}$. As expected, there are no issues with any coverage probabilities; this can be seen in Tables A.3 and A.4.

Table A.3: Results frequentist dynamic MSM; $B = 1000; n = 1000; R = 500$.

Parameter	$\gamma_3 = \gamma_4$	Coverage	Estimate	Bias	SD
γ_0	0	0.954	1.0038	0.0038	0.4194
γ_1		0.952	0.9597	-0.0403	1.5826
γ_2		0.948	0.9913	-0.0087	0.7859
γ_3		0.958	0.0439	0.0439	1.2823
γ_4		0.944	-0.0046	-0.0046	0.8172
γ_0	0.001	0.954	1.0037	0.0037	0.4194
γ_1		0.952	0.9597	-0.0403	1.5826
γ_2		0.948	0.9913	-0.0087	0.7859
γ_3		0.958	0.0448	0.0438	1.2823
γ_4		0.944	-0.0036	-0.0046	0.8172
γ_0	1	0.956	0.9921	-0.0079	0.5344
γ_1		0.946	0.9944	-0.0056	2.0336
γ_2		0.958	1.0081	0.0081	0.9960
γ_3		0.946	1.0114	0.0114	1.6185
γ_4		0.956	0.9900	-0.0100	1.0225

*indicates significant difference from 0.95

Table A.4: Results Bayesian dynamic MSM. $B = 1000; n = 1000; R = 500$.

Parameter	$\gamma_3 = \gamma_4$	Coverage	Estimate	Bias	SD
γ_0	0.000	0.950	0.9922	-0.0078	0.4183
γ_1		0.934	1.0303	0.0303	1.6094
γ_2		0.952	1.0099	0.0099	0.7619
γ_3		0.938	-0.0270	-0.0270	1.2993
γ_4		0.952	-0.0092	-0.0092	0.7893
γ_0	0.001	0.950	0.9922	-0.0078	0.4182
γ_1		0.934	1.0302	0.0302	1.6095
γ_2		0.952	1.0099	0.0099	0.7618
γ_3		0.938	-0.0260	-0.0270	1.2994
γ_4		0.952	-0.0082	-0.0092	0.7892
γ_0	1.000	0.966	1.0130	0.0130	0.5264
γ_1		0.956	0.9613	-0.0387	2.0571
γ_2		0.956	0.9826	-0.0174	0.9608
γ_3		0.958	1.0305	0.0305	1.6333
γ_4		0.952	1.0102	0.0102	0.9893

*indicates significant difference from 0.95

In what follows, we examine how inference is impacted as $\gamma_3 = \gamma_4$ get further away from the non-regular case. For Table A.5, we see that as we get further from non-regularity, the closer to nominal coverage becomes in Q-learning. Note that only results for the γ_{10} parameter are shown as this is the parameter that most clearly exhibits issues with non-regularity in Q-learning. From Tables A.6 and A.7, we see that the frequentist and Bayesian bootstrap yield adequate inference with the dynamic MSM, regardless of proximity to the non-regular case.

Table A.5: Results of Q-Learning for different levels of non-regularity; $B = 500, n = 1000, R = 500$.

Parameter	$\gamma_{23} = \gamma_{24}$	p-value	Coverage	Mean Estimate	Bias	SD
γ_{10}	0	< 0.001	0.854	1.0064	0.0064	0.0099
γ_{10}	0.001	< 0.001	0.858	1.0069	0.0056	0.0098
γ_{10}	0.005	< 0.001	0.906	1.0101	0.0031	0.0097
γ_{10}	0.010	0.031	0.928	1.0156	0.0017	0.0097
γ_{10}	0.050	0.051	0.930	1.0700	0.0002	0.0104
γ_{10}	0.100	0.473	0.942	1.1397	0.0001	0.0121
γ_{10}	1.000	0.356	0.940	2.3963	0.0007	0.0672

Table A.6: Frequentist dynamic MSM; $B = 500, n = 1000, R = 500$.

	$\gamma_2 = \gamma_4$	p-value	Coverage	Mean Estimate	Bias	SD
γ_0	0	0.356	0.960	0.9777	-0.0223	0.4193
γ_1		0.608	0.944	1.1461	0.1461	1.5732
γ_2		0.758	0.954	1.0355	0.0355	0.7879
γ_3		0.608	0.944	-0.1176	-0.1176	1.2878
γ_4		0.608	0.944	-0.0476	-0.0476	0.8202
γ_0	0.001	0.356	0.960	0.9777	-0.0223	0.4193
γ_1		0.608	0.944	1.1461	0.1461	1.5732
γ_2		0.758	0.954	1.0355	0.0355	0.7879
γ_3		0.608	0.944	-0.1165	-0.1175	1.2878
γ_4		0.608	0.944	-0.0465	-0.0475	0.8202
γ_0	0.005	0.356	0.960	0.9778	-0.0222	0.4193
γ_1		0.608	0.944	1.1459	0.1459	1.5730
γ_2		0.758	0.954	1.0352	0.0352	0.7879
γ_3		0.608	0.944	-0.1124	-0.1174	1.2875
γ_4		0.758	0.946	-0.0423	-0.0473	0.8202
γ_0	0.010	0.259	0.962	0.9780	-0.0220	0.4193
γ_1		0.608	0.944	1.1457	0.1457	1.5728
γ_2		0.758	0.954	1.0350	0.0350	0.7879
γ_3		0.608	0.944	-0.1073	-0.1173	1.2872
γ_4		0.758	0.946	-0.0370	-0.0470	0.8203
γ_0	0.050	0.259	0.962	0.9791	-0.0209	0.4198
γ_1		0.758	0.946	1.1441	0.1441	1.5725
γ_2		0.918	0.948	1.0328	0.0328	0.7887
γ_3		0.608	0.944	-0.0663	-0.1163	1.2857
γ_4		0.918	0.948	0.0053	-0.0447	0.8214
γ_0	0.100	0.259	0.962	0.9804	-0.0196	0.4210
γ_1		0.918	0.952	1.1421	0.1421	1.5743
γ_2		0.758	0.946	1.0302	0.0302	0.7909
γ_3		0.608	0.944	-0.0149	-0.1149	1.2856
γ_4		0.608	0.944	0.0582	-0.0418	0.8240
γ_0	1.000	0.918	0.952	1.0052	0.0052	0.5519
γ_1		0.608	0.956	1.1056	0.1056	1.9877
γ_2		0.473	0.958	0.9821	-0.0179	1.0391
γ_3		0.356	0.960	0.9086	-0.0914	1.5885
γ_4		0.608	0.956	1.0101	0.0101	1.0772

Table A.7: Bayesian dynamic MSM; $B = 500, n = 1000, R = 500$.

Parameter	$\gamma_3 = \gamma_4$	p-val	percent	Estimate	Bias	SD
γ_0	0	0.051	0.930	0.9857	-0.0143	0.4521
γ_1		0.259	0.962	1.0905	0.0905	1.5410
γ_2		0.081	0.932	1.0211	0.0211	0.8449
γ_3		0.608	0.956	-0.0709	-0.0709	1.2483
γ_4		0.356	0.940	-0.0363	-0.0363	0.8702
γ_0	0.001	0.051	0.930	0.9857	-0.0143	0.4521
γ_1		0.259	0.962	1.0906	0.0906	1.5411
γ_2		0.081	0.932	1.0211	0.0211	0.8449
γ_3		0.608	0.956	-0.0699	-0.0709	1.2484
γ_4		0.356	0.940	-0.0353	-0.0363	0.8702
γ_0	0.005	0.051	0.930	0.9856	-0.0144	0.4521
γ_1		0.259	0.962	1.0909	0.0909	1.5414
γ_2		0.081	0.932	1.0212	0.0212	0.8449
γ_3		0.608	0.956	-0.0661	-0.0711	1.2486
γ_4		0.356	0.940	-0.0314	-0.0364	0.8702
γ_0	0.010	0.051	0.930	0.9855	-0.0145	0.4521
γ_1		0.259	0.962	1.0912	0.0912	1.5418
γ_2		0.081	0.932	1.0212	0.0212	0.8449
γ_3		0.608	0.956	-0.0613	-0.0713	1.2489
γ_4		0.356	0.940	-0.0264	-0.0364	0.8701
γ_0	0.050	0.051	0.930	0.9847	-0.0153	0.4524
γ_1		0.356	0.960	1.0941	0.0941	1.5459
γ_2		0.081	0.932	1.0218	0.0218	0.8455
γ_3		0.473	0.958	-0.0233	-0.0733	1.2519
γ_4		0.473	0.942	0.0131	-0.0369	0.8704
γ_0	0.100	0.051	0.930	0.9838	-0.0162	0.4534
γ_1		0.356	0.960	1.0977	0.0977	1.5531
γ_2		0.051	0.930	1.0225	0.0225	0.8473
γ_3		0.259	0.962	0.0243	-0.0757	1.2573
γ_4		0.182	0.936	0.0624	-0.0376	0.8719
γ_0	1.000	0.051	0.930	0.9668	-0.0332	0.5651
γ_1		0.918	0.948	1.1624	0.1624	2.0210
γ_2		0.356	0.940	1.0355	0.0355	1.0660
γ_3		0.918	0.948	0.8805	-0.1195	1.6245
γ_4		0.259	0.938	0.9507	-0.0493	1.0883

A.3 Considerations for Doubly Robust Estimator

In this section, we present additional details related to ideas discussed in Section 3 of the main paper. This includes details about how to fit outcome models in the doubly robust estimator.

A.3.1 Outcome Models

In this section, we provide details about how to fit outcome models for the doubly robust estimator. Recall that for $k = K$, ϕ^*_{K+1} is defined as

$$\phi^*_{K+1}(\bar{x}^*_K) = E_{\mathcal{O}}[y^*|\bar{x}^*_K, \bar{z}^*_K = \bar{g}_K(\bar{x}_K), \bar{b}],$$

and for $k = K-1, ..., 1$, ϕ^*_{k+1} is defined as

$$\phi^*_{k+1}(\bar{x}^*_k) = E_{\mathcal{O}}[\phi^*_{k+2}(\bar{x}_{k+1})|\bar{x}^*_k, \bar{z}^*_k = \bar{g}_k(\bar{x}^*_k), \bar{b}].$$

First, note that based on the prior we have selected (which yields the non-parametric Bayesian bootstrap as the posterior), it is enough to fit these models on the observed data, conditional on a draw from the Dirichlet weights. In a regression setting, the weights would just be incorporated into the *weights* argument in the *lm* function. We now focus on how to pose these models, based on the data generating mechanism in the single threshold simulation, which can be found in Appendix C. The outcome is generated via $y = x_1 - (-\theta^{opt} + x_1)(\mathbb{1}_{x_1>\theta^{opt}} - z_1) - (-\theta^{opt} + x_2)(\mathbb{1}_{x_2>\theta^{opt}} - z_2) + \sqrt{0.5}\epsilon$. Note that θ^{opt} is a constant and $\epsilon \sim N(0, 1)$. Then, we may look to fit the following model:

$$\begin{aligned} E[y|\bar{x}, \bar{z}] =& \beta_{21}x_1 + \beta_{22}\mathbb{1}_{x_1>\theta^{opt}} + \beta_{23}z_1 + \beta_{24}x_1\mathbb{1}_{x_1>\theta^{opt}} + \beta_{25}x_1z_1 \\ &+ \beta_{26}\mathbb{1}_{x_2>\theta^{opt}} + \beta_{27}z_2 + \beta_{28}x_2\mathbb{1}_{x_2>\theta^{opt}} + \beta_{29}x_2z_2. \end{aligned} \tag{A.3}$$

We use this model to compute $\psi_2 = E[y|\bar{x}, z_1, z_2 = g(x_2)]$. We then seek to fit a model conditional on just stage one information. This requires marginalizing over x_2 when $z_2 = g(x_2)$ in equation A.3. For this, we need to compute a few quantities:

$$\begin{aligned}
&1)E[\mathbb{1}_{x_2>\theta^{opt}}|x_1, z_1] = p(\theta^{opt} - z_1 - 0.5x_1 < \epsilon|x_1, z_1)\\
&\quad = 1 - \Phi(\theta^{opt} - z_1 - 0.5x_1)\\
&\quad := T_1(x_1, z_1)\\
&2)E[g(x_2)|x_1, z_1] = E[\mathbb{1}_{x_2>\theta}|x_1, z_1]\\
&\quad = 1 - \Phi(\theta - z_1 - 0.5x_1)\\
&\quad := T_2(x_1, z_1)\\
&3)E[x_2\mathbb{1}_{x_2>\theta^{opt}}|x_1, z_1] = E[(z_1 + 0.5x_1 + \epsilon)\mathbb{1}_{\theta^{opt}<z_1+0.5x_1+\epsilon}|x_1, z_1]\\
&\quad := T_3(x_1, z_1)\\
&4)E[x_2 g(x_2)|x_1, z_1] = E[(z_1 + 0.5x_1 + \epsilon)\mathbb{1}_{\theta<z_1+0.5x_1+\epsilon}|x_1, z_1]\\
&\quad := T_4(x_1, z_1)
\end{aligned}$$

T_1, T_2, T_3 may be approximated numerically through quick draws of a normal distribution. This leads us to the model:

$$\begin{aligned}
&E[\psi_2|x_1, z_1] = \beta_{11}x_1 + \beta_{12}\mathbb{1}_{x_1>\theta^{opt}} + \beta_{13}z_1 + \beta_{14}x_1\mathbb{1}_{x_1>\theta^{opt}} + \beta_{15}x_1z_1\\
&+ \beta_{16}T_1(x_1, z_1) + \beta_{17}T_2(x_1, z_1) + \beta_{18}T_3(x_1, z_1) + \beta_{19}T_4(x_1, z_1).
\end{aligned}$$

When $\theta = \theta^{opt}$, then we have $T_1 = T_2$ and $T_3 = T_4$, and so two of these terms must be taken out of the model in this special case. Note that marginalization becomes slightly complex as, x_2 depends on both z_1 and x_1. If it only dependent on z_1 which is binary, things would be simplified as stage 1 terms would absorb any marginalization terms. Of course, in practice it is difficult to correctly specify these models, but one would hope that specifying a flexible enough model would lead to improved results with regard to efficiency. Once these two models have been fit, we may compute

$$\begin{aligned}\phi_2(x_1) = \beta_{11}x_1 + \beta_{12}\mathbb{1}_{x_1>\theta^{opt}} + \beta_{13}g(x_1) + \beta_{14}x_1\mathbb{1}_{x_1>\theta^{opt}} + \beta_{15}x_1g(x_1) + \beta_{16}T_1(x_1, g(x_1))\\ + \beta_{17}T_2(x_1, g(x_1)) + \beta_{18}T_3(x_1, g(x_1)) + \beta_{19}T_4(x_1, g(x_1))\end{aligned}$$

and

$$\begin{aligned}\phi_3(\bar{x}_2) =& \beta_{21}x_1 + \beta_{22}\mathbb{1}_{x_1>\theta^{opt}} + \beta_{23}g(x_1) + \beta_{24}x_1\mathbb{1}_{x_1>\theta^{opt}} + \beta_{25}x_1g(x_1)\\ &+ \beta_{26}\mathbb{1}_{\theta^{opt}>x_2} + \beta_{27}g(x_2) + \beta_{28}x_2\mathbb{1}_{\theta^{opt}>x_2} + \beta_{29}x_2g(x_2).\end{aligned}$$

We then use these last two expressions in the doubly robust estimator.

A.3.2 Bayesian Double Robustness

If we are able to show the equivalence between expressions (10) and (11) in the main article, then we will have demonstrated the double robustness property. Consider:

$$\begin{aligned}&\phi^*_2(\bar{x}^*_0) + \sum_{k=2}^{K} w^*_{k-1}(\phi^*_{k+1}(\bar{x}^*_k) - \phi^*_k(\bar{x}^*_{k-1})) + w^*_K(y^* - \phi^*_{K+1}(\bar{x}^*_K)))\\ &= \phi^*_2(\bar{x}^*_0) + w^*_K(y^* - \phi^*_{K+1}(\bar{x}^*_K)) + \sum_{k=2}^{K} w^*_{k-1}\phi^*_{k+1}(\bar{x}^*_k) - \sum_{k=1}^{K-1} w^*_k\phi^*_{k+1}(\bar{x}^*_k)\\ &= \phi^*_2(\bar{x}^*_0) + w^*_K(y^* - \phi^*_{K+1}(\bar{x}^*_K)) + w^*_{K-1}\phi^*_{K+1}(\bar{x}^*_{K-1}) - w^*_1\phi^*_2(\bar{x}^*_1) - \sum_{k=2}^{K-1}(w^*_k - w^*_{k-1})\phi^*_{k+1}(\bar{x}^*_k)\\ &= w^*_Ky^* - \sum_{k=1}^{K}(w^*_k - w^*_{k-1})\phi^*_{k+1}(\bar{x}^*_k) - \sum_{k=2}^{K+1} w^*_{k-1}(h(\bar{B}) - h(\bar{b}))\\ &= w^*_Ky^* - w^*_Kh(\bar{b}) + w^*_0h(\bar{b}) - \sum_{k=1}^{K}(w^*_k - w^*_{k-1})\phi^*_{k+1}(\bar{X}^*_k) + \sum_{k=1}^{K}(w^*_k - w^*_{k-1})h(\bar{b})\\ &= h(\bar{b}) + w^*_K(y^* - h(\bar{b})) - \sum_{k=1}^{K}(w^*_k - w^*_{k-1})(\phi^*_{k+1}(\bar{x}^*_k) - h(\bar{b})),\end{aligned}$$

recalling that $w^*_0 = 0$ and that $h(\bar{b}) = E_g[y^*|\bar{b}]$. From the first expression we may see that this is an unbiased estimator when the conditional means are correctly specified. This is obtained from an iterated expectation argument, and by showing that $E[w^*_{k-1}(\phi^*_{k+1}(\bar{x}^*_k) - \phi^*_k(\bar{x}^*_{k-1}))|\bar{x}_{k-1}] = 0$. The full argument can be seen in Orellana et al. [2010b]. The last expression allows us to see this is unbiased when the treatment assignment models are correctly specified by noting that $E[w^*_k|\bar{x}^*_k, \bar{z}_{k-1}] = w^*_{k-1}$ and by again using an iterated expectation.

A.4 Simulation Details

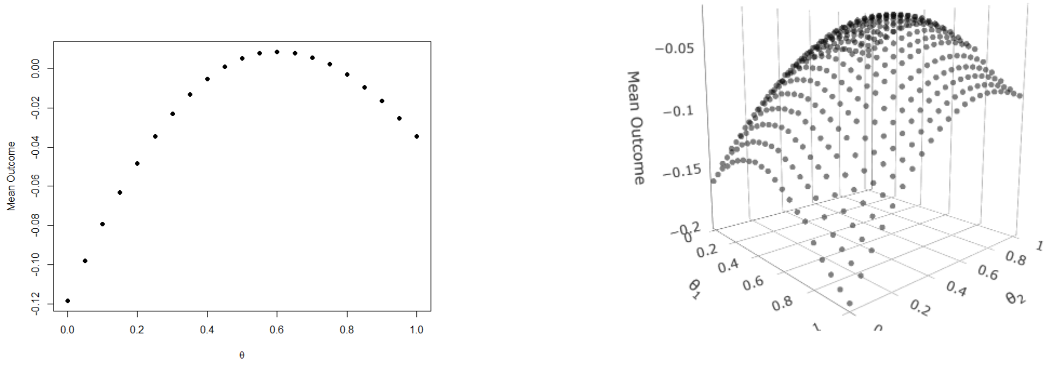
This appendix explores inference for the following regimes types: 1) single threshold regimes, 2) double threshold regimes, 3) weighted regimes, and 4) weighted regimes where the threshold depends on a binary baseline covariate. Simulations 2) and 4) are those in the main paper.

A.4.1 Thresholding DTRs

The data generating mechanism is given by:

- $x_1 \sim N(0,1)$, $x_2 \sim N(0, z_1 + 0.5x_1)$
- $z_1 \sim Bern(p = expit(1.5x_1))$, $z_2 \sim Bern(p = expit(2x_2 - 0.5z_1))$
- $y = x_1 - (-\theta_{1opt} + x_1)(\mathbb{1}_{\theta_{1opt} > z_1} - z_1) - (-\theta_{2opt} + x_2)(\mathbb{1}_{\theta_{2opt} > z_2} - z_2) + \sqrt{0.5}\epsilon$, $\epsilon \sim N(0,1)$ and $\theta_{1opt}, \theta_{2opt}$ the location of the desired optima. In the single threshold simulation, we have that $\theta_{1opt} = \theta_{2opt} = \theta_{opt}$.

Note that "expit" is the inverse logit function. For the out of sample prediction, we used a population of $n = 10,000$ and $x_1 \sim N(0.6, 1), x_2 \sim N(0.1 + z_1, 1)$. Figure A.1 plots the expected outcome under the DTRs considered in this section.



(a) (b)

Figure A.1: (a) Response surface for single threshold simulation with Normal covariates; (b) Response surface for double threshold simulation with Normal covariates.

Single Threshold Simulation

Table A.8 shows results for single threshold simulation, under a sample size of $n = 500$. This contrasts the sample size of $n = 1000$ shown in Table A.9. Here, $\theta_{opt} = 0.6$, and the value at the optimum is 0. Generally, the results follow the same pattern though with an overall loss of precision corresponding to the reduction in sample size.

Table A.8: Results for single threshold simulation (Normal covariates; $n = 500$; 500 Monte Carlo replicates).

Method	Model	$\hat{\theta}$	Estimated	Coverage	Mean
	Correct		Outcome	Probability	Outcome
			Train Pop.	θ	Test Pop.
Frequentist	None	0.416 (0.110)	0.217 (0.120)	—	0.587 (0.013)
Frequentist	Treat	0.637 (0.189)	0.038 (0.070)	—	0.589 (0.014)
Frequentist	Outcome	0.580 (0.176)	0.015 (0.069)	—	0.591 (0.014)
Frequentist	Both	0.618 (0.159)	0.013 (0.057)	—	0.593 (0.010)
Frequentist	IPW	0.638 (0.183)	0.029 (0.066)	—	0.590 (0.013)
Bayesian	None	0.414 (0.118)	0.232 (0.119)	0.664	0.586 (0.014)
Bayesian	Treat	0.648 (0.201)	0.057 (0.068)	0.976	0.587 (0.015)
Bayesian	Outcome	0.573 (0.188)	0.026 (0.068)	0.980	0.590 (0.016)
Bayesian	Both	0.624 (0.168)	0.021 (0.057)	0.972	0.592 (0.012)
Bayesian	IPW	0.641 (0.196)	0.045 (0.065)	0.976	0.588 (0.015)

Table A.9: Results for single threshold simulation (Normal covariates; $n = 1000$; 500 Monte Carlo replicates).

Method	Model Correct	$\hat{\theta}$	Estimated Outcome Train Pop.	Coverage Probability θ	Mean Outcome Test Pop.
Frequentist	None	0.419 (0.093)	0.209 (0.084)	—	0.588 (0.011)
Frequentist	Treat	0.635 (0.172)	0.024 (0.047)	—	0.591 (0.013)
Frequentist	Outcome	0.599 (0.122)	0.012 (0.044)	—	0.596 (0.006)
Frequentist	Both	0.608 (0.122)	0.010 (0.038)	—	0.596 (0.007)
Frequentist	IPW	0.624 (0.155)	0.018 (0.045)	—	0.593 (0.011)
Bayesian	None	0.418 (0.097)	0.218 (0.083)	0.516	0.588 (0.012)
Bayesian	Treat	0.642 (0.178)	0.038 (0.045)	0.976	0.590 (0.014)
Bayesian	Outcome	0.597 (0.132)	0.018 (0.044)	0.980	0.595 (0.008)
Bayesian	Both	0.611 (0.128)	0.016 (0.038)	0.972	0.596 (0.008)
Bayesian	IPW	0.634 (0.172)	0.030 (0.044)	0.968	0.591 (0.013)

Note: Standard deviations are Monte Carlo standard deviations

We also investigate the results when intermediary covariates are Gamma-distributed as follows: $x_1 \sim Gamma(\alpha = 2, \beta = 2)$, $x_2 \sim Gamma(\alpha = z_1 + 0.5x_1, \beta = 1)$. The known mean outcome under the optimal threshold is 1 in the training population. In the test population, the distribution of intermediary covariates was changed to be $x_1 = Gamma(\alpha = 1.5, \beta = 1)$ and $x_2 = Gamma(\alpha = z_1 + 0.5x_1, \beta = 2)$. The exploration grid was the same as the Normal setup except that the thresholds started at 0.05, given that Gamma covariates are positive. The results mostly parallel the already observed results, see Table A.10 and A.11. Notable is that the resulting credible intervals appear to be slightly more conservative.

Table A.10: Results for single threshold simulation (Gamma covariates; $n = 500$; 500 Monte Carlo replicates).

Method	Model Correct	$\hat{\theta}$	Estimated Outcome Train Pop.	Coverage Probability θ	Mean Outcome Test Pop.
Frequentist	None	0.127 (0.165)	1.065 (0.058)	—	1.444 (0.016)
Frequentist	Treat	0.609 (0.160)	1.024 (0.054)	—	1.489 (0.011)
Frequentist	Outcome	0.578 (0.192)	1.044 (0.084)	—	1.485 (0.014)
Frequentist	Both	0.624 (0.136)	1.020 (0.047)	—	1.490 (0.010)
Frequentist	IPW	0.654 (0.167)	1.044 (0.059)	—	1.486 (0.015)
Bayesian	None	0.181 (0.219)	1.132 (0.062)	0.774	1.450 (0.021)
Bayesian	Treat	0.632 (0.167)	1.092 (0.052)	0.998	1.487 (0.014)
Bayesian	Outcome	0.614 (0.182)	1.155 (0.087)	1	1.486 (0.013)
Bayesian	Both	0.649 (0.138)	1.080 (0.046)	0.998	1.489 (0.011)
Bayesian	IPW	0.706 (0.173)	1.124 (0.061)	0.962	1.482 (0.018)

Table A.11: Results for simulation I (Gamma covariates; $n = 1000$; 500 Monte Carlo replicates).

Method	Model Correct	$\hat{\theta}$	Estimated Outcome Train Pop.	Coverage Probability θ	Mean Outcome Test Pop.
Frequentist	None	0.076 (0.079)	1.051 (0.045)	—	1.439 (0.008)
Frequentist	Treat	0.608 (0.131)	1.015 (0.037)	—	1.491 (0.008)
Frequentist	Outcome	0.587 (0.146)	1.027 (0.065)	—	1.490 (0.009)
Frequentist	Both	0.610 (0.106)	1.010 (0.033)	—	1.493 (0.005)
Frequentist	IPW	0.632 (0.153)	1.025 (0.042)	—	1.488 (0.012)
Bayesian	None	0.106 (0.146)	1.085 (0.044)	0.546	1.442 (0.014)
Bayesian	Treat	0.625 (0.131)	1.062 (0.036)	1	1.491 (0.009)
Bayesian	Outcome	0.608 (0.143)	1.107 (0.065)	1	1.490 (0.009)
Bayesian	Both	0.626 (0.111)	1.052 (0.032)	1	1.492 (0.007)
Bayesian	IPW	0.671 (0.160)	1.080 (0.044)	0.974	1.486 (0.014)

Double Threshold Simulation

In Table A.12 we examine the results of the double threshold simulation with normal covariates and a larger sample size than presented in the main paper. Here, $\theta_{1opt} = 0.4, \theta_{2opt} = 0.8$, and the value at the optimum is 0. There is a general gain in precision due to the larger sample size; additionally, the coverage of the confidence intervals deviates slightly farther from the nominal coverage.

Table A.12: Results for double threshold simulation (Normal covariates; $n = 1000$; 500 Monte Carlo replicates).

Method	Model Correct	$\hat{\theta}_1$	$\hat{\theta}_2$	Estimated Outcome Train Pop.	Coverage Probability θ_1, θ_2	Mean Outcome Test Pop.
Frequentist	None	0.254 (0.097)	0.677 (0.142)	0.236 (0.086)	—	0.591 (0.008)
Frequentist	Treat	0.470 (0.204)	0.788 (0.175)	0.031 (0.045)	—	0.588 (0.015)
Frequentist	Outcome	0.393 (0.164)	0.783 (0.156)	0.016 (0.043)	—	0.593 (0.008)
Frequentist	Both	0.416 (0.152)	0.801 (0.145)	0.013 (0.037)	—	0.594 (0.007)
Frequentist	IPW	0.443 (0.179)	0.790 (0.180)	0.023 (0.044)	—	0.590 (0.012)
Bayesian	None	0.252 (0.104)	0.682 (0.154)	0.250 (0.085)	0.770, 0.918	0.590 (0.008)
Bayesian	Treat	0.473 (0.217)	0.795 (0.179)	0.047 (0.043)	0.970, 0.988	0.587 (0.016)
Bayesian	Outcome	0.390 (0.171)	0.787 (0.179)	0.026 (0.043)	0.986, 0.992	0.591 (0.010)
Bayesian	Both	0.419 (0.159)	0.809 (0.148)	0.021 (0.037)	0.982, 0.982	0.593 (0.008)
Bayesian	IPW	0.456 (0.191)	0.798 (0.183)	0.036 (0.043)	0.978, 0.988	0.589 (0.014)

Note: Standard deviations are Monte Carlo standard deviations

Next, we can examine the results when intermediary covariates are Gamma-distributed as described in the previous section. Tables A.13 and A.14 show the results for this setup. Overall, we observe that the optimal threshold are unbiasedly estimated, and that credible intervals are somewhat conservative leading to higher coverage probabilities. Part of this is due to the choice of increments: larger increments leading to higher coverage. The value at the optimal thresholds is also unbiased.

Table A.13: Results for double threshold simulation (Gamma covariates; $n = 500$; 500 Monte Carlo replicates).

Method	Model Correct	$\hat{\theta}_1$	$\hat{\theta}_2$	Estimated Outcome Train Pop.	Coverage Probability θ_1, θ_2	Mean Outcome Test Pop.
Frequentist	None	0.129 (0.074)	0.751 (0.210)	1.145 (0.076)	—	1.481 (0.005)
Frequentist	Treat	0.379 (0.181)	0.791 (0.173)	1.038 (0.049)	—	1.488 (0.010)
Frequentist	Outcome	0.401 (0.211)	0.757 (0.177)	1.055 (0.068)	—	1.485 (0.014)
Frequentist	Both	0.406 (0.168)	0.792 (0.149)	1.024 (0.043)	—	1.490 (0.009)
Frequentist	IPW	0.456 (0.197)	0.785 (0.188)	1.050 (0.052)	—	1.485 (0.016)
Bayesian	None	0.136 (0.087)	0.757 (0.216)	1.197 (0.072)	0.810, 0.974	1.481 (0.005)
Bayesian	Treat	0.393 (0.190)	0.806 (0.177)	1.099 (0.049)	0.998, 0.964	1.487 (0.012)
Bayesian	Outcome	0.446 (0.208)	0.760 (0.186)	1.131 (0.069)	0.994, 0.988	1.484 (0.017)
Bayesian	Both	0.426 (0.171)	0.807 (0.150)	1.076 (0.043)	1.000, 0.994	1.489 (0.010)
Bayesian	IPW	0.494 (0.213)	0.800 (0.186)	1.128 (0.053)	1.000, 0.952	1.482 (0.021)

Table A.14: Results for double threshold simulation (Gamma covariates; $n = 1000$; 500 Monte Carlo replicates).

Method	Model Correct	$\hat{\theta}_1$	$\hat{\theta}_2$	Estimated Outcome Train Pop.	Coverage Probability θ_1, θ_2	Mean Outcome Test Pop.
Frequentist	None	0.109 (0.034)	0.780 (0.163)	1.131 (0.060)	—	1.480 (0.002)
Frequentist	Treat	0.386 (0.150)	0.805 (0.150)	1.024 (0.037)	—	1.491 (0.006)
Frequentist	Outcome	0.383 (0.180)	0.790 (0.140)	1.035 (0.052)	—	1.489 (0.009)
Frequentist	Both	0.401 (0.125)	0.809 (0.128)	1.014 (0.032)	—	1.493 (0.005)
Frequentist	IPW	0.419 (0.150)	0.784 (0.164)	1.030 (0.040)	—	1.490 (0.009)
Bayesian	None	0.110 (0.038)	0.785 (0.175)	1.159 (0.058)	0.542 0.980	1.480 (0.003)
Bayesian	Treat	0.387 (0.162)	0.821 (0.152)	1.066 (0.036)	1.000 0.970	1.490 (0.007)
Bayesian	Outcome	0.426 (0.188)	0.786 (0.151)	1.088 (0.053)	1.000 0.990	1.488 (0.013)
Bayesian	Both	0.408 (0.136)	0.819 (0.127)	1.049 (0.031)	1.000 0.990	1.492 (0.006)
Bayesian	IPW	0.451 (0.164)	0.805 (0.168)	1.082 (0.040)	1.000 0.962	1.489 (0.011)

An analogous individualized decision rule graph can be produced for the thresholds in this simulation, however this is no more instructive than the figure for the single threshold rule.

A.4.2 Weighted DTRs Simulation

Next, we explore one additional simulation. For this family of regimes, patients are treated in stage one if $\psi_1 x_{1,1} + \psi_2 x_{1,2} > 0.5$ and in stage two if $\psi_1 x_{2,1} + \psi_2 x_{2,2} > 0.5$. Here, $\psi_1, \psi_2 > 0$ such that $\psi_1 + \psi_2 = 1$. The optimal parameters are chosen to be $\psi_{1opt} = \psi_{2opt} = 0.5$. The response surface is this setting is similar to that in Simulation II. The data generating mechanism proceeds as follows:

- $x_{1,1} \sim N(1,1)$, $x_{1,2} \sim N(0,1)$
- $z_1 \sim Bern(expit(1.5x_{1,2} + 2x_{1,1}))$
- $x_{2,1} \sim N(0.2z_1 + 0.1x_{1,1}, 1)$, $x_{2,2} \sim N(0.5z_1 + 0.1x_{1,2}, 1)$
- $z_2 \sim \text{Bern}(p = expit(1.5x_{2,2} - 0.6z_1 + 2x_{2,1}))$
- $z_{1,opt} = 0.5x_{1,1} + 0.5x_{1,2} > 0.5$, $z_{2,opt} = 0.5x_{2,1} + 0.5x_{2,2} > 0.5$,
- $y = x_{11} + x_{12} - (0.5x_{11} + 0.5x_{12} - 0.5)(z_{1,opt} - z_1) - (0.5x_{21} + 0.5O_{22} - 0.5)(z_{2,opt} - z_2) + \sqrt{0.5}\epsilon$,

  $\epsilon \sim N(0,1)$

The value at the optimal threshold can bee seen to be 1. For the test population, we used a population size of $n = 10,000$ and $x_{1,1} \sim N(0.1,1), x_{1,2} \sim N(0.5,1), x_{2,1} \sim N(0.1 + 0.2z_1 + 0.1x_{1,1}, 1), x_{2,2} \sim N(0.5 + 0.5z_1 + 0.1x_{1,2}, 1)$. Results are presented in the Table A.15 and A.16, and we observed that we obtain unbiased results. Surprisingly, even with both nuisance models misspecified, the estimator performs quite well in terms of coverage, though it estimates the outcome under the optimal regime with high bias. Note that although there are two-parameters in this decision rule, the condition that $\psi_1 + \psi_2 = 1$, makes it so that it is enough to evaluate the coverage probability of only one parameter; this is also why the Monte Carlo standard errors in the ψ_1, ψ_2 columns are the same.

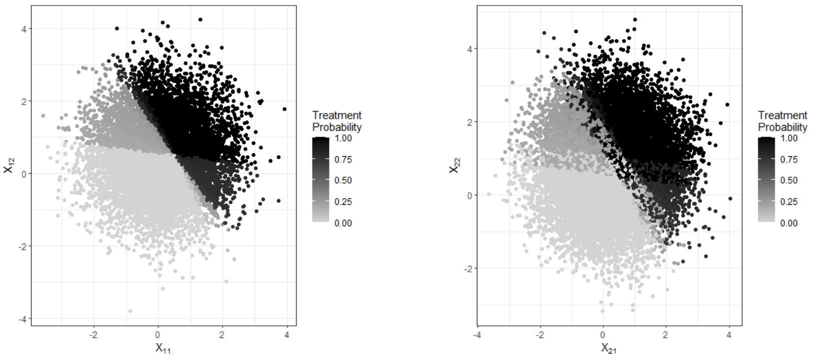
Table A.15: Frequentist and Bayesian results ($n = 500$; 500 Monte Carlo replicates).

Method	Model Correct	$\hat{\psi}_1$	$\hat{\psi}_2$	Estimated Outcome Train Pop.	Coverage Probability ψ_1	Mean Outcome Test Pop.
Frequentist	None	0.719 (0.251)	0.282 (0.251)	0.208 (0.206)	—	0.509 (0.078)
Frequentist	Treat	0.477 (0.193)	0.523 (0.193)	1.110 (0.274)	—	0.551 (0.045)
Frequentist	Outcome	0.518 (0.122)	0.482 (0.122)	1.022 (0.096)	—	0.571 (0.027)
Frequentist	Both	0.474 (0.117)	0.526 (0.117)	1.038 (0.124)	—	0.571 (0.027)
Frequentist	IPW	0.464 (0.208)	0.536 (0.208)	1.215 (0.559)	—	0.545 (0.049)
Bayesian	None	0.754 (0.258)	0.247 (0.258)	0.258(0.201)	0.944	0.496 (0.078)
Bayesian	Treat	0.473 (0.199)	0.527 (0.199)	1.142(0.272)	0.964	0.550 (0.048)
Bayesian	Outcome	0.523 (0.131)	0.477 (0.131)	1.034(0.095)	0.980	0.569 (0.030)
Bayesian	Both	0.476 (0.119)	0.524 (0.119)	1.047(0.119)	0.968	0.571 (0.027)
Bayesian	IPW	0.460 (0.211)	0.540 (0.211)	1.264(0.563)	0.950	0.544 (0.050)

Table A.16: Frequentist and Bayesian results ($n = 1000$; 500 Monte Carlo replicates).

Method	Model Correct	$\hat{\psi}_1$	$\hat{\psi}_2$	Estimated Outcome Train Pop.	Coverage Probability ψ	Mean Outcome Test Pop.
Frequentist	None	0.731 (0.249)	0.269 (0.249)	0.184 (0.140)	—	0.507 (0.079)
Frequentist	Treat	0.462 (0.170)	0.538 (0.170)	1.061 (0.140)	—	0.557 (0.041)
Frequentist	Outcome	0.521 (0.104)	0.479 (0.104)	1.015 (0.062)	—	0.575 (0.024)
Frequentist	Both	0.469 (0.109)	0.531 (0.109)	1.021 (0.066)	—	0.573 (0.026)
Frequentist	IPW	0.454 (0.190)	0.545 (0.190)	1.117 (0.242)	—	0.551 (0.046)
Bayesian	None	0.759 (0.250)	0.241 (0.250)	0.217 (0.136)	0.958	0.497 (0.079)
Bayesian	Treat	0.455 (0.173)	0.545 (0.173)	1.086 (0.138)	0.988	0.556 (0.042)
Bayesian	Outcome	0.526 (0.112)	0.474 (0.112)	1.023 (0.062)	0.972	0.573 (0.027)
Bayesian	Both	0.471 (0.107)	0.529 (0.107)	1.029 (0.064)	0.970	0.573 (0.025)
Bayesian	IPW	0.448 (0.193)	0.552 (0.193)	1.156 (0.236)	0.968	0.549 (0.047)

Now, we may examine the individualized inference for this scenario. Figure A.2 shows us that there are combinations of x_{k1}, x_{k2} where there is high certainty about following the optimal regime and areas of low certainty.



(a) (b)

Figure A.2: Weighted DTR simulation individualized treatment probabilities using doubly robust estimator; (a) Stage 1 treatment (b) Stage 2 treatment.

A.4.3 Weighted DTRs with Binary Covariate Simulation

Here, we provide the details for simulation II in the main paper. In this family of regimes, patients are treated if $\psi_1 x_{k1} + \psi_2 x_{k2} > 0.5 - 3\psi_3 u, k = 1, .., 4$, where $\psi_1 + \psi_2 = 1$, $\psi_1, \psi_2 > 0$. The exploration grid is given by $\psi_1, \psi_2 \in [0.2, 0.8]$ in increments of 0.05 and $\psi_3 \in$ [-0.3,0.3] in increments of 0.1. This yields a grid of 91 points, with known optima $\psi_{1opt} = 0.5, \psi_{2opt} = 0.5, \psi_{3opt} = 0.1$. The specific data generating mechanism used is given by:

- $x_{11} \sim N(1, 1)$, $x_{12} \sim N(0, 1)$, $u \sim \cdot Bern(0.5)$, $z_1 \sim Bern(expit(0.5x_{12} + x_{11}))$
- $x_{k1} \sim N(0.2z_{k-1} + 0.1x_{k-1,1}, 1)$, $x_{k2} \sim N(0.5z_{k-1} + 0.1x_{k-1,2}, 1), k = 2, 3, 4$
- $z_k \sim \text{Bern}(p = expit(0.5x_{k2} - 0.6z_{k-1} + x_{k1})), k = 2, 3, 4$
- $z_{k,opt} = 0.5x_{k1} + 0.5x_{k2} + 0.3u > 0.5, k = 1, .., 4$
- $y = x_{11} + x_{12} - \sum_{k=1}^{4}(0.5x_{k1} + 0.5x_{k2} + 0.3u - 0.5)(z_{k,opt} - z_k) + \sqrt{0.1}\epsilon,\ \epsilon \sim N(0, 1)$

Table A.17 shows the results for a sample size of $n = 1000$. Generally, we observe a gain in precision as compared to the $n = 500$ table in the main paper. Additionally, we note that when all models are correct, we estimate ψ_{3opt} very well. This reflects the fact that the value function is more peaked in this direction as compared to other parameters. For the test population, x_{k1}, x_{k2} were shifted by 0.1 and 0.5, respectively and $u \sim Bern(0.7)$; we observe that the doubly robust estimator yields the highest value, as expected.

Table A.17: Frequentist and Bayesian results ($n = 1000$; 500 Monte Carlo replicates).

Method	Model Correct	$\hat{\psi}_1$	$\hat{\psi}_3$	Estimated Outcome Train Pop.	Coverage Probability ψ_1, ψ_3	Mean Outcome Test Pop.
Frequentist	None	0.570 (0.099)	0.092 (0.091)	1.871 (0.282)	—	0.546 (0.051)
Frequentist	Treat	0.472 (0.136)	0.107 (0.101)	1.096 (0.111)	—	0.544 (0.051)
Frequentist	Outcome	0.503 (0.040)	0.100 (0.004)	1.002 (0.048)	—	0.583 (0.006)
Frequentist	Both	0.502 (0.025)	0.100 (0.000)	0.999 (0.045)	—	0.585 (0.004)
Frequentist	IPW	0.478 (0.137)	0.099 (0.110)	1.120 (0.139)	—	0.543 (0.051)
Bayesian	None	0.571 (0.108)	0.097 (0.085)	1.995 (0.272)	0.95 0.996	0.558 (0.021)
Bayesian	Treat	0.465 (0.133)	0.105 (0.103)	1.164 (0.100)	0.986 1	0.547 (0.026)
Bayesian	Outcome	0.501 (0.036)	0.100 (0.000)	1.006 (0.049)	0.992 1	0.590 (0.003)
Bayesian	Both	0.499 (0.022)	0.100 (0.000)	1.001 (0.045)	1 1	0.592 (0.002)
Bayesian	IPW	0.459 (0.142)	0.102 (0.105)	1.206 (0.117)	0.984 1	0.544 (0.026)

A.5 Details of the NA-ACCORD Analysis

In what follows, we describe the procedure used to create the data, the analysis plan, the specific models utilized, and we address questions of positivity, individualized inference, and balance.

A.5.1 Data Creation

Study Start: Study initiation (time zero) is the first instance of ART treatment on or after 2004 in the NA-ACCORD database.

- Study start is not enrollment date as many patients have a long lag between cohort enrollment and ART initiation.

Censoring: Last ART record that has continuous follow-up from study start and that has CD4 and viral load measurements available. This entails the following:

1. There is a monthly ART record from month one up until the month of study exit.

    - Note: some patients have no records for several months and then continuous follow-up resumes. Study exit for these patients is the last month of the first instance of continuous follow-up.

    - There is one exception to the above: If patients have four or fewer months of ART records missing and then continuous follow-up begins again, these months are filled with the last observed treatment. This approach is reasonable as patients do not switch treatment very often.

2. Each record can be associated with a viral load and CD4 cell count measurement.

    - Associate each ART record with CD4 and viral load measurement by taking closest measurement date to ART record date, and using last observation carried forward.

    - With the exception of missing baseline lab values, patients who have missing lab values are censored at the first instance of missingness.

    - Patients who have missing lab values at study start are kept in the study and we create a *status* variable which indicates baseline missingness.

Stage-specific Censoring Details:

- **Stage 1:** Patients lost to follow up after stage 1 covariates are observed but before stage 2 covariates are observed are censored at stage 1.

- **Stage 2:** Patients lost to follow up after stage 2 covariates are observed but before stage 3 covariates are observed are censored at stage 2.

- **Stage 3:** Patients lost to follow-up after stage 3 covariates are observed but before final outcome is observed are censored at stage 3.

Study End: Study end is 18 months after study start; the outcome FIB4 is taken to be the first FIB4 measurement recorded after study end, within 12 months.

Details of the follow-up can be observed in the following diagram:

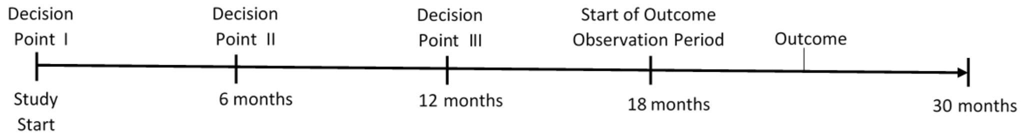


Figure A.3: Study Stages

Treatment: We dichotomize all ART treatments to PI based or another ART medication. Some patients receive dual therapy in combination with PI; these are included in "other ART" group.

Treatment Decisions: We consider 6-month observation intervals thereby leading to three treatment decision points: one in the first month of the study, one in the 7th month of the study, and one at 13th month.

Augmented Data Creation

The regimes we explore are of the family: start on a non-PI based ART therapy and switch into PI when $FIB4 > \theta$. Refer to a dataset with information about patients adhering to a regime in this family by R_θ. In addition to the censoring described in the section above, we must take care to keep track of artificial censoring in R_θ. A patient is artificially censored with respect to a regime with threshold θ when they stop adhering to the regime. If they never adhere to the regime, then they are artificially censored at baseline. Adherence to R_θ can be determined based on the following category of patients:

1. Indicated to Switch but did Not Switch (ISNS): Artificial censoring at Indicated switch date.

2. Indicated to Switch and Switched (ISS): No artificial censoring. If patient switches more than once during the study period, then they are artificially censored at the time

of their second switch.

3. Not Indicated to Switch and did Not Switch (NISNS): No artificial censoring.

4. Not Indicated to Switch but Switched (NISS): Artificial censoring at switch date.

5. No Regime (NR): Initial therapy was PI; artificial censoring at baseline.

Note on creating R_θ:

- Each R_θ dataset will contain all patients in the study population. Even patients who are artificially censored at baseline will contribute to fitting outcome models, and toward the fit of the doubly robust estimator.

- To determine the θ that will be used in the data augmentation, look at the distribution of FIB4 measurements at baseline and create equally spaced increments of 0.2. Based on the data, it turned out that the starting value was 0.4.

Final Datasets: At the end of the above data creation we should have two datasets:

- DATA in long format constitutes of patients in the study population up until their censoring or the study end date. This dataset does not contain any variables that reference regime adherence.

- AUGDATA is the stacked R_θ datasets. Each R_θ datasets is a long-format dataset of patients who adhere to regime R_θ with threshold θ, for the full follow-up period. Each of these dataset have an additional variable providing the regime *index* θ.

A.5.2 Analysis

For simplicity, we first describe the frequentist analysis, and then describe the Bayesian adaptation.

Treatment Propensity Models: Use DATA to fit a logistic regression model for each stage. Possible time-varying confounders include CD4 cell count and Viral load. These variables are certainly used to assign treatment, and were proxies for level of HIV infection. There is some evidence to suggest that HIV is associated with decreased liver health. Therefore, these variables may also mediate previous treatment effects.

Censoring Models: Fit censoring models for each decision point.

Outcome Models: The first conditions on baseline information; the second conditions on information up to stage 2; the third conditions on information up stage 3.

Weight Construction:

- Estimate stage-specific treatment and censoring models.
- For all patients in AUGDATA use the treatment propensity model to compute the probability that they received their observed treatment at each time point.
- Invert each of these probabilities to obtain a weight for each patient for each decision point. Collapse AUGDATA into one observation per patient per regime, and multiply all patient weights in order to create a final weight variable for each patient.

Inverse Probability Weighting Analysis: This analysis is only performed on the subset of cases who are neither censored nor artificially censored. Fit a weighted regression with FIB4 as the outcome and with regime *index* as the predictor. The weights are the ones calculated in the above step. This fit yields the normalized IPW estimator.

Doubly Robust Analysis: Make use of doubly robust estimator. This estimator makes used of all observations censored or uncensored (up to the censoring point).

Bayesian Inference Adaptation:

- Draw a vector of Dirichlet weights for as many patients as in DATA. Assign one of these weights to each patient by adding a Dirichlet weight variable to DATA. Note that this variable will not have variation within patients. Additionally, merge these weights into AUGDATA.
- Fit the treatment propensity, censoring, and outcome models as above, but this time

incorporate the Dirichlet weights into the fitting. Construct the weights for the collapsed data as before, using the predictions from the treatment propensity model.

- For the IPW analysis, fit the marginal mean model by multiplying the final weights in the collapsed AUGDATA by the Dirichlet weights of each person in AUGDATA.
- For the doubly robust analysis, run regression, where outcome for each patient is the person-specific contribution to equation (12) in the main paper, and where the predictor is *index*.
- Repeat this over many iterations in order to obtain the posterior distribution of interest.

Analysis Models:

We now specify the models used for analysis.

Censoring Models:

$$\begin{aligned} Stage\ 1 : &\ status + status \times rcs(log(CD4)) + AgeBaseline + Insurance + AtRiskAlcohol \\ &+ Sex + Smoking + DrugUse + Race + CalendarYear \end{aligned}$$

$$Stage\ 2 : rcs(log(CD4)) + AtRiskAlcohol + Smoking + DrugUse + Race + CalendarYear$$

$$Stage\ 3 : rcs(log(CD4)) + AtRiskAlcohol + Smoking + DrugUse + Race + CalendarYear$$

Treatment Models:

$$\begin{aligned} Stage\ 1 : &\ status + status \times rcs(log(CD4)) + AgeBaseline + Insurance + AtRiskAlcohol \\ &+ Sex + HCV + Race + CalendarYear \end{aligned}$$

$$Stage\ 2 : rcs(log(CD4)) + Sex + Insurance + HCV + Stage1Treat + Race + CalendarYear$$

$$Stage\ 3 : rcs(log(CD4)) + Sex + Insurance + HCV + Stage2Treat + Race + CalendarYear$$

Note: *rcs* denotes a restricted cubic spline; Stage1Treat denotes stage 1 treatment and Stage2Treat denotes stage 2 treatment. Some patients have missing lab values at baseline; this is indicated by the *status* variable in the models above.

Outcome Models:

$$\begin{aligned}
Stage\ 1: &\ index + index \times (Sex + AgeBaseline + Smoking + DrugUse + HBV + HCV \\
&+ Insurance + Treat + status \times rcs(log(CD4)) + status \times rcs(log(ViralLoad))) \\
Stage\ 2: &\ index + index \times (Sex + AgeBaseline + Smoking + DrugUse + HBV + HCV \\
&+ Insurance + Stage1Treat + Treat + rcs(log(CD4)) + rcs(log(ViralLoad))) \\
Stage\ 3: &\ Sex + AgeBaseline + Smoking + DrugUse + HBV + HCV + Insurance \\
&+ Stage1Treat + Stage2Treat + Treat + rcs(log(CD4)) + rcs(log(ViralLoad))
\end{aligned}$$

Note: The *index* variable in the models above is fit as a categorical variable, denoting the regime index.

Sensitivity Analyses: The following sensitivity analyses were performed:

- Sensitivity Analysis I: All models the same, except that outcome model restricted cubic splines are replaced with $log(CD4)$ and $log(ViralLoad)$ terms.
- Sensitivity Analysis II: All models the same, except for outcome model restricted cubic splines are replaced with $rcs(log(TimeBetween \times CD4))$ and $rcs(log(TimeBeteween \times ViralLoad))$ terms. This model attempts to account for the fact that not all lab measurements are taken within the same amount of time of the decision point.
- Sensitivity Analysis III: All models the same, except for outcome model restricted cubic splines are replaced with $log(TimeBetween \times CD4)$ and $log(TimeBeteween \times ViralLoad)$ terms.

Conclusion of sensitivity analysis: results changed only minimally across models.

A.5.3 Positivity

Two types of positivity violations are of concern: structural positivity and practical positivity [Petersen et al., 2012]. The former refers to when patients with specific sets of characteristics are precluded from receiving a treatment; we do not think this is an issue here. The latter refers to the fact that we do not observe all treatments covariate combinations, due to a finite sample size. This is of concern in our setting, as therapeutic switches were infrequent. Zhu et al. [2021] mention that if propensity scores (PS) are used for achieving balance, then the focus should be on assessing PS overlap between treatment groups. We assessed positivity *for each candidate regime* by checking whether the distribution of the propensity score at each interval for the modeled treatment are similar in the regime adherent group and the regime non-adherent group. This must be done separately for each regime of interest (each θ). In the first stage, all regimes start by evaluating the hypothetical world in which all patients start on a non-PI regimen. Therefore, at this stage the treatment was the probability of receiving PI. For this reason, we only need to perform one comparison across all regimes for this stage (there is no dependence on θ at this stage). We observe that there is overlap from Table A.18. For the second and third stage, the propensity of interested was in those who switched treatment. Therefore, we compared the probability that a patient switched into PI in the adherent group vs. the non-adherent group; these comparisons are specific to a threshold θ and are presented for a subset of regimes in Table A.18. Propensity score overlap indicated that patients who adhered have similar covariate distributions to those who did not adhere. Therefore the types of patients who switch in the regime-enforced world are well represented in the observational world. The propensity to switch treatment was generally small, highlighting that relatively few individuals contribute to the estimation of our regime of interest – a limitation that must be acknowledged.

Table A.18: Propensity score overlap between patients who adhered to a specific regime and patients who did not adhere for a subset of regimes. (Adh.="Adherent").

Regime θ	Group	0%	10%	25%	50%	75%	90%	100%
	Adh. Stage 1	0.198	0.440	0.515	0.606	0.693	0.747	0.835
	Non-Adh. Stage 1	0.179	0.383	0.453	0.537	0.628	0.698	0.820
0.4	Adh. Stage 2	0.012	0.018	0.023	0.037	0.056	0.071	0.119
0.4	Non-Adh. Stage 2	0.011	0.018	0.023	0.035	0.052	0.067	0.130
0.4	Adh. Stage 3	0.006	0.010	0.013	0.018	0.030	0.041	0.050
0.4	Non-Adh. Stage 3	0.006	0.010	0.012	0.017	0.027	0.036	0.062
1.0	Adh. Stage 2	0.012	0.017	0.022	0.033	0.051	0.067	0.123
1.0	Non-Adh. Stage 2	0.011	0.019	0.024	0.037	0.053	0.070	0.130
1.0	Adh. Stage 3	0.006	0.010	0.013	0.018	0.027	0.037	0.062
1.0	Non-Adh. Stage 3	0.006	0.011	0.013	0.020	0.029	0.037	0.062
1.6	Adh. Stage 2	0.011	0.018	0.022	0.034	0.051	0.066	0.123
1.6	Non-Adh. Stage 2	0.012	0.020	0.026	0.039	0.056	0.073	0.130
1.6	Adh. Stage 3	0.006	0.010	0.013	0.019	0.028	0.037	0.062
1.6	Non-Adh. Stage 3	0.007	0.011	0.014	0.021	0.030	0.039	0.062
2.2	Adh. Stage 2	0.011	0.018	0.022	0.034	0.051	0.067	0.123
2.2	Non-Adh. Stage 2	0.012	0.021	0.028	0.041	0.059	0.075	0.130
2.2	Adh. Stage 3	0.006	0.010	0.013	0.019	0.028	0.037	0.062
2.2	Non-Adh. Stage 3	0.007	0.011	0.015	0.022	0.032	0.040	0.061
2.8	Adh. Stage 2	0.011	0.018	0.022	0.034	0.051	0.066	0.123
2.8	Non-Adh. Stage 2	0.012	0.021	0.029	0.043	0.062	0.077	0.130
2.8	Adh. Stage 3	0.006	0.010	0.013	0.019	0.028	0.037	0.062
2.8	Non-Adh. Stage 3	0.007	0.012	0.016	0.023	0.032	0.042	0.061

A.5.4 Normalization of Weights

In real data analyses, the variability of the estimators is an important consideration. One approach to arrive at more robust estimates is to use weight normalization, as this can reduce the variability of the resulting weights. A discussion of weight normalization can be found in Chapter 12 of Hernán and Robins [2020], and it has been further explored in the literature for example in Xiao et al. [2010]. For a sample of Dirichlet weights $\pi = (\pi_1, ..., \pi_n)$, the normalized IPW estimator for the value of a regime g^r is:

$$\frac{\sum_{i=1}^{n} \frac{\pi_i \mathbb{1}_{g^r(\bar{x}_i)}(\bar{z}_i) y_i}{\prod_{j=1}^{K} p_{\mathcal{O}}(z_{ij}|\bar{z}_{ij-1},\bar{x}_{ij})}}{\sum_{i=1}^{n} \frac{\pi_i \mathbb{1}_{g^r(\bar{x}_i)}(\bar{z}_i)}{\prod_{j=1}^{K} p_{\mathcal{O}}(z_{ij}|\bar{z}_{ij-1},\bar{x}_{ij})}}$$

Taking the expectation in the numerator and the denominator across Π, yields the familiar frequentist estimator. The same approach can be taken with the weights in the doubly robust estimator.

A.5.5 Balance Diagnostics

Next, we assess the balance obtained from the resulting weighting We used standardized mean differences to assess balance. Table A.20 shows the treatment balance assessment at each stage, using the full weights. Some standardized mean differences are moderately large, even after weighting, but this must be considered in the context of having a finite sample size and several probabilities contributing to the weighting of each observation.

A.5.6 Results for Individualized Inference

By looking at Figure 3 in the main paper, it may be tempting to conclude that there is no benefit to tailoring. This is actually not the case. We remind the reader that we are after the computation: $\theta_{min} = \arg\min(E_{\theta_1}[Y], ..., E_{\theta_{13}}[Y])$. From Figure A.4, we note that across draws of Π, the expected outcome under regime θ follows a predictable pattern. That is, for small θ the outcome tends to be lower than for high values of θ. We conclude that Figure 3 in the main paper does not display all necessary information.

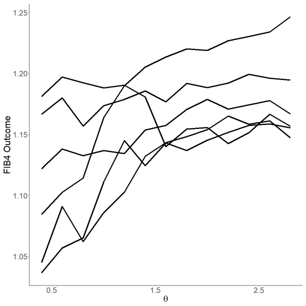


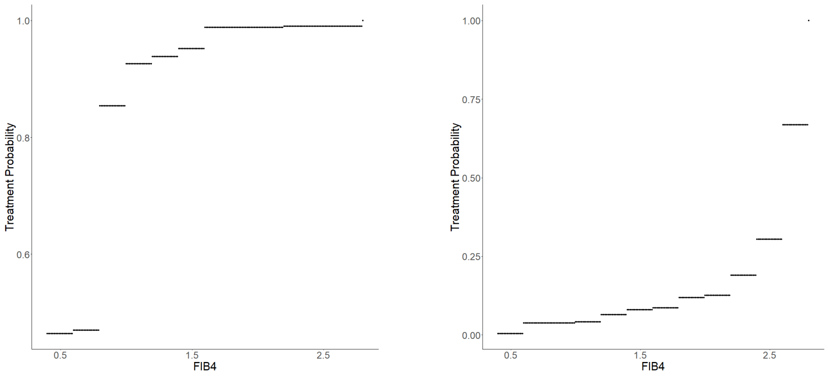
Figure A.4: Values for six different samples of the posterior distribution

To enrich our analysis, we consider the posterior distribution of two types of θ: one is θ_{min}, which was the original target and which is thought to minimize end-stage FIB4; the second is θ_{max}, which corresponds to the worst decision rule we can obtain by maximizing end-stage FIB4. We now see from Table A.19 that the outcome-minimizing and outcome-maximizing threshold are not equiprobable. Consequently this does allow us to consider individualized inference, though we should realize that even if we can identify an optimal threshold, it is still clear that the expected change in final FIB4 is minimal and therefore the resulting optimal decision rule will have limited clinical value.

Table A.19: Posterior Distribution of outcome minimizing/maximizing regimes (500 posterior draws).

Threshold	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2	2.2	2.4	2.6	2.8
θ_{min}	232	3	192	36	6	7	18	0	0	1	0	0	5
θ_{max}	2	17	0	2	11	8	3	16	4	32	57	182	166

From Figure A.5(a), we see that the when a patient's FIB4 score is at 0.8 or greater, they should switch into PI if they hope to follow optimal therapy. From figure A.5(b), we see that we should be careful regarding when to switch into PI. Operationalizing a rule that says switch when FIB4 is greater than 2.6 means that we might actually be following the least optimal regime. Of course, we remind the reader that the difference in effect size that each of these regimes yield is small.



(a)

(b)

Figure A.5: Cases study individualized treatment probabilities using doubly robust estimator; (b) Treatment based on θ_{min} (a) Treatment based on θ_{max}.

Table A.20: Balance diagnostics on the weighted sample: NA-ACCORD.

Stage 1	No PI	PI	SMD
n	7438.8	5182.2	
Smoking (%)	4112.9 (55.3)	2888.7 (55.7)	0.009
At Risk Alcohol (%)	1971.3 (26.5)	1478.1 (28.5)	0.045
Drug Use (%)	1495.4 (20.1)	992.7 (19.2)	0.024
Sex (%)	1258.4 (16.9)	1240.3 (23.9)	0.175
Age at Baseline (mean (SD))	40.07 (11.05)	40.89 (10.56)	0.076
Race Group(%)			0.067
Black	2874.1 (38.6)	1875.4 (36.2)	
Missing	533.6 (7.2)	444.5 (8.6)	
Other	405.3 (5.4)	273.6 (5.3)	
White	3625.7 (48.7)	2588.7 (50.0)	
Insurance (%)	3148.4 (42.3)	1946.2 (37.6)	0.097
HCV at Baseline (%)	736.3 (9.9)	772.8 (14.9)	0.153
HBV at Baseline (%)	381.5 (5.1)	359.6 (6.9)	0.076
Stage 2	No PI	PI	SMD
n	7138.2	5482.8	
Smoking (%)	3921.4 (54.9)	2921.8 (53.3)	0.033
At Risk Alcohol (%)	1871.6 (26.2)	1466.8 (26.8)	0.012
Drug Use (%)	1381.3 (19.4)	900.8 (16.4)	0.076
Sex (%)	1267.1 (17.8)	1333.8 (24.3)	0.162
Age at Baseline (mean (SD))	40.77 (11.02)	41.41 (10.37)	0.060
Race Group (%)			0.100
Black	2792.0 (39.1)	2063.4 (37.6)	
Missing	500.8 (7.0)	535.6 (9.8)	
Other	386.5 (5.4)	301.1 (5.5)	
White	3458.9 (48.5)	2582.8 (47.1)	
Insurance (%)	2959.3 (41.5)	1975.0 (36.0)	0.112
HCV at Baseline (%)	723.5 (10.1)	847.7 (15.5)	0.160
HBV at Baseline (%)	366.9 (5.1)	398.0 (7.3)	0.088
Stage 3	No PI	PI	SMD
n	7156.6	5464.4	
Smoking (%)	3946.3 (55.1)	2895.3 (53.0)	0.043
At Risk Alcohol(%)	1863.6 (26.0)	1445.0 (26.4)	0.009
druguse (%)	1393.1 (19.5)	884.4 (16.2)	0.086
Sex(%)	1287.5 (18.0)	1365.2 (25.0)	0.171
Age at Baseline (mean (SD))	40.78 (11.00)	41.24 (10.34)	0.043
Race Group (%)			0.105
Black	2792.9 (39.0)	2111.3 (38.6)	
Missing	504.9 (7.1)	541.4 (9.9)	
Other	391.9 (5.5)	298.8 (5.5)	
White	3466.9 (48.4)	2513.0 (46.0)	
Insurance(%)	3010.2 (42.1)	1940.8 (35.5)	0.135
HCV at Baseline (%)	732.6 (10.2)	844.8 (15.5)	0.157
HBV at Baseline (%)	372.2 (5.2)	387.4 (7.1)	0.079

A.6 Acknowledgements

The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The NA-ACCORD was supported by National Institutes of Health grants U01AI069918, F31AI124794, F31DA037788, G12MD007583, K01AI093197, K01AI131895, K23EY013707, K24AI065298, K24AI118591, K24DA000432, KL2TR000421, N01CP01004, N02CP055504, N02CP91027, P30AI027757, P30AI027763, P30AI027767, P30AI036219, P30AI050409, P30AI050410, P30AI094189, P30AI110527, P30MH62246, R01AA016893, R01DA011602, R01DA012568, R01 AG053100, R24AI067039, U01AA013566, U01AA020790, U01AI038855, U01AI038858, U01AI068634, U01AI068636, U01AI069432, U01AI069434, U01DA03629, U01DA036935, U10EY008057, U10EY008052, U10EY008067, U01HL146192, U01HL146193, U01HL146194, U01HL146201, U01HL146202, U01HL146203, U01HL146204, U01HL146205, U01HL146208, U01HL146240, U01HL146241, U01HL146242, U01HL146245, U01HL146333, U24AA020794,U54MD007587, UL1RR024131, UL1TR000004, UL1TR000083, Z01CP010214 and Z01CP010176; contracts CDC-200-2006-18797 and CDC-200-2015-63931 from the Centers for Disease Control and Prevention, USA; contract 90047713 from the Agency for Healthcare Research and Quality, USA; contract 90051652 from the Health Resources and Services Administration, USA; grants CBR-86906, CBR-94036, HCP-97105 and TGF-96118 from the Canadian Institutes of Health Research, Canada; Ontario Ministry of Health and Long Term Care; and the Government of Alberta, Canada. Additional support was provided by the National Institute Of Allergy And Infectious Diseases (NIAID), National Cancer Institute (NCI), National Heart, Lung, and Blood Institute (NHLBI), Eunice Kennedy Shriver National Institute Of Child Health & Human Development (NICHD), National Human Genome Research Institute (NHGRI), National Institute for Mental Health (NIMH) and National Institute on Drug Abuse (NIDA), National Institute On Aging (NIA), National Institute Of Dental & Craniofacial Research (NIDCR), National Institute Of Neurological Disorders And Stroke (NINDS), National Institute Of Nursing Research (NINR), National Institute on Alcohol Abuse and Alcoholism (NIAAA), National Institute on Deafness and Other Communication Disorders (NIDCD), and National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK).

APPENDIX B

Appendix to Manuscript 2

B.1 Heteroskedasticity

In this section, we further explore possible reasons as to why heteroskedasticity can arise in the estimation surface. For all examples related to this exploration, we focus on regimes of the form treat if "$x > \psi$". As described in the main text, when going from regime g^{ψ_1} to regime g^{ψ_2} ($\psi_1 < \psi_2$), only treated patients can become non-compliant and only untreated patients can become compliant.

B.1.1 Heteroskedasticity Type I

The first source of heteroskedasticity relates to the fact that patient-level responses can exhibit substantially higher levels of variability than the estimation surface. We consider the following data generating mechanism for the purposes of illustration.

- $n = 10000$, $x \sim U(-1, 1)$, $z_1 \sim bin(1, 0.5)$
- $Y = 1000(-x + (x + 0.8)x(x - 0.2)(x - 0.4)(x - 0.8)z_1)$

Note that $y|x_1, z_1$ is deterministically generated so as to avoid other sources of variability.

In panel (a) of Figure B.1, we see the mean response for the treated and untreated patients, as well as the value function. We note that in this scale, the value function looks flat, but it is actually multimodal. In panel (b) of Figure B.1, we observe the normalized IPW-surface, and we see that there are clear signs of heteroskedasticity. Most notably, there is much more variability in the edges of the plot than in the middle. This variability depends mainly on three elements: 1) how close are the treated and untreated lines to the IPW surface, 2) how many patients become compliant, and 3) how many patients become non-compliant.

It is clear from Figure B.1 that the IPW-surface exhibits less variability around zero. This corresponds to when the treated and untreated lines are closest to the IPW-surface. Conversely, when the lines are far from the IPW surface, more variability is observed. We may ask ourselves why there should be more variability if, in some cases, the treated line and the untreated lines are proximal to each other. In an unconfounded setting, for an increase in ψ, the number of lost non-compliers is the same as the number of new compliers, on average. However, for a finite sample there may be more/less new compliers than non-compliers, and it is these differences that lead to variability in the curve, even if treated and untreated response lines are similar. This source of heteroskedasticity is mainly a finite sample consideration; as more patients contribute to the IPW estimator, it becomes harder to shift the mean by a large amount when patients become compliant/non-compliant.

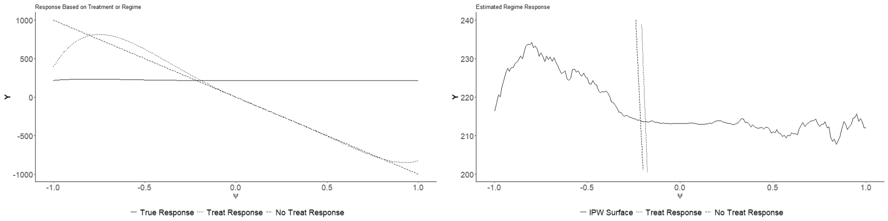


Figure B.1: Heteroskedasticity Type I: (a) Value surface and treated/untreated response curves (b) IPW-surface with treated/untreated response

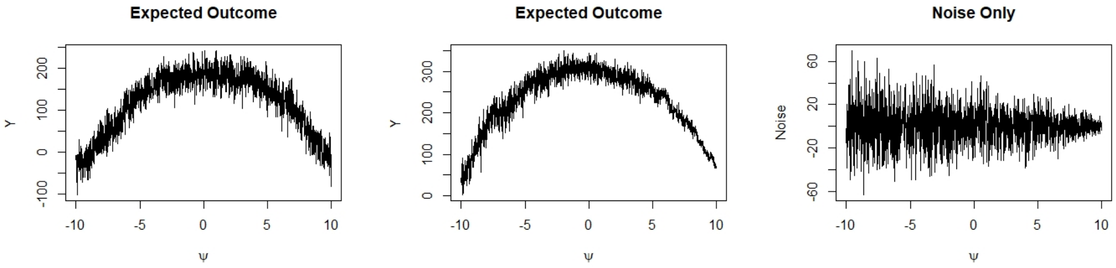
B.1.2 Heteroskedasticity Type II

The second type of heteroskedasticity that we consider is heteroskedasticity that arises at the individual level. For this we consider the following data-generating mechanism.

- $n = 1000$, $\epsilon_1 = N(0, 5)$, $\epsilon_2 = N(0, 0.5)$
- $x \sim U(-10.5, 10.5)$, $z \sim binom(1, 0.5)$
- Homoskedastic Mechanism: $y = 100(-2x + xz + \epsilon_1)$
- Heteroskedastic Mechanism: $y = 100(-2x + xz + z\epsilon_1 + (1 - z_1)\epsilon_2)$.

Note that there is again no confounding in this mechanism. This ensures that the number of units adherent to each regime is approximately equal across ψ. For the IPW estimator, conditional on known treatment propensities, variance is related to the number of units adherent to a treatment; we refer to this as the effective sample size. As we will see in Case III, confounding plays a role in the effective sample size and consequently on the variability structure.

From Figure B.2, we observe that the variability of the IPW-surface is heteroskedastic. From panel (a) we see that when there is homoskedastic noise in the person-level mechanism, this transfers to homoskedastic noise in the IPW-surface. The noise only plot in panel (c) is created by taking the weighted mean of the noise terms only. It is quite straightforward to see why heteroskedasticity comes about for regimes of type treat if "$x > \psi$". We have already established that, as we move from left to right in the plot, we lose treated patients and gain untreated patients. This means that we are losing observations with high variability (ϵ_1) and gaining observations with low variability (ϵ_2). This has the consequence of affecting the variability of the resulting estimator.



(a) (b) (c)

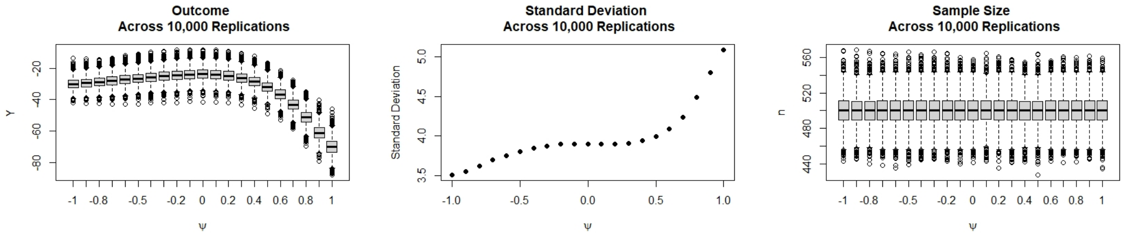
Figure B.2: Heteroskedasticity Type II: n=1000 (a) IPW estimator; homoskedastic case (b) IPW estimator; heteroskedastic case; (c) Noise Component of heteroskedastic case

B.1.3 Heteroskedasticity Type III

We now illustrate how variability is a function of sample size, in particular, we note that a given set of data may be more informative about one regime versus another, leading to the notion of effective sample size. One way to influence the effective sample size is by changing the confounding structure in the problem. Consider the following data generating mechanism:

- $z \sim binom(1, p),\ x \sim (\Gamma(\alpha = 1/3, \beta = 2) - 1)$.
- Confounding case: $p = expit(2.5x)$, Non-confounding case: $p = 0.5$.
- $y = 100(-2x + xz)$.

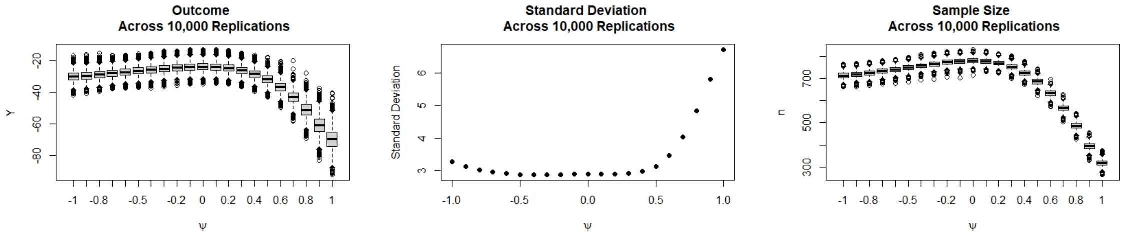
We contrast confounded and unconfounded data because even unconfounded data, which result in equal effective sample sizes across regimes, can lead to heteroskedasticity. This heteroskedasticity is likely arising from Case I. To examine whether effective sample size induces heteroskedasticity, we must compare the variability in the unconfounded case with the variability in the confounded case; if there is a non-constant change in variability of the estimator across regimes, then we can determine that further heteroskedasticity has been induced. Panels (b) and (c) in Figure B.3 show that the effective sample size is the same across all regimes and that there is heteroskedasticity.



(a) (b) (c)

Figure B.3: IPW estimator under no confounding setting (a) Expected outcome for varying thresholds (b) Standard deviation of estimator for varying thresholds (c) Sample size for varying thresholds.

Now if we induce confounding, as a way to manage the effective sample size, we obtain the results in Figure B.4. We see that for higher values of ψ, the effective sample size falls. For these same values the standard deviation increases. Comparing the standard deviation in Figure B.3 and B.4, we see that this change in variability is not uniform, thereby informing us that effective sample size leads to further heteroskedasticity. To understand how this drop in effective sample size comes about, note that the x distribution is concentrated on the right side of the plot. Furthermore, note that as we move from left to right, we encounter more data and a larger and larger portion of these patients are receiving treatment due to the confounding structure. Treated patients are exactly the patient we lose as we move right, and we are only gaining a small number of untreated patients. This results in the observed drop in effective sample size.



(a) (b) (c)

Figure B.4: IPW estimator under confounding setting (a) Expected outcome for varying thresholds (b) Standard deviation of estimator for varying thresholds (c) Sample size for varying thresholds.

B.2 Re-Interpolation

Here we show that the posterior mean in the re-interpolating process is the same as the posterior mean in the original process, be it in a homoskedastic or heteroskedastic setting. That is, we show that $\mu_{\hat{\upsilon}_{m+1}} = \mu_{\upsilon_{m+1}}$. For this, we first show that the empirical Bayes estimates for the prior means are the same, that is $\mu_{0\hat{\upsilon}} = \mu_{0\upsilon}$. The empirical Bayes expression for $\mu_{0\upsilon}$ is given by:

$$\mu_{0\upsilon} = \mu_{0f} = \frac{\mathbf{1}^T(K+S)^{-1}\upsilon}{\mathbf{1}^T(K+S)\mathbf{1}}. \tag{B.1}$$

This equation tells us that $\mathbf{1}(K+S)^{-1}\mathbf{1}\mu_{0\upsilon} = \mathbf{1}(K+S)^{-1}\upsilon$. Now based on the mean given in equation B.1, we are able to write the vector of predicted values as:

$$\hat{\upsilon} = \mathbf{1}\mu_{0\upsilon} + K(K+S)^{-1}(\upsilon - \mathbf{1}\mu_{0\upsilon}). \tag{B.2}$$

Then substituting this into the expression for $\mu_{0\hat{\upsilon}}$, analogous to expression B.1, we obtain:

$$\begin{aligned}
\mu_{0\hat{\upsilon}} =& \frac{\mathbf{1}^T K^{-1}\hat{\upsilon}}{\mathbf{1}^T K \mathbf{1}} \\
=& \frac{\mathbf{1}^T K^{-1}\mathbf{1}\mu_{0\upsilon} + \mathbf{1}^T(K+S)^{-1}(\upsilon - \mathbf{1}\mu_{0f})}{\mathbf{1}^T K^{-1}\mathbf{1}} \\
=& \mu_{0\upsilon} + \frac{\mathbf{1}(K+S)^{-1}\upsilon - \mathbf{1}(K+S)^{-1}\mathbf{1}\mu_{0\upsilon}}{\mathbf{1}^T K^{-1}\mathbf{1}} \\
=& \mu_{0\upsilon} + 0.
\end{aligned} \tag{B.3}$$

Plugging in the expression for $\hat{\upsilon}$ into the expression for the posterior mean in the re-interpolating model we get

$$\begin{aligned}
\mu_{\hat{\upsilon}_{m+1}} &= \mu_{0\hat{\upsilon}} + k^T K(\hat{\upsilon} - \mathbf{1}\hat{\mu}_{0\upsilon}) \\
&= \mu_{0\hat{\upsilon}} + k^T K^{-1}(\mathbf{1}\mu_{0\upsilon} + K(K+S)^{-1}(y - \mathbf{1}\mu_{0f}) - \mathbf{1}\mu_{0\hat{\upsilon}}) \\
&= \mu_{0\hat{\upsilon}} + k^T K^{-1}\mathbf{1}\mu_{0\upsilon} + k^T(K+S)^{-1}(y - \mathbf{1}\mu_{0f}) - k^T K^{-1}\mathbf{1}\mu_{0\upsilon} \qquad \text{(B.4)} \\
&= \mu_{0\upsilon} + k^T(K+S)^{-1}(\upsilon - \mathbf{1}\mu_{0\upsilon}) \\
&= \mu_{\upsilon_{m+1}}.
\end{aligned}$$

This was what we were required to show. Additionally, we recall that the Kriging variance at a new sample point is $\sigma^2_{\hat{\upsilon}_{m+1}} = k(\psi_{m+1}, \psi_{m+1}) - k^T K^{-1} k$ and it has the desired property of having zero error at already sampled points. All details of this approach can be found in Forrester et al. [2006].

B.3 Simulation I

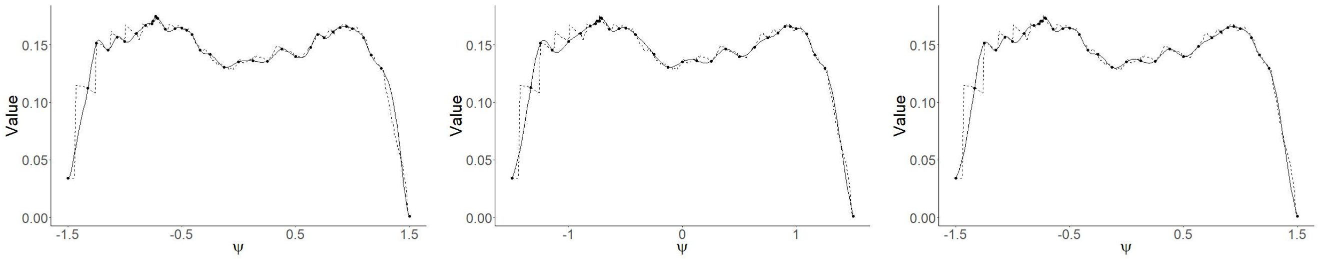
In this section, we explore other simulation settings to determine whether the obtained results are sensitive to modeling assumptions. This includes examining the performance of a $Matern_{3/2}$ covariance. Generally, the Matérn covariance family is identified by parameters ν and κ. This family of covariances has a particularly nice form when $\nu = \kappa + 1/2$. Generally κ is known to control how fast the correlation decays with distance, which determines the low frequency (Coarse-scale) behaviour. ν is known to control the high frequency (fine-scale) smoothness of the sample paths. Before examining the results for other simulation settings, we observe from Table B.1 that for a sample size of $n = 1000$, the grid search improves substantially as compared to the $n = 500$ setting.

Table B.1: Simulation I: Results for grid-search with increments of 0.01 and $n = 1000$. True $\psi_{opt} = 0.9$; true value at optimum 0.165.

Statistic	$\hat{\psi}_{opt}$	Value at Optimum
Mean (SD)	0.559 (0.691)	0.169 (0.025)
Median (IQR)	0.880 (0.133)	0.169 (0.020)

B.3.1 Simulation I: $Matern_{5/2}$ Covariance; Sample Size $n = 500$

From Figure B.5, we see the emulated curves for a single replicates when $n = 500$. These are the curves shown in Figure 2 of the main text, but now the domain has not been restricted.



(a) (b) (c)

Figure B.5: Simulation I: Emulation surfaces at +25 points overlaid over the IPW-surface (a) Int$\mathcal{GP}$ (b) HM$\mathcal{GP}$ (c) HE$\mathcal{GP}$.

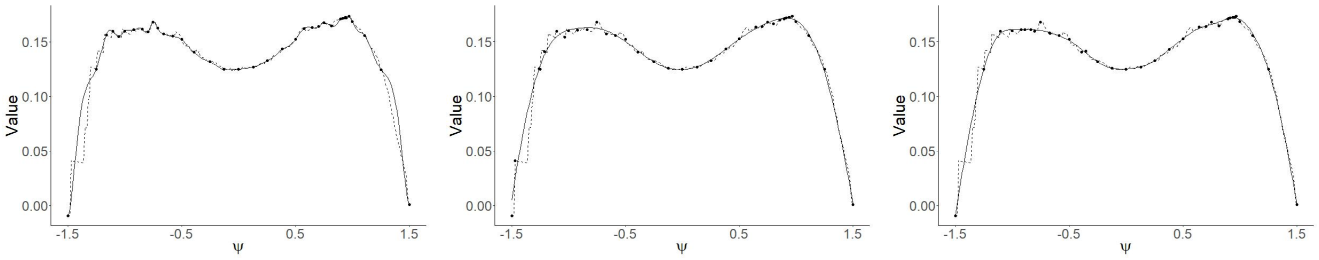
The results in Table B.2 are in line with the results (median IQR) presented in the main paper, and we see that all modeling approaches estimate well the value at the optimum.

Table B.2: Simulation I: Estimated value at $\hat{\psi}_opt$ optimum after $+m$ points, mean (SD); $n = 1000$ with 13 design points over 500 replicates. True value at optimum: 0.165.

$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Int$\mathcal{GP}$	0.169 (0.021)	0.171 (0.021)	0.171 (0.021)	0.171 (0.021)	0.171 (0.022)	0.172 (0.021)
HM$\mathcal{GP}$	0.169 (0.021)	0.170 (0.021)	0.170 (0.021)	0.170 (0.021)	0.170 (0.021)	0.170 (0.021)
HE$\mathcal{GP}$	0.169 (0.021)	0.170 (0.021)	0.170 (0.021)	0.170 (0.021)	0.170 (0.021)	0.171 (0.021)

B.3.2 Simulation I: $Matern_{5/2}$ Covariance; Sample Size $n = 1000$

From Figure B.6, we see that the IPW-surface exhibits more smoothness than the $n = 500$ scenario. Consequently, the emulation surfaces are very smooth. Frame (a) makes clear why the Int$\mathcal{GP}$ may identify the incorrect optimum more often than the other methods; there are spikes in the fit around the local maximum at -0.8.



(a) (b) (c)

Figure B.6: Simulation I: Contour plot at +25 points (a) Int$\mathcal{GP}$ (b) HM$\mathcal{GP}$ (c) HE$\mathcal{GP}$.

From Table B.3, we see that the mean is still biased for the optimal threshold. If we focus on the medians, we see that all three $\mathcal{GP}$ modeling approaches work well, with the HM$\mathcal{GP}$ having slightly more precision than the other methods. Inference is improved slightly by sampling additional points. From Table B.4, we see that all modeling approaches perform well in computing the value at the optimum.

Table B.3: Simulation I: Estimated optimal ψ after $+m$ points; $n = 1000$ with 13 design points over 500 replicates and $Matern_{5/2}$ covariance. True $\psi_{opt} = 0.9$.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	0.875 (0.169)	0.880 (0.126)	0.881 (0.124)	0.881 (0.122)	0.882 (0.124)	0.882 (0.126)
Med. (IQR)	HM$\mathcal{GP}$	0.880 (0.167)	0.883 (0.116)	0.884 (0.118)	0.882 (0.121)	0.884 (0.118)	0.883 (0.124)
Med. (IQR)	HE$\mathcal{GP}$	0.877 (0.167)	0.884 (0.124)	0.884 (0.118)	0.885 (0.117)	0.885 (0.121)	0.884 (0.120)
Mean (SD)	Int$\mathcal{GP}$	0.548 (0.693)	0.582 (0.668)	0.586 (0.668)	0.577 (0.676)	0.578 (0.677)	0.572 (0.681)
Mean (SD)	HM$\mathcal{GP}$	0.555 (0.692)	0.596 (0.659)	0.577 (0.674)	0.580 (0.674)	0.581 (0.673)	0.576 (0.676)
Mean (SD)	HE$\mathcal{GP}$	0.547 (0.696)	0.586 (0.667)	0.582 (0.672)	0.589 (0.668)	0.581 (0.675)	0.577 (0.677)

Table B.4: Simulation I: Estimated value at $\hat{\psi}_{opt}$ after $+m$ points; $n = 1000$ with 13 design points over 500 replicates and $Matern_{5/2}$ covariance. True value at ψ_{opt}: 0.165.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	0.167 (0.020)	0.168 (0.020)	0.168 (0.020)	0.169 (0.020)	0.169 (0.020)	0.169 (0.020)
Med. (IQR)	HM$\mathcal{GP}$	0.167 (0.019)	0.168 (0.020)	0.168 (0.020)	0.168 (0.020)	0.168 (0.020)	0.168 (0.020)
Med. (IQR)	HE$\mathcal{GP}$	0.167 (0.019)	0.168 (0.020)	0.168 (0.020)	0.168 (0.020)	0.168 (0.020)	0.168 (0.020)
Mean (SD)	Int$\mathcal{GP}$	0.167 (0.015)	0.168 (0.015)	0.168 (0.015)	0.168 (0.015)	0.168 (0.015)	0.168 (0.015)
Mean (SD)	HM$\mathcal{GP}$	0.167 (0.015)	0.167 (0.015)	0.167 (0.015)	0.167 (0.015)	0.167 (0.015)	0.168 (0.015)
Mean (SD)	HE$\mathcal{GP}$	0.167 (0.015)	0.167 (0.015)	0.167 (0.015)	0.167 (0.015)	0.168 (0.015)	0.168 (0.015)

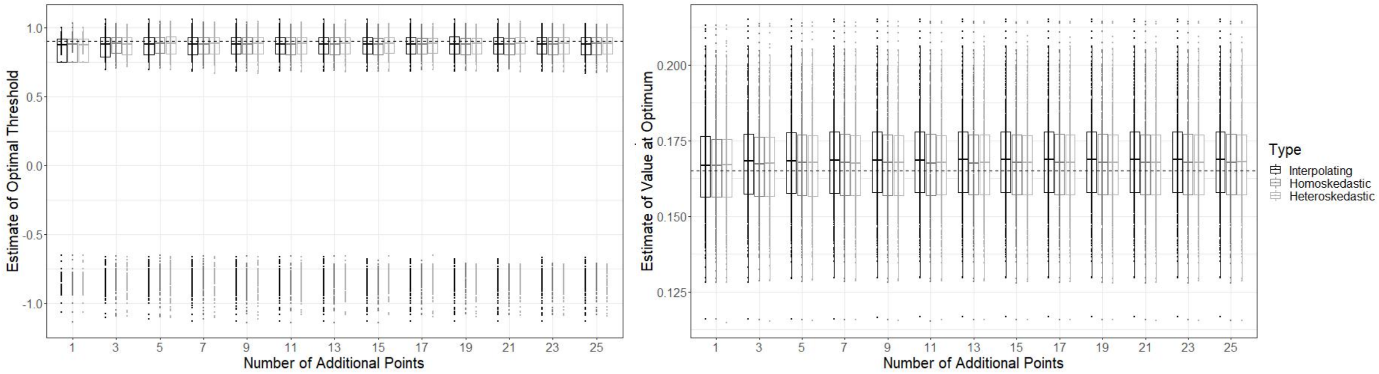


Figure B.7: Simulation I: Boxplot at $+m$ points; $n = 1000$ with 13 design points and $Matern_{5/2}$ covariance (a) Optimal ψ (b) Value at optimum.

B.3.3 Simulation I: $Matern_{3/2}$ Covariance; Sample Size $n = 1000$

We see from Figure B.8 that the posterior means of the $\mathcal{GP}$s very much resemble those in the $Matern_{5/2}$ scenario, and that they capture well the value surface. From Table B.5 and B.6, we see that the results for this setting are almost unchanged.

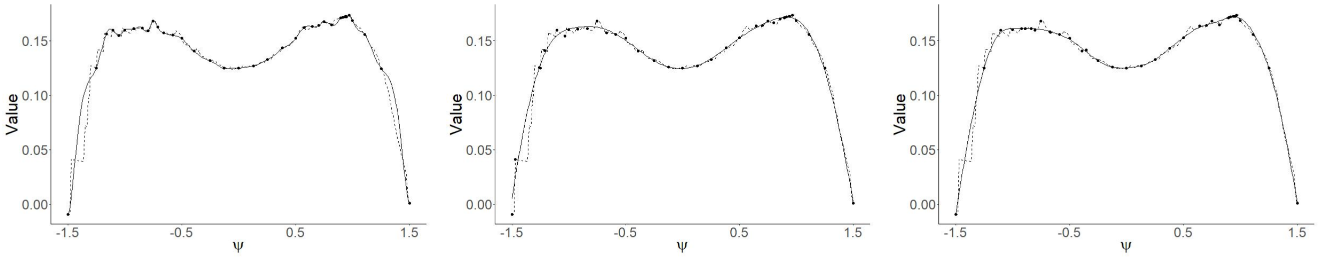


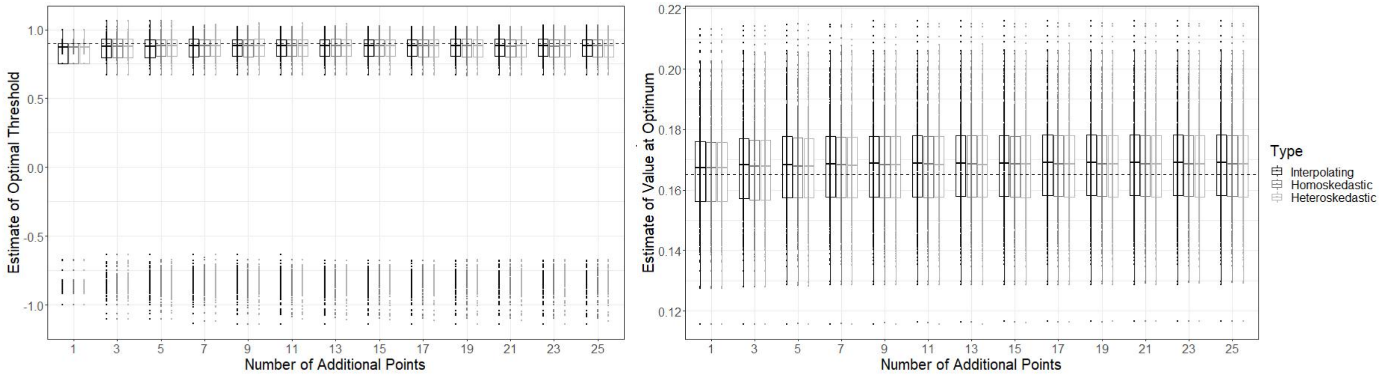
Figure B.8: Simulation I: Contour plot at +25 points (a) Int$\mathcal{GP}$ (b) HM$\mathcal{GP}$ (c) HE$\mathcal{GP}$.

Table B.5: Simulation I: Estimated optimal ψ after $+m$ points; $n = 1000$ with 13 design points over 500 replicates and $Matern_{3/2}$ covariance. True ψ_{opt} =0.9.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	0.875 (0.149)	0.880 (0.131)	0.882 (0.122)	0.881 (0.124)	0.882 (0.125)	0.881 (0.124)
Med. (IQR)	HM$\mathcal{GP}$	0.875 (0.149)	0.882 (0.122)	0.882 (0.122)	0.884 (0.123)	0.882 (0.126)	0.881 (0.124)
Med. (IQR)	HE$\mathcal{GP}$	0.875 (0.149)	0.881 (0.120)	0.884 (0.120)	0.885 (0.123)	0.882 (0.124)	0.882 (0.124)
Mean (SD)	Int$\mathcal{GP}$	0.549 (0.690)	0.570 (0.678)	0.579 (0.673)	0.573 (0.681)	0.572 (0.681)	0.575 (0.680)
Mean (SD)	HM$\mathcal{GP}$	0.552 (0.687)	0.579 (0.671)	0.578 (0.675)	0.576 (0.678)	0.57 (0.682)	0.570 (0.682)
Mean (SD)	HE$\mathcal{GP}$	0.549 (0.689)	0.583 (0.668)	0.582 (0.673)	0.581 (0.673)	0.577 (0.677)	0.572 (0.680)

Table B.6: Simulation I: Estimated value at $\hat{\psi}_{opt}$ after $+m$ points; $n = 1000$ with 13 design points over 500 replicates and $Matern_{3/2}$ covariance. True value at ψ_{opt}: 0.165.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	0.167 (0.020)	0.168 (0.020)	0.169 (0.020)	0.169 (0.020)	0.169 (0.020)	0.169 (0.020)
Med. (IQR)	HM$\mathcal{GP}$	0.167 (0.019)	0.168 (0.020)	0.168 (0.020)	0.169 (0.020)	0.169 (0.020)	0.169 (0.020)
Med. (IQR)	HE$\mathcal{GP}$	0.167 (0.019)	0.168 (0.020)	0.168 (0.020)	0.169 (0.020)	0.169 (0.020)	0.169 (0.020)
Mean (SD)	Int$\mathcal{GP}$	0.167 (0.015)	0.168 (0.015)	0.168 (0.015)	0.169 (0.015)	0.169 (0.015)	0.169 (0.015)
Mean (SD)	HM$\mathcal{GP}$	0.167 (0.015)	0.168 (0.015)	0.168 (0.015)	0.168 (0.015)	0.168 (0.015)	0.168 (0.015)
Mean (SD)	HE$\mathcal{GP}$	0.167 (0.015)	0.168 (0.015)	0.168 (0.015)	0.168 (0.015)	0.168 (0.015)	0.168 (0.015)

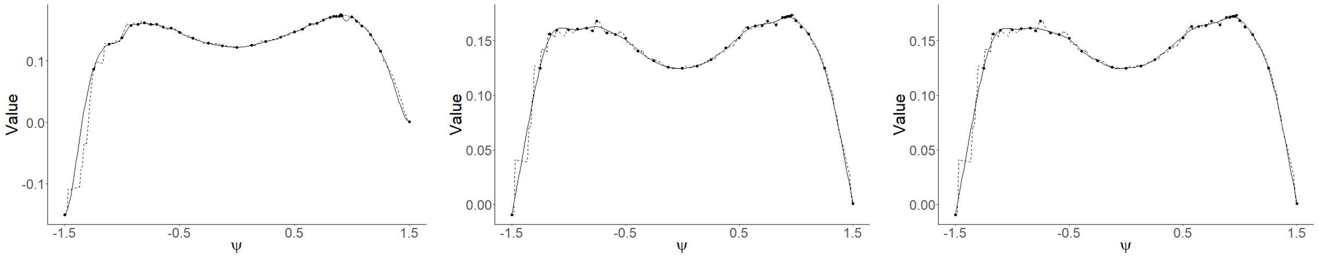


(a) (b)

Figure B.9: Simulation I: Boxplot at $+m$ points; $n = 1000$ with 13 design points and $Matern_{3/2}$ covariance (a) Optimal ψ (b) Value at optimum.

B.3.4 Simulation I: $Matern_{5/2}$ Covariance; Sample Size $n = 1000$; Log-Normal Prior

We see from the results below that the Log-Normal prior on the covariance parameter θ_f with a $Matern_{5/2}$ covariance does not change the results substantially. The hyperparameters in the Log-Normal prior were selected so that the $5th$ and $95th$ percentiles of the distribution have a correlation of 0.05 and 0.95, respectively, for a 10% increase in the range of ψ. This makes this prior rather non-informative.



(a) (b) (c)

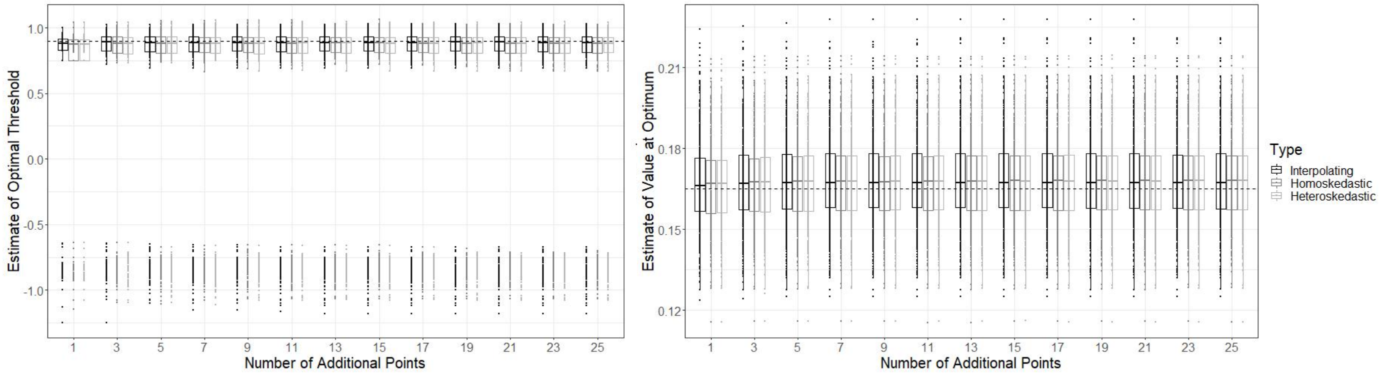
Figure B.10: Simulation I: Contour plot at +25 points: (a) Int$\mathcal{GP}$ (b) HM$\mathcal{GP}$ (c) HE$\mathcal{GP}$.

Table B.7: Simulation I: Estimated optimal ψ_1 after $+m$ points; $n = 1000$ with 13 design points over 500 replicates, $Matern_{5/2}$ covariance and Log-Normal prior. True $\psi_{op}t = 0.9$.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	0.881 (0.085)	0.888 (0.112)	0.888 (0.112)	0.888 (0.107)	0.890 (0.108)	0.886 (0.121)
Med. (IQR)	HM$\mathcal{GP}$	0.873 (0.157)	0.880 (0.123)	0.881 (0.113)	0.884 (0.115)	0.882 (0.117)	0.882 (0.119)
Med. (IQR)	HE$\mathcal{GP}$	0.873 (0.157)	0.880 (0.124)	0.882 (0.123)	0.880 (0.121)	0.883 (0.118)	0.882 (0.116)
Mean (SD)	Int$\mathcal{GP}$	0.609 (0.651)	0.604 (0.659)	0.606 (0.659)	0.601 (0.664)	0.606 (0.660)	0.578 (0.683)
Mean (SD)	HM$\mathcal{GP}$	0.551 (0.691)	0.589 (0.664)	0.586 (0.668)	0.580 (0.675)	0.574 (0.678)	0.574 (0.678)
Mean (SD)	HE$\mathcal{GP}$	0.552 (0.688)	0.585 (0.667)	0.578 (0.674)	0.577 (0.676)	0.577 (0.675)	0.577 (0.676)

Table B.8: Simulation I: Estimated value at $\hat{\psi}_{opt}$ after $+m$ points; $n = 1000$ with 13 design points over 500, $Matern_{5/2}$ covariance and Log-Normal prior. True value at ψ_{opt}: 0.165.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	0.166 (0.020)	0.167 (0.020)	0.167 (0.020)	0.167 (0.020)	0.167 (0.020)	0.167 (0.021)
Med. (IQR)	HM$\mathcal{GP}$	0.167 (0.020)	0.168 (0.020)	0.168 (0.020)	0.168 (0.020)	0.168 (0.020)	0.168 (0.020)
Med. (IQR)	HE$\mathcal{GP}$	0.167 (0.019)	0.168 (0.020)	0.168 (0.020)	0.168 (0.020)	0.168 (0.020)	0.168 (0.020)
Mean (SD)	Int$\mathcal{GP}$	0.167 (0.016)	0.168 (0.016)	0.169 (0.016)	0.169 (0.016)	0.169 (0.016)	0.168 (0.016)
Mean (SD)	HM$\mathcal{GP}$	0.167 (0.015)	0.167 (0.015)	0.167 (0.015)	0.168 (0.015)	0.168 (0.015)	0.168 (0.015)
Mean (SD)	HE$\mathcal{GP}$	0.167 (0.015)	0.167 (0.015)	0.168 (0.015)	0.168 (0.015)	0.168 (0.015)	0.168 (0.015)



(a) (b)

Figure B.11: Simulation I: Boxplot at $+m$ points; $n = 1000$ with 13 design points, $Matern_{5/2}$ covariance and Log-Normal prior (a) Optimal ψ (b) Value at optimum.

B.4 Simulation II

From Table B.9, we see that the grid search results for $n = 1000$. These reveal much more precise results than the grid search for $n = 500$. In the following sections we examine the method performance under a variety of modeling assumptions.

Table B.9: Simulation II: Grid search results with increments of 0.05 and $n = 1000$. True $(\psi_{1opt}, \psi_{2opt}) = (1.8, -0.3)$; true value at optimum 0.241.

Statistic	Simulation	$\hat{\psi}_{1opt}$	$\hat{\psi}_{2opt}$	Value at Optimum
Median (IQR)	Homoskedastic	1.800 (0.150)	-0.300 (0.225)	0.260 (0.081)
Median (IQR)	Heteroskedastic	1.800 (0.150)	-0.300 (0.225)	0.267 (0.094)
Mean (SD)	Homoskedastic	1.488 (0.730)	-0.360 (0.233)	0.260 (0.065)
Mean (SD)	Heteroskedastic	1.401 (0.888)	-0.366 (0.290)	0.267 (0.073)

B.4.1 Simulation II: $Matern_{5/2}$ Covariance; Sample Size $n = 500$

The following tables present the results for main simulation as measured by the mean and standard deviation, having presented the results for the median and interquartile range in the main paper. As expected, the results reveal the multi-modality of the problem.

Table B.10: Simulation II: Estimated optimal ψ_1 after $+m$ points; n = 500 with 16 design points over 500 replicates, and $Matern_{5/2}$ covariance. True ψ_{1opt} =1.8.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Mean (SD)	Int$\mathcal{GP}$	-0.111 (1.209)	0.429 (1.378)	0.745 (1.309)	0.854 (1.266)	0.904 (1.247)	0.912 (1.261)
Mean (SD)	HM$\mathcal{GP}$	0.302 (1.589)	0.599 (1.560)	0.862 (1.390)	0.999 (1.313)	1.085 (1.249)	1.139 (1.194)
Mean (SD)	HE$\mathcal{GP}$	0.324 (1.548)	0.577 (1.499)	0.886 (1.356)	0.996 (1.306)	1.105 (1.227)	1.126 (1.204)

Table B.11: Simulation II: Estimated optimal ψ_2 after $+m$ points; n = 500 with 16 design points over 500 replicates, and $Matern_{5/2}$ covariance. True ψ_{2opt} = -0.3.

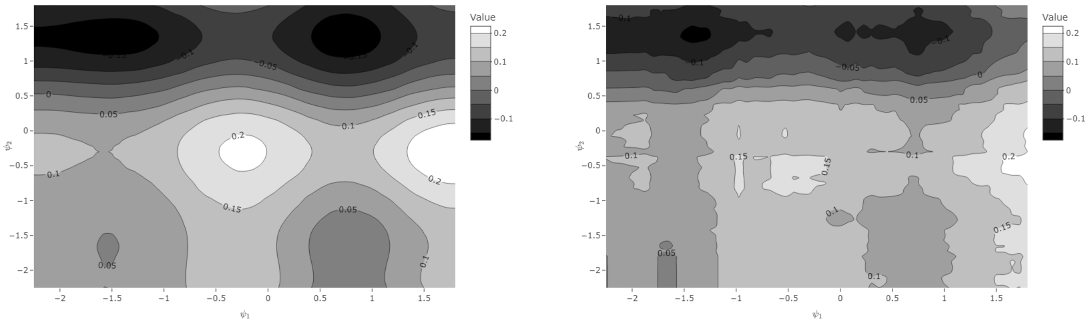
Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Mean (SD)	Int$\mathcal{GP}$	-0.389 (0.571)	-0.413 (0.483)	-0.422 (0.439)	-0.429 (0.422)	-0.426 (0.423)	-0.429 (0.437)
Mean (SD)	HM$\mathcal{GP}$	-0.391 (0.569)	-0.420 (0.489)	-0.429 (0.466)	-0.416 (0.422)	-0.410 (0.421)	-0.411 (0.429)
Mean (SD)	HE$\mathcal{GP}$	-0.378 (0.564)	-0.411 (0.471)	-0.435 (0.442)	-0.428 (0.429)	-0.430 (0.455)	-0.426 (0.450)

Table B.12: Simulation II: Estimated value at $\hat{\psi}_{1opt}, \hat{\psi}_{2opt}$ after $+m$ points; n = 500 with 16 design points over 500 replicates, and $Matern_{5/2}$ covariance. True value at ψ_{1opt}, ψ_{2opt}: 0.241.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Median (SD)	Int$\mathcal{GP}$	0.199 (0.095)	0.242 (0.101)	0.259 (0.100)	0.264 (0.099)	0.266 (0.097)	0.269 (0.096)
Median (SD)	HM$\mathcal{GP}$	0.200 (0.102)	0.233 (0.103)	0.250 (0.099)	0.259 (0.097)	0.263 (0.096)	0.265 (0.095)
Median (SD)	HE$\mathcal{GP}$	0.203 (0.101)	0.237 (0.103)	0.253 (0.098)	0.260 (0.097)	0.264 (0.095)	0.266 (0.095)

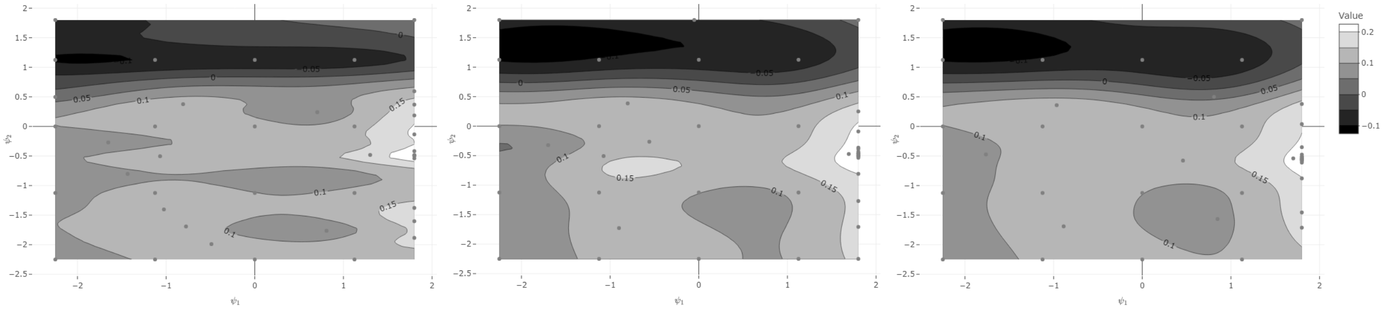
B.4.2 Simulation II: $Matern_{5/2}$ Covariance; Sample size $n = 1000$

Figure B.12 shows that, visually, the IPW-surface generally captures the main characteristics of the value function. This can also be seen in the Interactive Supplement. Additionally, Figure B.13 shows the estimated value function resulting from each of the three $\mathcal{GP}$ modeling approaches, after 25 additional sampled points. All three approaches explore the region around the global optimum; only the HM$\mathcal{GP}$ gives some possibility that there is another local maxima around the center of the domain.



(a) (b)

Figure B.12: Simulation II: (a) Value function (b) Estimated value function using normalized IPW.



(a) (b) (c)

Figure B.13: Simulation II: Contour plot at +25 points: (a) Int$\mathcal{GP}$ (b) HM$\mathcal{GP}$ (c) HE$\mathcal{GP}$.

If we examine what results after 500 replications of this analysis, Table B.13 and B.14 show that relatively good performance can be attained using all three methods after sufficient exploration, though the HM$\mathcal{GP}$ and HE$\mathcal{GP}$ yield even better performance, specifically with regard to the ψ_1 parameter, as they have a lower IQR. When compared to the grid search, the HM$\mathcal{GP}$ and HE$\mathcal{GP}$ yield similar performance with regard to ψ_2, and better performance with regard to ψ_1.

Table B.13: Simulation II: Estimated optimal ψ_1 after $+m$ points; $n = 1000$ with 16 design points over 500 replicates, and $Matern_{5/2}$ covariance. True ψ_{1opt} =1.8.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	0.000 (0.536)	0.359 (2.003)	1.800 (1.933)	1.800 (0.219)	1.800 (0.124)	1.800 (0.104)
Med. (IQR)	HM$\mathcal{GP}$	1.800 (1.800)	1.800 (1.520)	1.800 (0.000)	1.800 (0.000)	1.800 (0.000)	1.800 (0.000)
Med. (IQR)	HE$\mathcal{GP}$	1.800 (1.80)	1.800 (1.741)	1.800 (0.000)	1.800 (0.000)	1.800 (0.000)	1.800 (0.000)
Mean (SD)	Int$\mathcal{GP}$	0.089 (1.015)	0.609 (1.246)	1.046 (1.158)	1.233 (1.039)	1.352 (0.930)	1.413 (0.866)
Mean (SD)	HM$\mathcal{GP}$	0.759 (1.510)	1.013 (1.396)	1.212 (1.181)	1.402 (0.962)	1.506 (0.818)	1.501 (0.813)
Mean (SD)	HE$\mathcal{GP}$	0.671 (1.598)	0.969 (1.408)	1.237 (1.191)	1.343 (1.062)	1.452 (0.907)	1.466 (0.899)

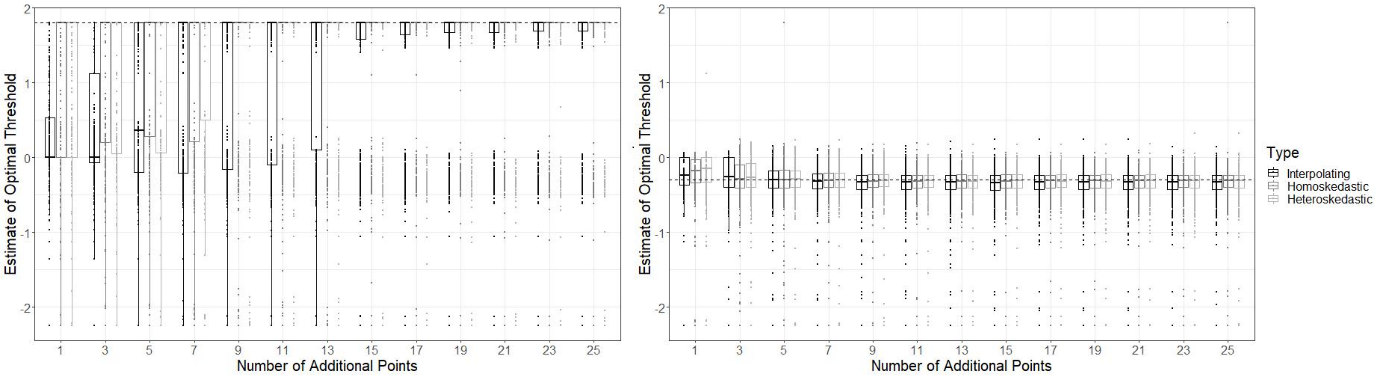
Table B.14: Simulation II: Estimated optimal ψ_2 after $+m$ points; $n = 1000$ with 16 design points over 500 replicates, and $Matern_{5/2}$ covariance. True $\psi_{2opt} = 0.3$.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	-0.237 (0.365)	-0.297 (0.231)	-0.331 (0.193)	-0.335 (0.198)	-0.333 (0.193)	-0.330 (0.194)
Med. (IQR)	HM$\mathcal{GP}$	-0.181 (0.313)	-0.303 (0.236)	-0.323 (0.176)	-0.315 (0.175)	-0.314 (0.174)	-0.310 (0.161)
Med. (IQR)	HE$\mathcal{GP}$	-0.150 (0.332)	-0.291 (0.224)	-0.309 (0.156)	-0.309 (0.171)	-0.315 (0.176)	-0.313 (0.177)
Mean (SD)	Int$\mathcal{GP}$	-0.274 (0.419)	-0.330 (0.309)	-0.370 (0.287)	-0.363 (0.254)	-0.360 (0.242)	-0.359 (0.232)
Mean (SD)	HM$\mathcal{GP}$	-0.249 (0.385)	-0.316 (0.313)	-0.345 (0.248)	-0.344 (0.232)	-0.349 (0.226)	-0.343 (0.248)
Mean (SD)	HE$\mathcal{GP}$	-0.232 (0.369)	-0.301 (0.292)	-0.338 (0.223)	-0.343 (0.227)	-0.348 (0.233)	-0.347 (0.231)

Table B.15: Simulation II: Estimated value at $\hat{\psi}_{1opt}, \hat{\psi}_{2opt}$ after $+m$ points; $n = 1000$ with 16 design points over 500 replicates, and $Matern_{5/2}$ covariance. True value at ψ_{1opt}, ψ_{2opt}: 0.241.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	0.176 (0.087)	0.225 (0.108)	0.249 (0.094)	0.254 (0.086)	0.256 (0.082)	0.257 (0.082)
Med. (IQR)	HM$\mathcal{GP}$	0.205 (0.103)	0.238 (0.095)	0.249 (0.089)	0.250 (0.082)	0.254 (0.084)	0.255 (0.080)
Med. (IQR)	HE$\mathcal{GP}$	0.205 (0.112)	0.235 (0.100)	0.247 (0.096)	0.252 (0.093)	0.255 (0.082)	0.254 (0.080)
Mean (SD)	Int$\mathcal{GP}$	0.180 (0.068)	0.224 (0.076)	0.246 (0.073)	0.252 (0.070)	0.254 (0.067)	0.256 (0.067)
Mean (SD)	HM$\mathcal{GP}$	0.202 (0.078)	0.229 (0.076)	0.243 (0.072)	0.248 (0.069)	0.252 (0.068)	0.253 (0.067)
Mean (SD)	HE$\mathcal{GP}$	0.201 (0.079)	0.229 (0.075)	0.243 (0.073)	0.249 (0.070)	0.253 (0.067)	0.253 (0.066)

We can also visualize these results via Figures B.14 and B.15. The superiority of the HM$\mathcal{GP}$ becomes apparent in estimating the ψ_1 parameter, as it identifies the global optimum much faster than the other two methods. Although all three approaches yield results that concentrate around the true value after a certain number of samples, the HM$\mathcal{GP}$ and HE$\mathcal{GP}$ approaches converge around the true value more quickly than the Int$\mathcal{GP}$.



(a)

(b)

Figure B.14: Simulation II: Boxplot at $+m$ points; $n = 1000$ with 16 design points, and $Matern_{5/2}$ covariance (a) Optimal ψ_1 (b) Optimal ψ_2.

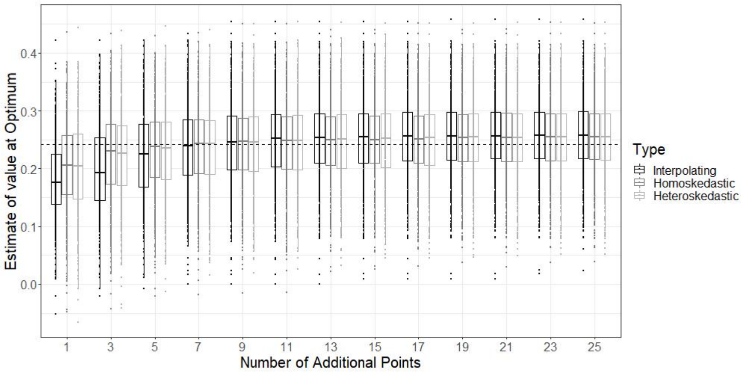
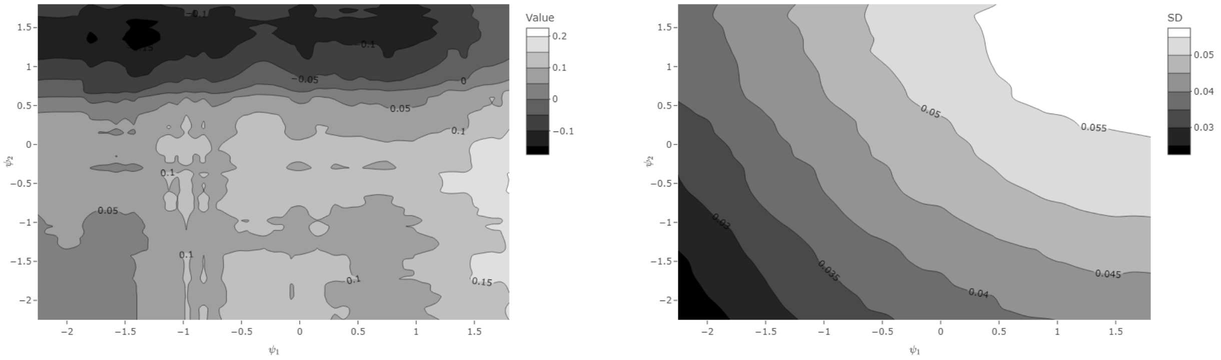


Figure B.15: Boxplot of value at optimum after $+m$ points; $n = 1000$ with 16 design points, and $Matern_{5/2}$ covariance.

B.4.3 Simulation II: $Matern_{5/2}$ Covariance; Sample Size $n = 1000$; Heteroskedastic Noise

In this section, we explore the consequences of having additive heteroskedastic noise as opposed to homoskedastic noise. We consider the following additive noise structure for the Simulation II data generating mechanism: $\epsilon = z_1\epsilon_{12} + (1 - z_1)\epsilon_{21} + z_2\epsilon_{12} + (1 - z_2)\epsilon_{22}$, where $\epsilon_{11}, \epsilon_{12} \sim N(0, 0.05^2)$ and $\epsilon_{21}, \epsilon_{22} \sim N(0, 0.45^2)$. We see from Figure B.16 (a) that the estimated value function is not as well captured as in the additive homoskedastic noise scenario. This can be further observed in the Interactive Supplement. From Figure B.16

(b), we see that the additive heteroskedastic noise introduced at the individual level results in further heteroskedasticity at the estimator level.

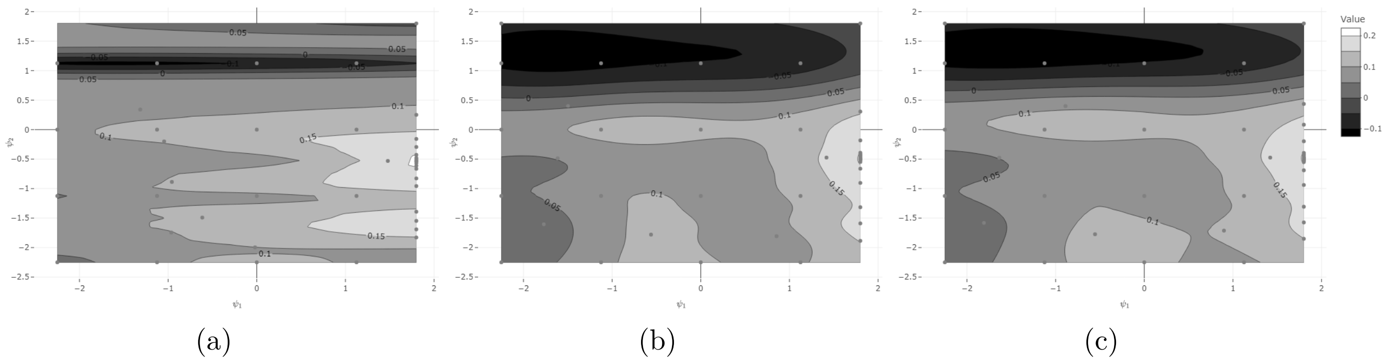


(a)

(b)

Figure B.16: Simulation II: (a) Estimated value function (b) Standard deviation of additive noise term, across replicates.

We see from Figure B.17 that the additive heteroskedastic noise has not changed the exploration substantially. From Table B.13, we see that the HM$\mathcal{GP}$ and HE$\mathcal{GP}$ produce better results than the interpolating and grid search approaches, specifically looking at the IQR. In this case, however, the grid search yields relatively good performance. It is surprising that the HE$\mathcal{GP}$ does not yield improved performance in this setting. In part, this is due to the fact that it is challenging to estimate the noise surface when we are exploring only very specific regions of the domain (where the maximum may be). Consequently, there is not enough data to obtain a precise estimate of the systematic structure in the heteroskedastic noise. When introducing this approach to HE$\mathcal{GP}$s, Zhang and Ni [2020] only examined it in one dimensional problems, and in a non-optimization context. This means that there was a considerable amount of data in all regions of the domain, thereby allowing for the precise estimation of a systematic component to the noise variance. A scatter plot of the residuals can be found in the Interactive Supplement. We see that in regions that are not well explored, the residuals are nearly zero. However, more exploration in those regions would certainly change the estimated residual structure.



(a) (b) (c)

Figure B.17: Simulation II: Contour plot at +25 points: (a) Int$\mathcal{GP}$ (b) HM$\mathcal{GP}$ (c) HE$\mathcal{GP}$.

Table B.16: Simulation II: Estimated optimal ψ_1 after $+m$ points; $n = 1000$ with 16 design points over 500 replicates, and $Matern_{5/2}$ covariance. True ψ_{1opt} = 1.8.

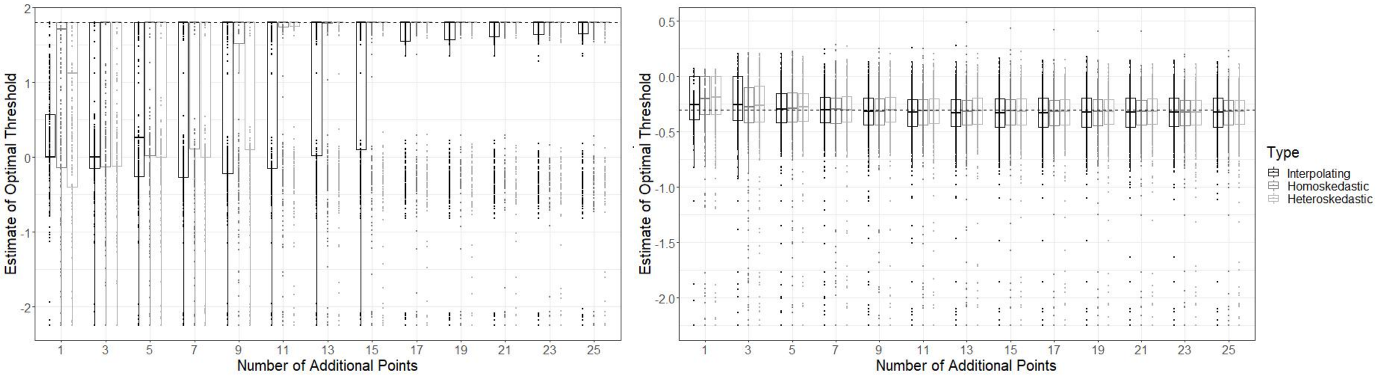
Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	0.00 (0.569)	0.262 (2.056)	1.800 (1.962)	1.800 (1.694)	1.800 (0.215)	1.800 (0.149)
Med. (IQR)	HM$\mathcal{GP}$	1.710 (1.935)	1.800 (1.776)	1.800 (0.165)	1.800 (0.000)	1.800 (0.000)	1.800 (0.000)
Med. (IQR)	HE$\mathcal{GP}$	1.125 (2.197)	1.800 (1.800)	1.800 (0.856)	1.800 (0.000)	1.800 (0.000)	1.800 (0.000)
Mean (SD)	Int$\mathcal{GP}$	0.017 (1.079)	0.560 (1.306)	0.954 (1.234)	1.156 (1.126)	1.247 (1.045)	1.326 (0.966)
Mean (SD)	HM$\mathcal{GP}$	0.580 (1.539)	0.906 (1.407)	1.134 (1.245)	1.300 (1.104)	1.387 (1.008)	1.417 (0.964)
Mean (SD)	HE$\mathcal{GP}$	0.481 (1.583)	0.817 (1.436)	1.139 (1.233)	1.315 (1.073)	1.416 (0.951)	1.452 (0.913)

Table B.17: Simulation II: Estimated optimal ψ_2 after $+m$ points; $n = 1000$ with 16 design points over 500 replicates, and $Matern_{5/2}$ covariance. True ψ_{2opt} = -0.3.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	-0.257 (0.392)	-0.294 (0.262)	-0.315 (0.248)	-0.329 (0.260)	-0.327 (0.265)	-0.324 (0.265)
Med. (IQR)	HM$\mathcal{GP}$	-0.201 (0.344)	-0.287 (0.267)	-0.312 (0.232)	-0.312 (0.230)	-0.320 (0.223)	-0.319 (0.228)
Med. (IQR)	HE$\mathcal{GP}$	-0.190 (0.343)	-0.275 (0.248)	-0.309 (0.218)	-0.313 (0.232)	-0.315 (0.227)	-0.315 (0.222)
Mean (SD)	Int$\mathcal{GP}$	-0.336 (0.530)	-0.349 (0.377)	-0.365 (0.336)	-0.369 (0.312)	-0.368 (0.307)	-0.367 (0.305)
Mean (SD)	HM$\mathcal{GP}$	-0.311 (0.505)	-0.326 (0.361)	-0.355 (0.316)	-0.367 (0.324)	-0.377 (0.336)	-0.370 (0.306)
Mean (SD)	HE$\mathcal{GP}$	-0.314 (0.521)	-0.344 (0.403)	-0.349 (0.319)	-0.366 (0.324)	-0.371 (0.320)	-0.377 (0.325)

Table B.18: Simulation II: Estimated value at $\hat{\psi}_{1opt}, \hat{\psi}_{2opt}$ after $+m$ points; $n = 1000$ with 16 design points over 500 replicates, and $Matern_{5/2}$ covariance. True value at ψ_{1opt}, ψ_{2opt}: 0.241.

Estimate	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	0.182 (0.098)	0.232 (0.113)	0.253 (0.107)	0.261 (0.096)	0.261 (0.096)	0.262 (0.093)
Med. (IQR)	HM$\mathcal{GP}$	0.202 (0.118)	0.229 (0.116)	0.249 (0.110)	0.256 (0.101)	0.259 (0.098)	0.261 (0.094)
Med. (IQR)	HE$\mathcal{GP}$	0.203 (0.118)	0.234 (0.116)	0.249 (0.109)	0.255 (0.098)	0.259 (0.097)	0.260 (0.093)
Mean (SD)	Int$\mathcal{GP}$	0.187 (0.073)	0.231 (0.082)	0.249 (0.080)	0.257 (0.077)	0.259 (0.076)	0.259 (0.074)
Mean (SD)	HM$\mathcal{GP}$	0.200 (0.085)	0.228 (0.085)	0.245 (0.082)	0.252 (0.078)	0.256 (0.076)	0.258 (0.074)
Mean (SD)	HE$\mathcal{GP}$	0.199 (0.084)	0.231 (0.085)	0.246 (0.080)	0.254 (0.077)	0.258 (0.074)	0.259 (0.074)



(a) (b)

Figure B.18: Simulation II: Boxplot after $+m$ points; $n = 1000$ with 16 design points, and $Matern_{5/2}$ covariance (a) Optimal ψ_1 (b) Optimal ψ_2.

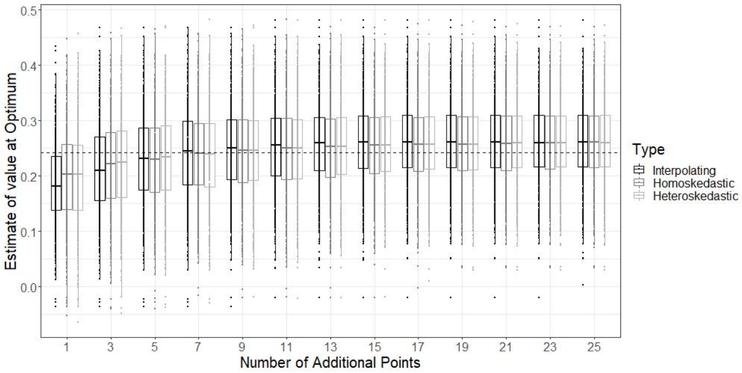


Figure B.19: Simulation II: Boxplot of value at optimum for $+m$ points; $n = 1000$ with 16 design points, and $Matern_{5/2}$ covariance.

B.4.4 Simulation II: $Matern_{3/2}$ Covariance; Sample Size $n = 1000$

We see from Figure B.20, that under the $Matern_{3/2}$ covariance, the local maximizer is still not captured and all methods focus exploration in the area near the global maximizer. From Tables B.19-B.21, we see that the $\mathcal{GP}s$ perform slightly worse than in the $Matern_{5/2}$ covariance setting. We see from Figure B.21 that with this covariance function it takes more additional samples for the HM$\mathcal{GP}$ to achieve increased precision.

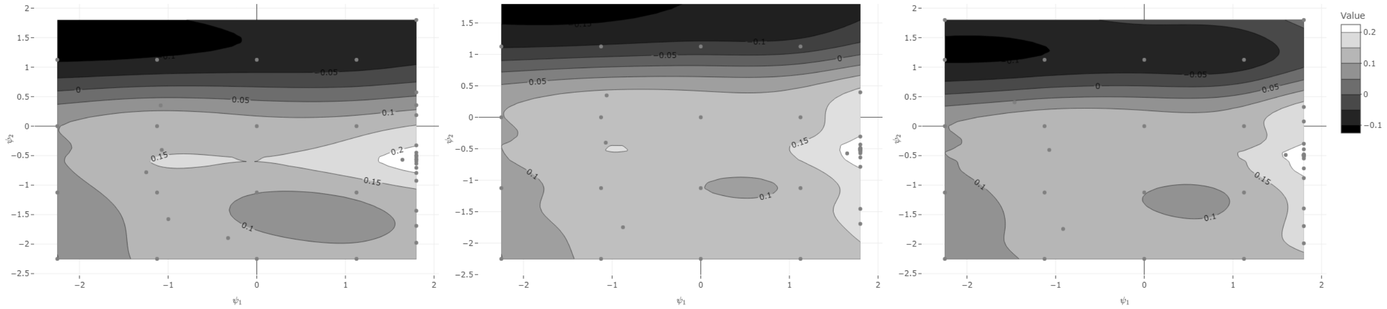


Figure B.20: Simulation II: Contour plot at +25 points: (a) Int$\mathcal{GP}$ (b) HM$\mathcal{GP}$ (c) HE$\mathcal{GP}$.

Table B.19: Simulation II: Estimated optimal ψ_1 after $+m$ points; $n = 1000$ with 16 design points over 500 replicates, and $Matern_{3/2}$ covariance. True $\psi_{1opt} = 1.8$.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	0.000 (0.506)	1.800 (1.958)	1.800 (1.761)	1.800 (0.055)	1.800 (0.054)	1.800 (0.058)
Med. (IQR)	HM$\mathcal{GP}$	0.247 (1.800)	1.800 (1.906)	1.800 (1.898)	1.800 (1.714)	1.800 (0.100)	1.800 (0.089)
Med. (IQR)	HE$\mathcal{GP}$	0.343 (1.800)	1.800 (1.852)	1.800 (1.859)	1.800 (0.101)	1.800 (0.076)	1.800 (0.071)
Mean (SD)	Int$\mathcal{GP}$	0.120 (1.057)	0.725 (1.273)	1.155 (1.142)	1.334 (1.008)	1.416 (0.912)	1.418 (0.906)
Mean (SD)	HM$\mathcal{GP}$	0.415 (1.260)	0.736 (1.334)	1.030 (1.181)	1.188 (1.078)	1.289 (0.998)	1.343 (0.947)
Mean (SD)	HE$\mathcal{GP}$	0.493 (1.365)	0.813 (1.346)	1.106 (1.172)	1.263 (1.033)	1.345 (0.953)	1.387 (0.920)

Table B.20: Simulation II: Estimated optimal ψ_2 after $+m$ points; $n = 1000$ with 16 design points over 500 replicates, and $Matern_{3/2}$ covariance. True ψ_{2opt} =-0.3.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	-0.279 (0.378)	-0.309 (0.238)	-0.327 (0.202)	-0.331 (0.195)	-0.330 (0.197)	-0.333 (0.201)
Med. (IQR)	HM$\mathcal{GP}$	-0.256 (0.362)	-0.298 (0.208)	-0.302 (0.195)	-0.312 (0.200)	-0.309 (0.204)	-0.317 (0.205)
Med. (IQR)	HE$\mathcal{GP}$	-0.217 (0.354)	-0.301 (0.223)	-0.304 (0.200)	-0.314 (0.190)	-0.315 (0.192)	-0.317 (0.191)
Mean (SD)	Int$\mathcal{GP}$	-0.302 (0.407)	-0.355 (0.297)	-0.364 (0.266)	-0.364 (0.260)	-0.363 (0.263)	-0.365 (0.256)
Mean (SD)	HM$\mathcal{GP}$	-0.285 (0.404)	-0.343 (0.291)	-0.349 (0.257)	-0.351 (0.256)	-0.353 (0.255)	-0.356 (0.248)
Mean (SD)	HE$\mathcal{GP}$	-0.262 (0.398)	-0.337 (0.283)	-0.350 (0.273)	-0.352 (0.264)	-0.354 (0.259)	-0.356 (0.255)

Table B.21: Simulation II: Estimated value at $\hat{\psi}_{1opt}, \hat{\psi}_{2opt}$ after $+m$ points; $n = 1000$ with 16 design points over 500 replicates, and $Matern_{3/2}$ covariance. True value at ψ_{1opt}, ψ_{2opt}: 1.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	0.181 (0.093)	0.234 (0.101)	0.253 (0.091)	0.259 (0.084)	0.259 (0.083)	0.261 (0.082)
Med. (IQR)	HM$\mathcal{GP}$	0.191 (0.100)	0.234 (0.102)	0.245 (0.092)	0.250 (0.092)	0.255 (0.090)	0.257 (0.089)
Med. (IQR)	HE$\mathcal{GP}$	0.195 (0.104)	0.235 (0.099)	0.248 (0.092)	0.253 (0.091)	0.256 (0.088)	0.258 (0.083)
Mean (SD)	Int$\mathcal{GP}$	0.186 (0.071)	0.233 (0.074)	0.250 (0.071)	0.256 (0.068)	0.258 (0.067)	0.259 (0.067)
Mean (SD)	HM$\mathcal{GP}$	0.193 (0.074)	0.228 (0.076)	0.242 (0.072)	0.247 (0.071)	0.252 (0.070)	0.254 (0.069)
Mean (SD)	HE$\mathcal{GP}$	0.198 (0.074)	0.231 (0.075)	0.245 (0.071)	0.251 (0.069)	0.254 (0.068)	0.256 (0.067)

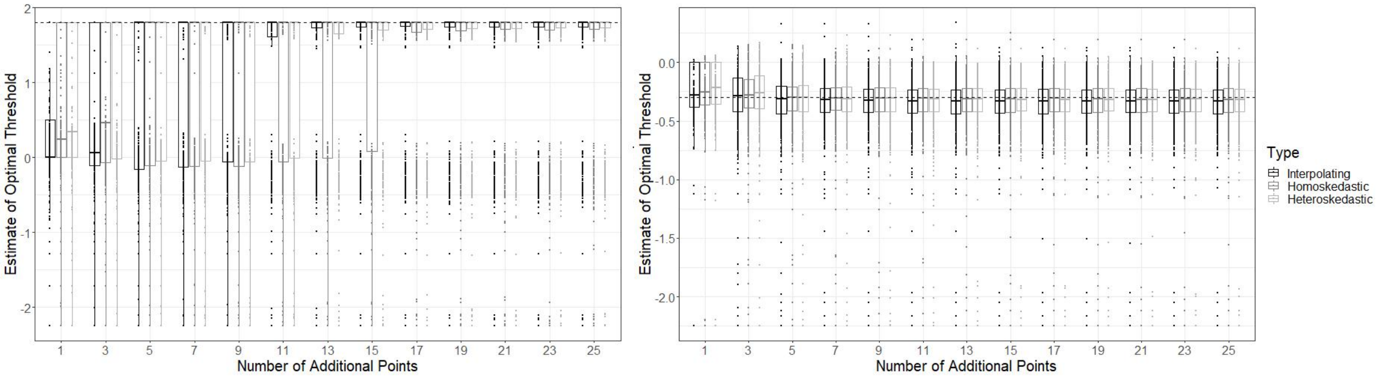


Figure B.21: Simulation II: Boxplot after $+m$ points; $n = 1000$ with 16 design points, and $Matern_{3/2}$ covariance (a) Optimal ψ_1 (b) Optimal ψ_2.

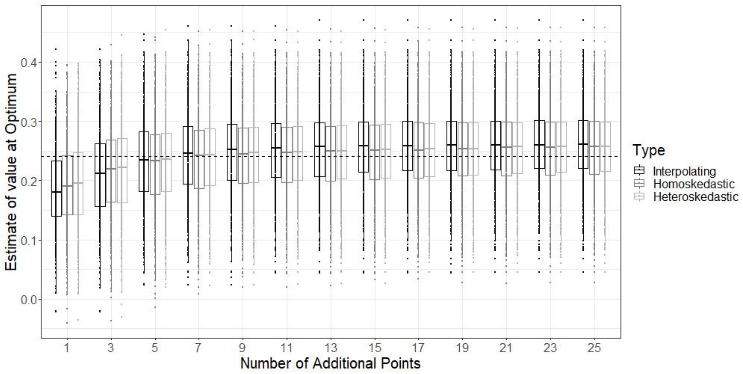
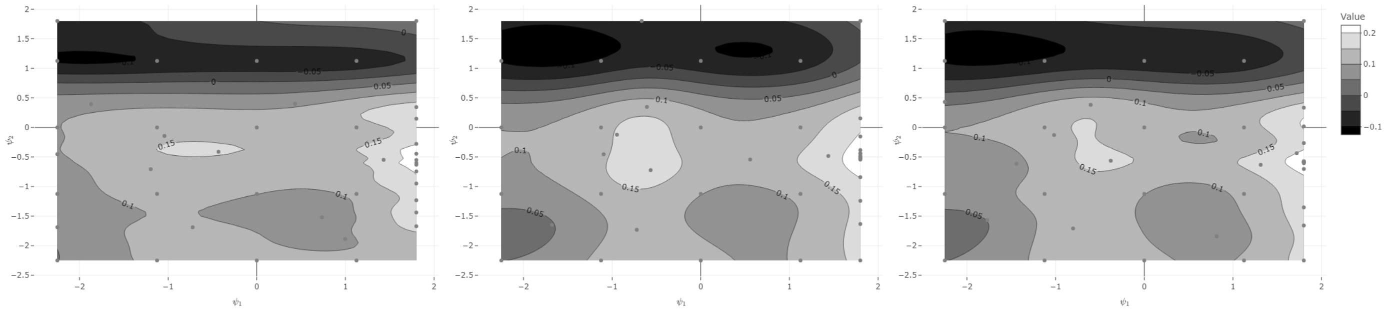


Figure B.22: Simulation II: Boxplot of value for $+m$ points; $n = 1000$ with 16 design points, and $Matern_{3/2}$ covariance.

B.4.5 Simulation II: $Matern_{5/2}$ Covariance; Sample Size $n = 1000$; Log-Normal Prior

We see from Figure B.23 that the HM$\mathcal{GP}$ and HE$\mathcal{GP}$, capture both the global maximum and local maximum. We see from Tables B.22-B.24 that at an additional 25 points, the results using the Log-Normal prior are very similar to those without a prior. From Figure B.24 and B.24, we see that the effect of additional points is very similar to that of the no prior setting. The prior hyperparameters were chosen using the same approach as in simulation I.



(a) (b) (c)

Figure B.23: Simulation II: Contour plot at +25 points (a) Int$\mathcal{GP}$ (b) HM$\mathcal{GP}$ (c) HE$\mathcal{GP}$.

Table B.22: Simulation II: Estimated optimal ψ_1 after +m points; $n = 1000$ with 16 design points over 500 replicates, $Matern_{5/2}$ covariance, and Log-Normal prior. True $\psi_{1opt} = 1.8$.

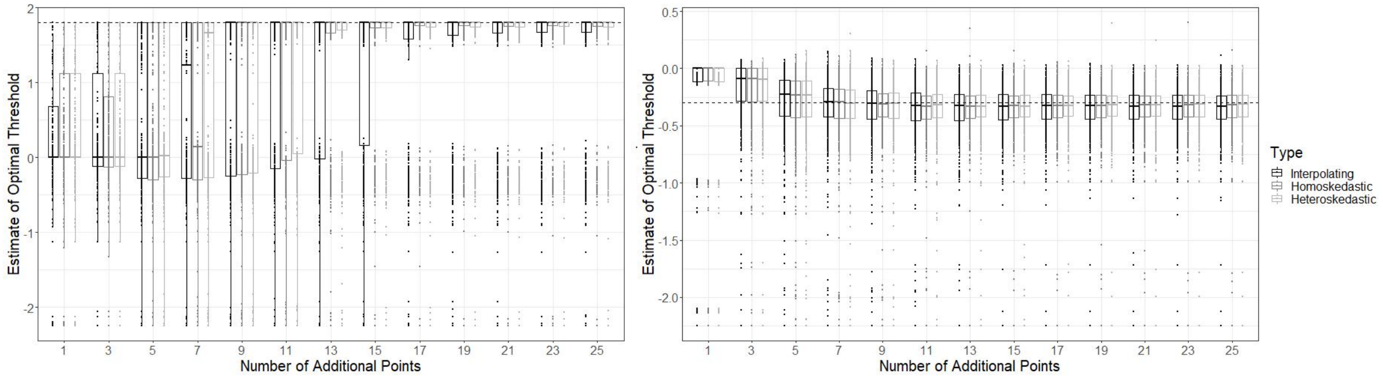
Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	0.000 (0.677)	0.000 (2.074)	1.800 (2.012)	1.800 (1.638)	1.800 (0.141)	1.800 (0.125)
Med. (IQR)	HM$\mathcal{GP}$	0.000 (1.125)	0.000 (2.097)	1.800 (1.964)	1.800 (0.067)	1.800 (0.035)	1.800 (0.043)
Med. (IQR)	HE$\mathcal{GP}$	0.000 (1.125)	0.019 (2.061)	1.800 (1.921)	1.800 (0.067)	1.800 (0.053)	1.800 (0.056)
Mean (SD)	Int$\mathcal{GP}$	0.054 (1.006)	0.394 (1.235)	0.933 (1.180)	1.190 (1.060)	1.296 (0.996)	1.348 (0.944)
Mean (SD)	HM$\mathcal{GP}$	0.074 (1.015)	0.296 (1.235)	1.047 (1.151)	1.383 (0.932)	1.492 (0.801)	1.500 (0.801)
Mean (SD)	HE$\mathcal{GP}$	0.074 (1.032)	0.410 (1.289)	1.086 (1.100)	1.404 (0.919)	1.479 (0.831)	1.502 (0.796)

Table B.23: Simulation II: Estimated optimal ψ_2 after +m points; $n = 1000$ with 16 design points over 500 replicates, $Matern_{5/2}$ covariance, and Log-Normal prior. True ψ_{2opt} =-0.3.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	0.000 (0.111)	-0.226 (0.309)	-0.316 (0.248)	-0.328 (0.219)	-0.328 (0.207)	-0.329 (0.203)
Med. (IQR)	HM$\mathcal{GP}$	0.000 (0.110)	-0.232 (0.324)	-0.317 (0.200)	-0.325 (0.192)	-0.318 (0.189)	-0.319 (0.195)
Med. (IQR)	HE$\mathcal{GP}$	0.000 (0.112)	-0.230 (0.316)	-0.317 (0.220)	-0.329 (0.190)	-0.321 (0.185)	-0.313 (0.191)
Mean (SD)	Int$\mathcal{GP}$	-0.247 (0.555)	-0.338 (0.425)	-0.366 (0.307)	-0.373 (0.271)	-0.373 (0.256)	-0.373 (0.255)
Mean (SD)	HM$\mathcal{GP}$	-0.248 (0.555)	-0.345 (0.421)	-0.365 (0.286)	-0.357 (0.243)	-0.357 (0.235)	-0.353 (0.230)
Mean (SD)	HE$\mathcal{GP}$	-0.251 (0.555)	-0.336 (0.404)	-0.363 (0.284)	-0.363 (0.273)	-0.352 (0.223)	-0.353 (0.236)

Table B.24: Simulation II: Estimated value at $\hat{\psi}_{1opt}, \hat{\psi}_{2opt}$ after +m points; $n = 1000$ with 16 design points over 500 replicates, $Matern_{5/2}$ covariance, and Log-Normal prior. True value at ψ_{1opt}, ψ_{2opt}: 1.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	0.167 (0.089)	0.217 (0.103)	0.246 (0.100)	0.254 (0.092)	0.255 (0.089)	0.256 (0.085)
Med. (IQR)	HM$\mathcal{GP}$	0.164 (0.092)	0.206 (0.107)	0.246 (0.091)	0.254 (0.086)	0.255 (0.082)	0.256 (0.080)
Med. (IQR)	HE$\mathcal{GP}$	0.164 (0.092)	0.214 (0.105)	0.246 (0.092)	0.252 (0.087)	0.254 (0.081)	0.256 (0.081)
Mean (SD)	Int$\mathcal{GP}$	0.173 (0.067)	0.216 (0.075)	0.241 (0.073)	0.251 (0.070)	0.255 (0.069)	0.256 (0.068)
Mean (SD)	HM$\mathcal{GP}$	0.171 (0.068)	0.208 (0.077)	0.242 (0.072)	0.251 (0.068)	0.254 (0.066)	0.255 (0.066)
Mean (SD)	HE$\mathcal{GP}$	0.172 (0.068)	0.214 (0.077)	0.242 (0.071)	0.252 (0.068)	0.255 (0.066)	0.255 (0.066)



(a)

(b)

Figure B.24: Simulation II: Boxplot after $+m$ points; $n = 1000$ with 16 design points, $Matern_{5/2}$ covariance, and Log-Normal prior (a) Optimal ψ_1 (b) Optimal ψ_2.

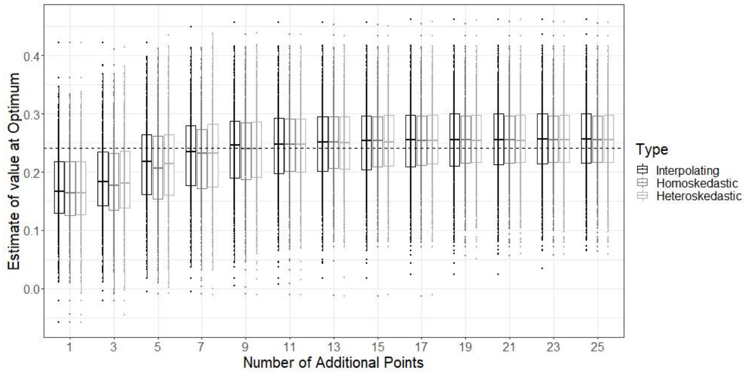


Figure B.25: Boxplot of value at optimum after $+m$ points; $n = 1000$ with 16 design points, $Matern_{5/2}$ covariance, and Log-Normal prior.

B.5 Simulation III

In this section, we present additional results for other simulation settings in simulation III. Before this, we provide the full data-generating mechanism:

- $x_{1,1} \sim N(1,1)$, $x_{1,2} \sim N(0,1)$
- $z_1 \sim Bern(p = expit(1.5x_{1,2} + 2x_{1,1}))$
- $x_{2,1} \sim N(0.2z_1 + 0.1x_{1,1}, 1)$, $x_{2,2} \sim N(0.5z_1 + 0.1x_{1,2}, 1)$
- $z_2 \sim \text{Bern}(p = expit(1.5x_{2,2} - 0.6z_1 + 2x_{2,1}))$
- $z_{1,opt} = 0.5x_{1,1} + 0.5x_{1,2} > 0.5$, $z_{2,opt} = 0.5x_{2,1} + 0.5x_{2,2} > 0.5$
- $y = x_{11} + x_{12} - (0.5x_{11} + 0.5x_{12} - 0.5)(z_{1,opt} - z_1) - (0.5x_{21} + 0.5Ox_{22} - 0.5)(z_{2,opt} - z_2) + \sqrt{0.5}\epsilon$,
  $\epsilon \sim N(0,1)$.

From Figure B.25, we see that the mean value estimates well the optimal parameters, and this reflects the fact that this is a uni-modal problem.

Table B.25: Simulation III: Grid search results for simulation III in increments of 0.01 and $n = 1000$. True $(\psi_{1opt}, \psi_{2opt}) = (0.5,\ 0.1)$; true value at optimum: 1.

Statistic	$\hat{\psi}_{1opt}$	$\hat{\psi}_{3opt}$	Value at Optimum
Mean (SD)	0.479 (0.132)	0.111 (0.100)	1.143 (0.103)
Median (IQR)	0.480 (0.170)	0.11 (0.130)	1.141 (0.131)

B.5.1 Simulation III: $Matern_{5/2}$ Covariance; Sample Size $n = 500$

The following tables present the means and standard deviations pertaining to the estimated parameters of interest. These complement the tables in the main paper, showing the medians and IQRs.

Table B.26: Simulation III: Estimated optimal ψ_1 after $+m$ points, mean (SD); $n = 500$ with 20 design points over 500 replicates, and $Matern_{5/2}$ covariance. True ψ_{1opt} =0.5.

$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Int$\mathcal{GP}$	0.479 (0.172)	0.480 (0.164)	0.482 (0.161)	0.477 (0.161)	0.478 (0.16)	0.479 (0.157)
HM$\mathcal{GP}$	0.477 (0.178)	0.488 (0.170)	0.485 (0.169)	0.483 (0.166)	0.482 (0.164)	0.483 (0.165)
HE$\mathcal{GP}$	0.480 (0.174)	0.481 (0.171)	0.476 (0.163)	0.479 (0.161)	0.478 (0.161)	0.479 (0.160)

Table B.27: Simulation III: Estimated optimal ψ_3 after $+m$ points, mean (SD); $n = 500$ with 20 design points over 500 replicates, and $Matern_{5/2}$ covariance. True $\psi_{3opt} = 0.1$.

$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Int$\mathcal{GP}$	0.107 (0.139)	0.105 (0.136)	0.102 (0.132)	0.099 (0.131)	0.100 (0.128)	0.098 (0.129)
HM$\mathcal{GP}$	0.114 (0.137)	0.111 (0.137)	0.106 (0.136)	0.107 (0.134)	0.104 (0.132)	0.100 (0.137)
HE$\mathcal{GP}$	0.113 (0.139)	0.113 (0.135)	0.106 (0.136)	0.103 (0.134)	0.104 (0.131)	0.103 (0.130)

Table B.28: Simulation III: Estimated value at $\hat{\psi}_{1opt}, \hat{\psi}_{3opt}$ after $+m$ points, mean (SD); $n = 500$ with 20 design points over 500 replicates, and $Matern_{5/2}$ covariance. True value at ψ_{1opt}, ψ_{3opt}: 1.

$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Int$\mathcal{GP}$	1.127 (0.150)	1.161 (0.154)	1.177 (0.151)	1.188 (0.151)	1.193 (0.149)	1.199 (0.147)
HM$\mathcal{GP}$	1.076 (0.154)	1.115 (0.159)	1.138 (0.157)	1.153 (0.155)	1.163 (0.154)	1.168 (0.152)
HE$\mathcal{GP}$	1.078 (0.153)	1.116 (0.158)	1.139 (0.156)	1.152 (0.154)	1.162 (0.151)	1.168 (0.151)

B.5.2 Simulation III: $Matern_{5/2}$ Covariance; Sample Size $n = 1000$

From Figure B.26, we see that the IPW-surface does not completely capture the shape of this value function, for a specific sample of size $n = 1000$.. Consequently, for this replicate, none of the $\mathcal{GP}$s capture well the optimal parameters after an additional 25 sampled points; this can be seen from Figure B.27.

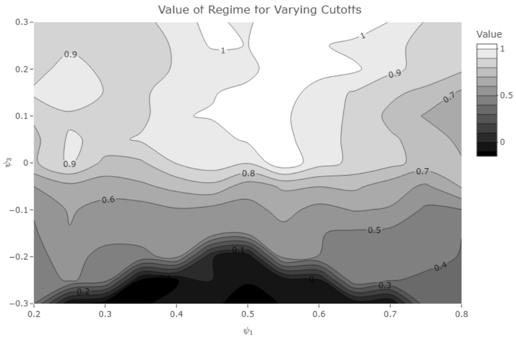
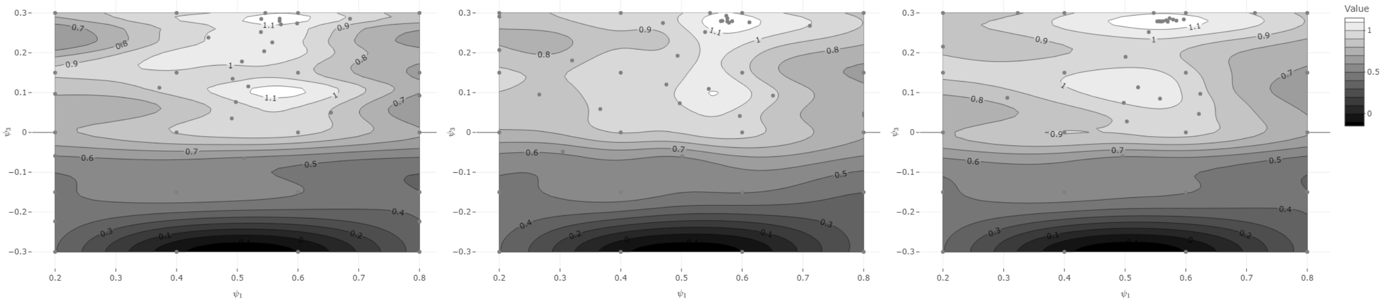


Figure B.26: Simulation III: Estimated value function using normalized IPW.



(a) (b) (c)

Figure B.27: Simulation III: Contour plot at +25 points (a) Int$\mathcal{GP}$ (b) HM$\mathcal{GP}$ (c) HE$\mathcal{GP}$.

We see from Tables B.29, B.30, and B.31 that in this case, unlike the multi-modal cases, all three $\mathcal{GP}$ yield similar results. The increase in sample size brings about more precision as compared to the $n = 500$ case. Additionally, we see from these settings that convergence happens rather quickly, and there is only a slight improvement when sampling additional points up to 25.

Table B.29: Simulation III: Estimated optimal ψ_1 after $+m$ points; $n = 1000$ with 20 design points over 500 replicates, and $Matern_{5/2}$ covariance. True value ψ_{1opt} =0.5.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	0.474 (0.200)	0.472 (0.175)	0.475 (0.170)	0.478 (0.173)	0.485 (0.172)	0.487 (0.178)
Med. (IQR)	HM$\mathcal{GP}$	0.491 (0.200)	0.489 (0.162)	0.478 (0.164)	0.486 (0.164)	0.485 (0.164)	0.484 (0.170)
Med. (IQR)	HE$\mathcal{GP}$	0.489 (0.189)	0.494 (0.165)	0.489 (0.164)	0.485 (0.169)	0.479 (0.172)	0.482 (0.173)
Mean (SD)	Int$\mathcal{GP}$	0.483 (0.141)	0.475 (0.137)	0.477 (0.136)	0.477 (0.137)	0.484 (0.136)	0.486 (0.137)
Mean (SD)	HM$\mathcal{GP}$	0.483 (0.150)	0.482 (0.141)	0.479 (0.138)	0.482 (0.135)	0.484 (0.134)	0.484 (0.135)
Mean (SD)	HE$\mathcal{GP}$	0.483 (0.144)	0.486 (0.140)	0.484 (0.136)	0.483 (0.137)	0.481 (0.137)	0.481 (0.136)

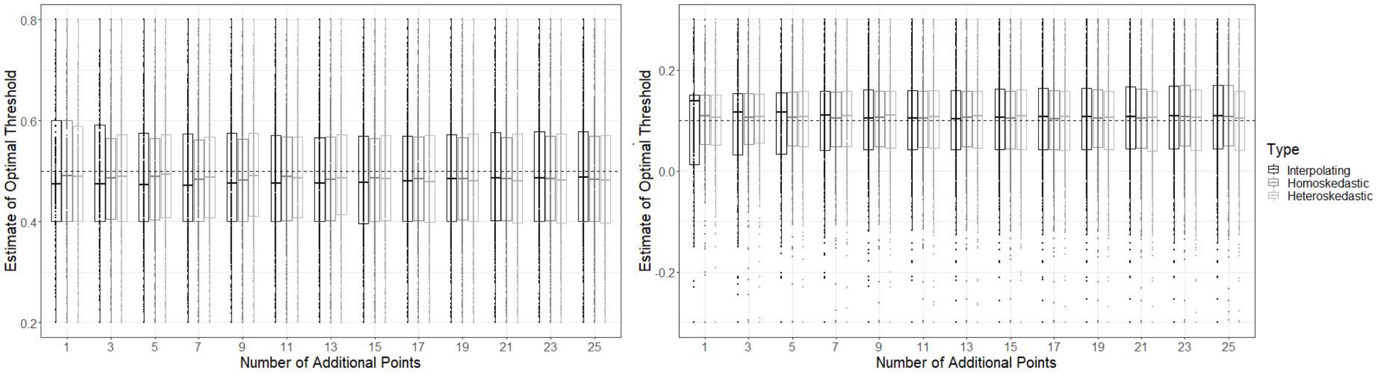
Table B.30: Simulation III: Estimated optimal ψ_3 after $+m$ points; $n = 1000$ with 20 design points over 500 replicates, and $Matern_{5/2}$ covariance. True value ψ_{3opt} =0.1.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	0.139 (0.138)	0.116 (0.121)	0.103 (0.118)	0.106 (0.121)	0.106 (0.121)	0.108 (0.125)
Med. (IQR)	HM$\mathcal{GP}$	0.109 (0.098)	0.106 (0.106)	0.105 (0.108)	0.104 (0.114)	0.105 (0.119)	0.108 (0.121)
Med. (IQR)	HE$\mathcal{GP}$	0.107 (0.098)	0.108 (0.109)	0.110 (0.113)	0.110 (0.119)	0.106 (0.119)	0.104 (0.116)
Mean (SD)	Int$\mathcal{GP}$	0.111 (0.110)	0.109 (0.108)	0.106 (0.107)	0.107 (0.106)	0.107 (0.106)	0.109 (0.105)
Mean (SD)	HM$\mathcal{GP}$	0.110 (0.103)	0.110 (0.101)	0.107 (0.100)	0.105 (0.100)	0.107 (0.100)	0.111 (0.101)
Mean (SD)	HE$\mathcal{GP}$	0.112 (0.105)	0.110 (0.105)	0.105 (0.105)	0.106 (0.105)	0.104 (0.107)	0.104 (0.107)

Table B.31: Simulation III: Estimated value at $\hat{\psi}_{1opt}, \hat{\psi}_{3opt}$ after $+m$ points; $n = 1000$ with 20 design points over 500 replicates, and $Matern_{5/2}$ covariance. True value at ψ_{1opt}, ψ_{3opt}: 1.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	1.060 (0.137)	1.086 (0.132)	1.103 (0.130)	1.109 (0.133)	1.114 (0.129)	1.117 (0.129)
Med. (IQR)	HM$\mathcal{GP}$	1.028 (0.134)	1.053 (0.134)	1.066 (0.140)	1.080 (0.130)	1.089 (0.132)	1.092 (0.133)
Med. (IQR)	HE$\mathcal{GP}$	1.033 (0.137)	1.062 (0.138)	1.073 (0.126)	1.080 (0.133)	1.086 (0.131)	1.091 (0.129)
Mean (SD)	Int$\mathcal{GP}$	1.064 (0.105)	1.091 (0.106)	1.106 (0.104)	1.112 (0.104)	1.115 (0.104)	1.118 (0.104)
Mean (SD)	HM$\mathcal{GP}$	1.028 (0.105)	1.056 (0.108)	1.071 (0.108)	1.080 (0.106)	1.088 (0.105)	1.093 (0.105)
Mean (SD)	HE$\mathcal{GP}$	1.032 (0.104)	1.060 (0.106)	1.075 (0.104)	1.083 (0.104)	1.089 (0.102)	1.093 (0.102)

From Figure B.28, we see that optimal indices are well identified across simulations; from Figure B.29, we see that the value is slightly less well estimated.



(a)

(b)

Figure B.28: Simulation III: Boxplot after $+m$ points; $n = 1000$ with 20 design points, and $Matern_{5/2}$ covariance (a) Optimal ψ_1 (b) Optimal ψ_3.

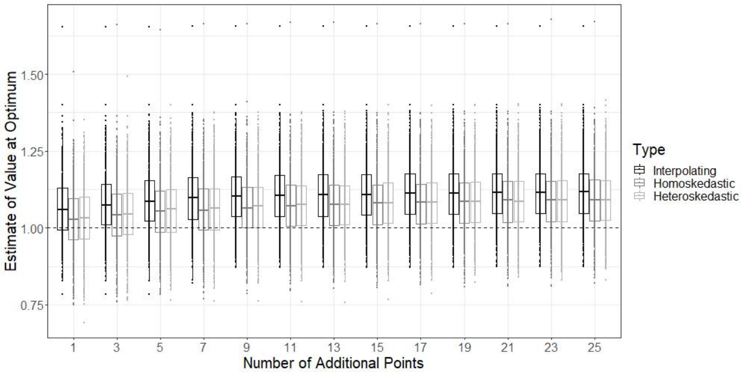
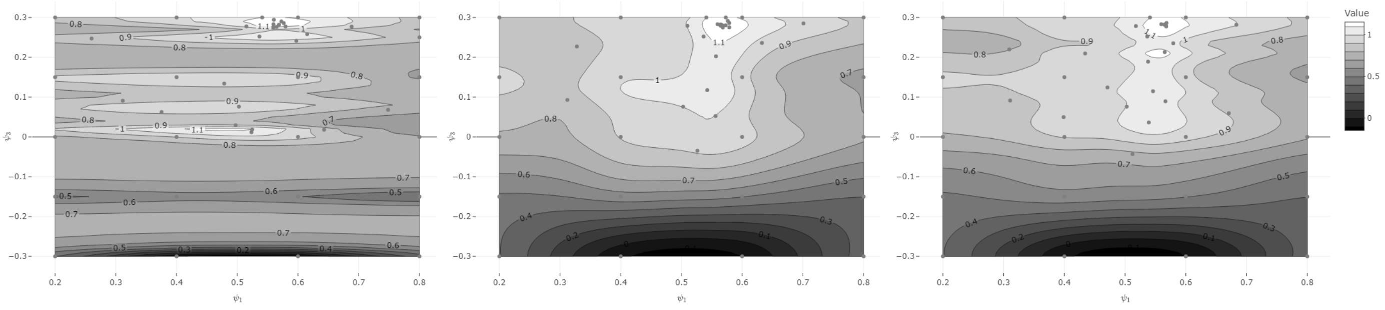


Figure B.29: Simulation III: Boxplot of value at optimum after $+m$ points; $n = 1000$ with 20 design points, and $Matern_{5/2}$ covariance.

B.5.3 Simulation III: $Matern_{3/2}$ Covariance; Sample Size $n = 1000$

We see from Figure B.30 that using a $Matern_{3/2}$ covariance does not improve the fit for this specific replicate.



(a) (b) (c)

Figure B.30: Simulation III: Contour plot at +25 points (a) Int$\mathcal{GP}$ (b) HM$\mathcal{GP}$ (c) HE$\mathcal{GP}$.

From Tables B.32-B.34, we note that the use of a Matérn 3/2 covariance provides slightly more precise results in this setting, but not sufficiently so to determine that the choice between the two covariance functions we explore is consequential in the estimation results.

Table B.32: Simulation III: Estimated optimal ψ_1 after $+m$ points; $n = 1000$ with 20 design points over 500 replicates, and $Matern_{3/2}$ covariance. True ψ_{1opt} =0.5.

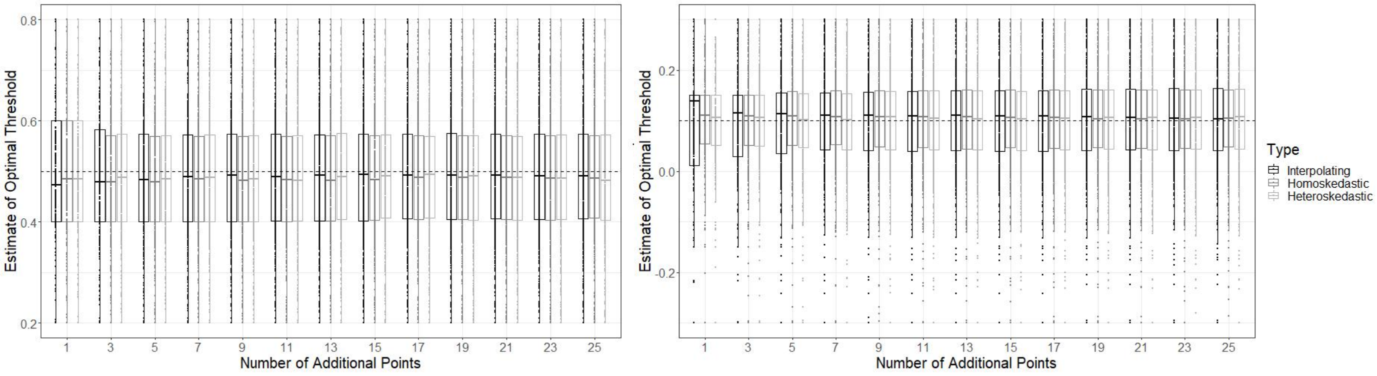
Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	0.473 (0.200)	0.483 (0.174)	0.490 (0.173)	0.493 (0.172)	0.493 (0.168)	0.491 (0.167)
Med. (IQR)	HM$\mathcal{GP}$	0.485 (0.200)	0.479 (0.170)	0.485 (0.169)	0.483 (0.168)	0.487 (0.168)	0.487 (0.163)
Med. (IQR)	HE$\mathcal{GP}$	0.485 (0.200)	0.485 (0.171)	0.487 (0.169)	0.490 (0.164)	0.488 (0.166)	0.482 (0.169)
Mean (SD)	Int$\mathcal{GP}$	0.482 (0.142)	0.481 (0.133)	0.484 (0.132)	0.486 (0.131)	0.489 (0.132)	0.488 (0.132)
Mean (SD)	HM$\mathcal{GP}$	0.480 (0.147)	0.481 (0.137)	0.482 (0.136)	0.482 (0.131)	0.483 (0.132)	0.485 (0.133)
Mean (SD)	HE$\mathcal{GP}$	0.481 (0.145)	0.482 (0.140)	0.484 (0.136)	0.487 (0.135)	0.484 (0.134)	0.484 (0.134)

Table B.33: Simulation III: Estimated optimal ψ_3 after $+m$ points; $n = 1000$ with 20 design points over 500 replicates, and $Matern_{3/2}$ covariance. True ψ_{3opt} =:0.1.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	0.139 (0.139)	0.114 (0.120)	0.110 (0.118)	0.109 (0.120)	0.106 (0.121)	0.103 (0.123)
Med. (IQR)	HM$\mathcal{GP}$	0.111 (0.096)	0.109 (0.108)	0.108 (0.111)	0.106 (0.114)	0.104 (0.115)	0.105 (0.113)
Med. (IQR)	HE$\mathcal{GP}$	0.106 (0.098)	0.101 (0.107)	0.105 (0.115)	0.103 (0.118)	0.106 (0.119)	0.107 (0.118)
Mean (SD)	Int$\mathcal{GP}$	0.111 (0.109)	0.109 (0.105)	0.107 (0.102)	0.108 (0.102)	0.108 (0.101)	0.108 (0.102)
Mean (SD)	HM$\mathcal{GP}$	0.113 (0.105)	0.110 (0.103)	0.107 (0.100)	0.107 (0.100)	0.106 (0.101)	0.107 (0.101)
Mean (SD)	HE$\mathcal{GP}$	0.112 (0.104)	0.104 (0.103)	0.105 (0.103)	0.104 (0.102)	0.106 (0.102)	0.108 (0.101)

Table B.34: Simulation III: Value at $\hat{\psi}_{1opt}, \hat{\psi}_{3opt}$ after $+m$ points; $n = 1000$ with 20 design points over 500 replicates, and $Matern_{3/2}$ covariance. True value at ψ_{1opt}, ψ_{3opt}: 1.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	1.061 (0.133)	1.089 (0.133)	1.103 (0.138)	1.113 (0.134)	1.117 (0.136)	1.121 (0.132)
Med. (IQR)	HM$\mathcal{GP}$	1.033 (0.136)	1.061 (0.134)	1.080 (0.132)	1.097 (0.133)	1.100 (0.134)	1.104 (0.133)
Med. (IQR)	HE$\mathcal{GP}$	1.037 (0.136)	1.069 (0.133)	1.087 (0.128)	1.095 (0.124)	1.102 (0.125)	1.107 (0.132)
Mean (SD)	Int$\mathcal{GP}$	1.064 (0.103)	1.092 (0.106)	1.107 (0.107)	1.115 (0.106)	1.120 (0.105)	1.125 (0.104)
Mean (SD)	HM$\mathcal{GP}$	1.031 (0.105)	1.062 (0.108)	1.080 (0.108)	1.093 (0.108)	1.100 (0.107)	1.106 (0.106)
Mean (SD)	HE$\mathcal{GP}$	1.035 (0.104)	1.067 (0.108)	1.086 (0.105)	1.096 (0.104)	1.103 (0.105)	1.107 (0.103)



(a) (b)

Figure B.31: Simulation III: Boxplot after $+m$ points; $n = 1000$ with 20 design points, and $Matern_{3/2}$ covariance (a) Optimal ψ_1 (b) Optimal ψ_3.

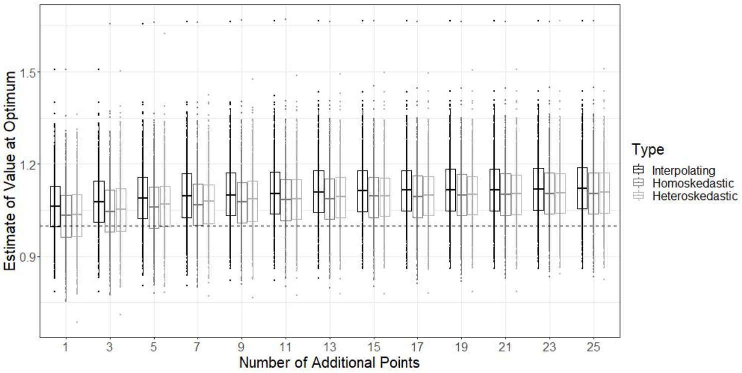


Figure B.32: Boxplot of value at optimum after $+m$ points; $n = 1000$ with 20 design points, and $Matern_{3/2}$ covariance.

B.5.4 Simulation III: $Matern_{5/2}$ Covariance; Sample Size $n = 1000$; Log-Normal Prior

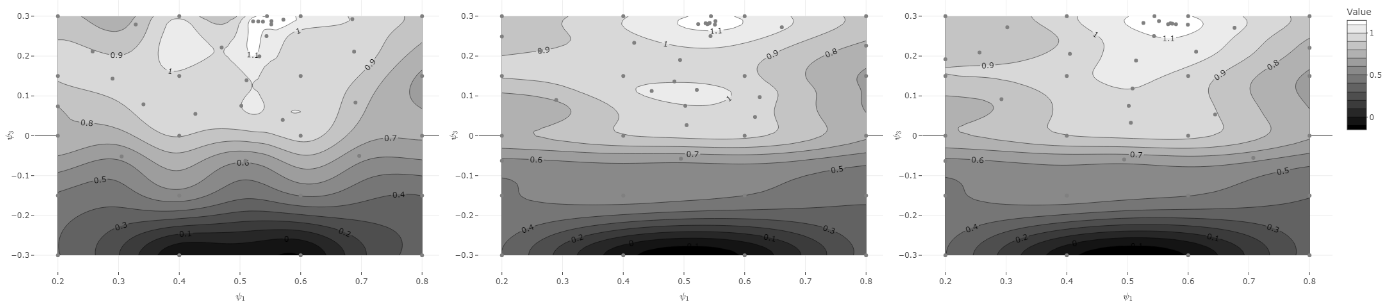


Figure B.33: Simulation III: Contour plot at +25 points: (a) Int$\mathcal{GP}$ (b) HM$\mathcal{GP}$ (c) HE$\mathcal{GP}$.

Table B.35: Simulation III: Estimated optimal ψ_1 after $+m$ points; $n = 1000$ with 20 design points over 500 replicates, $Matern_{5/2}$ covariance, and Log-Normal prior. True ψ_{1opt} =0.5.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	0.466 (0.200)	0.480 (0.177)	0.475 (0.170)	0.477 (0.170)	0.475 (0.167)	0.477 (0.169)
Med. (IQR)	HM$\mathcal{GP}$	0.456 (0.200)	0.476 (0.180)	0.483 (0.179)	0.482 (0.173)	0.485 (0.173)	0.485 (0.168)
Med. (IQR)	HE$\mathcal{GP}$	0.465 (0.200)	0.470 (0.169)	0.472 (0.168)	0.475 (0.166)	0.480 (0.169)	0.483 (0.167)
Mean (SD)	Int$\mathcal{GP}$	0.481 (0.142)	0.479 (0.138)	0.478 (0.139)	0.477 (0.137)	0.478 (0.136)	0.477 (0.135)
Mean (SD)	HM$\mathcal{GP}$	0.479 (0.143)	0.481 (0.134)	0.483 (0.136)	0.482 (0.134)	0.484 (0.133)	0.484 (0.132)
Mean (SD)	HE$\mathcal{GP}$	0.482 (0.141)	0.479 (0.136)	0.477 (0.135)	0.477 (0.136)	0.481 (0.135)	0.483 (0.134)

Table B.36: Simulation III: Estimated optimal ψ_3 after $+m$ points; $n = 1000$ with 20 design points over 500 replicates, $Matern_{5/2}$ covariance, and Log-Normal prior. True ψ_{3opt} =0.1.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	0.143 (0.132)	0.116 (0.126)	0.112 (0.120)	0.110 (0.118)	0.110 (0.120)	0.110 (0.122)
Med. (IQR)	HM$\mathcal{GP}$	0.139 (0.140)	0.111 (0.121)	0.111 (0.123)	0.109 (0.123)	0.103 (0.124)	0.101 (0.123)
Med. (IQR)	HE$\mathcal{GP}$	0.135 (0.135)	0.107 (0.119)	0.110 (0.117)	0.107 (0.113)	0.105 (0.121)	0.102 (0.122)
Mean (SD)	Int$\mathcal{GP}$	0.113 (0.108)	0.112 (0.104)	0.111 (0.104)	0.111 (0.104)	0.109 (0.105)	0.110 (0.105)
Mean (SD)	HM$\mathcal{GP}$	0.113 (0.108)	0.107 (0.101)	0.109 (0.102)	0.107 (0.103)	0.107 (0.104)	0.107 (0.102)
Mean (SD)	HE$\mathcal{GP}$	0.111 (0.110)	0.110 (0.101)	0.110 (0.101)	0.109 (0.100)	0.108 (0.100)	0.107 (0.100)

Table B.37: Simulation III: Estimated value at $\hat{\psi}_{1opt}, \hat{\psi}_{3opt}$ after $+m$ points; $n = 1000$ with 20 design points over 500 replicates, $Matern_{5/2}$ covariance, and Log-Normal prior. True value at ψ_{1opt}, ψ_{3opt}: 1.

Measure	$\mathcal{GP}$	+1	+5	+10	+15	+20	+25
Med. (IQR)	Int$\mathcal{GP}$	1.062 (0.132)	1.089 (0.127)	1.105 (0.135)	1.112 (0.134)	1.113 (0.131)	1.116 (0.132)
Med. (IQR)	HM$\mathcal{GP}$	1.055 (0.138)	1.071 (0.127)	1.084 (0.129)	1.096 (0.132)	1.098 (0.133)	1.099 (0.132)
Med. (IQR)	HE$\mathcal{GP}$	1.054 (0.134)	1.077 (0.126)	1.083 (0.125)	1.090 (0.126)	1.096 (0.131)	1.098 (0.134)
Mean (SD)	Int$\mathcal{GP}$	1.064 (0.105)	1.092 (0.107)	1.106 (0.105)	1.112 (0.105)	1.115 (0.105)	1.117 (0.104)
Mean (SD)	HM$\mathcal{GP}$	1.058 (0.104)	1.074 (0.106)	1.086 (0.106)	1.093 (0.105)	1.098 (0.106)	1.101 (0.105)
Mean (SD)	HE$\mathcal{GP}$	1.058 (0.102)	1.077 (0.106)	1.085 (0.103)	1.091 (0.101)	1.096 (0.101)	1.099 (0.102)

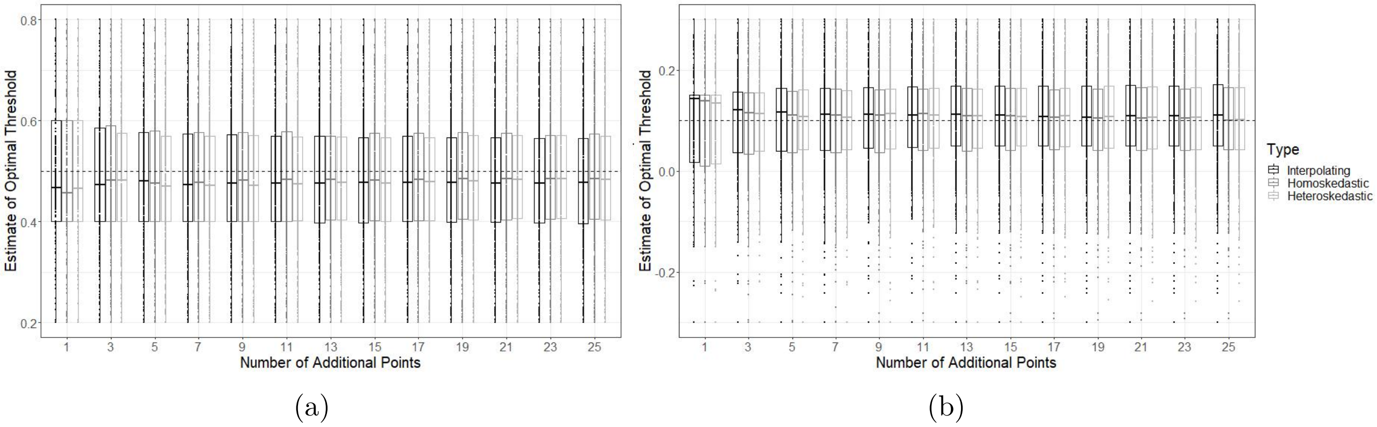


Figure B.34: Simulation III: Boxplot after $+m$ points; $n = 1000$ with 20 design points, $Matern_{5/2}$ covariance, and Log-Normal prior (a) Optimal ψ_1 (b) Optimal ψ_3.

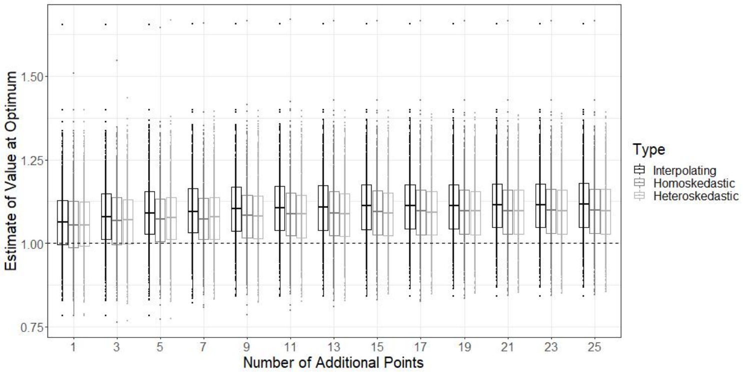


Figure B.35: Simulation III: Boxplot of value at optimum after $+m$ points; $n = 1000$ with 20 design points, $Matern_{5/2}$ covariance, and Log-Normal prior.

B.6 Additional Details on the HIV Treatment Analysis

In this section we explore some additional details related to the case study. We begin by discussing measures to improve efficiency in inferring about the optimal regime. Obtaining sample paths from the $\mathcal{GP}$ amounts to sampling from a multivariate Normal distribution; the dimension of the multivariate Normal depends on the number of points in the sample path. A very fine grid may be the obvious choice, but this can be computationally burdensome as the time it takes to sample from a multivariate Normal vector of dimension d does not grow linearly in d. One approach that helps in this regard is to sample not from the multivariate Normal variable directly but rather to use the property that the conditional and marginal distributions of a multivariate Normal vector are multivariate Normal and to sample from these lower dimensional variables. Although this fact may facilitate sampling from very high dimensional distributions, we must also keep in mind modeling constraints; two points that are very close together will be almost perfectly correlated thereby leading to covariance matrices that are near singular. The general approach should then be to use a grid that is sufficiently fine for the exploration of interest but also sufficiently coarse so that issues with covariance matrix singularity will not arise.

Another approach to speed up computation time when generating sample paths from the multivariate Normal may be to only sample the paths in the vicinity of the optimum, as identified by the posterior mean. The reasoning for this is that points far from the optimum will have low correlation, thereby making them unlikely candidates to be an optimum. Having low correlation across large distances is data-dependent. For this problem, the correlation remains high across the range of possible distances, therefore not allowing us to use examine this approach. Figure 35 shows the correlation for the homoskedastic fit after exploring +25 points, and we see that a strong correlation remains even at large distances.

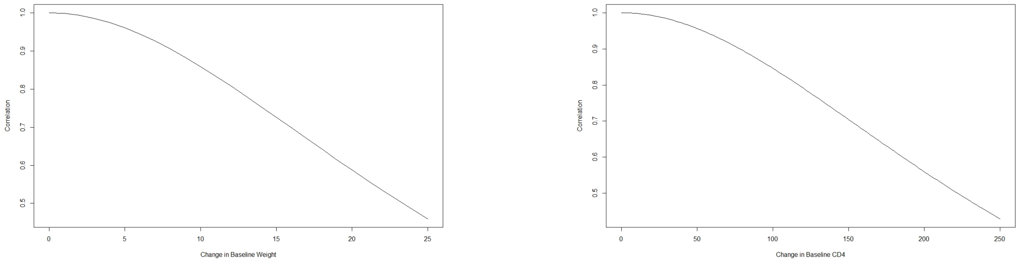


Figure B.36: HIV Study: $Matern_{5/2}$ correlation (a) change in weight (b) change in CD4.

Figure B.38 shows the contour plots for the posterior mean surface using each $\mathcal{GP}$ modeling approach with a $Matern_{3/2}$. We note that in this case, the Int$\mathcal{GP}$ is more in agreement with the other fitting procedures, in contrast with the $Matern_{5/2}$ scenario where it identified two troughs rather than one. We also see that the homogenous and heteroskedastic models follow very similar exploration paths, thereby suggesting that the heteroskedastic model has identified a low level of heteroskedasticity. The Interactive Supplement allows us to further explore the obtained response surfaces.

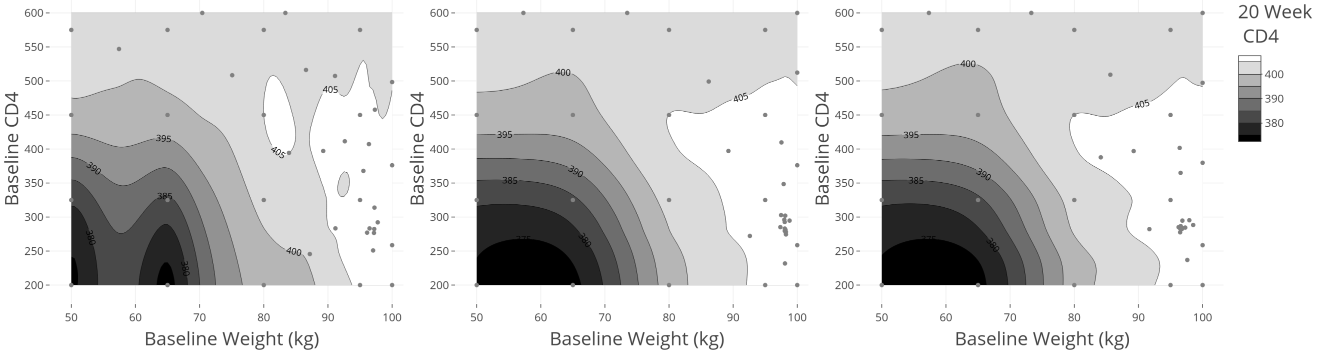


Figure B.37: HIV Study: Emulation surface with $Matern_{3/2}$ covariance after +25 points (a) Int$\mathcal{GP}$ (b) HM$\mathcal{GP}$ (c) HE$\mathcal{GP}$.

Table B.38: HIV Study: Estimates and 95% credible intervals for each $\mathcal{GP}$ modeling strategy; 250 sample paths; 500 Bayesian bootstrapped samples; $Matern_{3/2}$ covariance.

$\mathcal{GP}$	Variable	+1	+5	+15	+25
Int$\mathcal{GP}$	$\hat{\psi}_W^{opt}$	94 (58-98)	98 (58-98)	98 (58-98)	98 (58-98)
Int$\mathcal{GP}$	$\hat{\psi}_{CD4}^{opt}$	372.5 (200-597.5)	342.5 (200-597.5)	305 (200-597.5)	305 (200-597.5)
Int$\mathcal{GP}$	20 Week CD4	409.8 (397.7-422.8)	409.0 (397.5-421.2)	409.3 (398.0-422.8)	409.6 (397.8-422.8)
HM$\mathcal{GP}$	$\hat{\psi}_W^{opt}$	94 (58-98)	98 (58-98)	98 (74-98)	98 (78-98)
HM$\mathcal{GP}$	$\hat{\psi}_{CD4}^{opt}$	372.5 (200-597.5)	335 (200-597.5)	290 (200-575)	290 (200-530)
HM$\mathcal{GP}$	20 Week CD4	409.7 (397.7-422.7)	408.7 (397.2-420.9)	408.6 (397.0-420.7)	408.5 (397.5-421.0)
HE$\mathcal{GP}$	$\hat{\psi}_W^{opt}$	94 (58-98)	98 (58-98)	98 (74-98)	98 (78-98)
HE$\mathcal{GP}$	$\hat{\psi}_{CD4}^{opt}$	372.5 (200-597.5)	327.5 (200-597.5)	290 (200-545)	282.5 (200-530)
HE$\mathcal{GP}$	20 Week CD4	409.7 (397.6-422.5)	408.7 (397.36-420.8)	408.5 (397.0-420.7)	408.6 (397.0-421.1)

Increments for the sample paths were by $4kg$ in the ψ_W axis and by 7.5 cells/mm^3 in the ψ_{CD4} axis

Now we look at the results for the Log-Normal prior. We see from the figures and tables below that the results are essentially the same as those obtained without the prior.

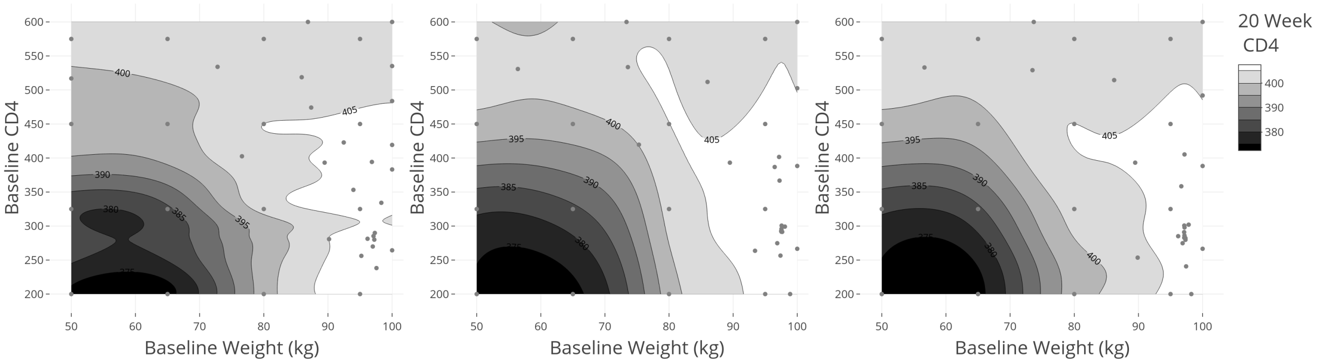


Figure B.38: HIV Study: Emulation surface after +25 points with $Matern_{5/2}$ covariance and Log-Normal prior (a) Int$\mathcal{GP}$ (b) HM$\mathcal{GP}$ (c) HM$\mathcal{GP}$.

Table B.39: HIV Study: Estimates and 95% credible intervals for each $\mathcal{GP}$ modeling strategy; 250 sample paths; 500 Bayesian bootstrapped samples; $Matern_{5/2}$ covariance; Log-Normal prior.

Method	Variable	+1	+5	+15	+25
Int$\mathcal{GP}$	$\hat{\psi}_W^{opt}$	94 (58-98)	98 (58-98)	94 (54-98)	94 (54-98)
Int$\mathcal{GP}$	$\hat{\psi}_{CD4}^{opt}$	372.5 (200-597.5)	335 (200-597.5)	365 (200-597.5)	365 (200-597.5)
Int$\mathcal{GP}$	20 Week CD4	409.5 (397.5-422.4)	408.7 (397.4-421.3)	409.8 (397.5-424.6)	410.1 (397.8-425.1)
HM$\mathcal{GP}$	$\hat{\psi}_W^{opt}$	94 (50-98)	98 (58-98)	98 (78-98)	98 (78-98)
HM$\mathcal{GP}$	$\hat{\psi}_{CD4}^{opt}$	372.5 (200-597.5)	312.5 (200-97.5)	290 (200-530)	290 (200-485)
HM$\mathcal{GP}$	20 Week CD4	412.3 (397.9-436.4)	408.7 (397.3-421.1)	408.3 (397.1-420.8)	408.4 (397.3-420.9)
HE$\mathcal{GP}$	$\hat{\psi}_W^{opt}$	94 (54-98)	98 (58-98)	98 (78-98)	98 (78-98)
HE$\mathcal{GP}$	$\hat{\psi}_{CD4}^{opt}$	372.5 (200-597.5)	305 (200-597.5)	290 (200-522.5)	290 (200-507.5)
HE$\mathcal{GP}$	20 Week CD4	411.0 (397.7-431.4)	408.6 (397.2-421.3)	408.4 (397.03-420.6)	408.4 (397.3-420.7)

APPENDIX C

Appendix to Manuscript 3

In this Appendix, we more systematically lay out the parameters required to use the functions in the `BayesDTR` package.

C.1 Parameters for BayesMSM function

We begin by examining the parameters in the `BayesMSM` function.

```
BayesMSM(PatID,Data,Outcome_Var,Treat_Vars,Treat_M_List,Outcome_M_List,
  MSM_Model,G_List,Psi,Bayes=TRUE,DR=FALSE,Normalized=FALSE,B=100,Bayes_Seed=1)
```

Aim 1: Marginal Structural Model

- **Required**: `PatID`, `Data`, `Outcome_Var`, `Treat_Vars`, `Treat_M_List`, `MSM_Model`, `G_List`, `Psi`
- **Optional**: `Bayes`, `B`, `Bayes_Seed`
- **Unavailable**: `Normalized`

Aim 2: Grid-Search IPW Estimator

- **Required**: PatID, Data, Outcome_Var, Treat_Vars, Treat_M_List, G_List, Psi
- **Optional**: Bayes, Normalized, B, Bayes_Seed

Aim 3: Grid-Search Doubly Robust Estimator

- **Required**: PatID, Data, Outcome_Var, Treat_Vars, Treat_M_List, Outcome_M_List, G_List, Psi, DR=TRUE
- **Optional**: Bayes, Normalized, B, Bayes_Seed

C.2 Parameters for Gaussian Process Functions

We now examine the parameters required to perform the Gaussian Process optimization.

```
DesignFit(PatID,Data,Outcome_Var,Treat_Vars,Treat_M_List,Outcome_M_List,
         Normalized=TRUE,DR=FALSE,G_List,Psi,
         Covtype,Numbr_Samp,IthetasU,IthetasL,Likelihood_Limits=NA,
         Prior_List=NULL,Prior_Der_List=NULL)
```

- **Required**: PatID, Data, Outcome_Var, Treat_Vars, Treat_M_List, G_List Numbr_Samp, IthetasU, IthetasL, Covtype
  - Note: The default is to use the normalized IPW estimator.
- **Optional**: Outcome_M_List, Normalize, DR, Likelihood_Limits, Prior_List, Prior_Der_List
  - Note: The doubly robust estimator can be used by specifying the Outcome_M_List and DR parameters.

```
SequenceFit(Previous_Fit,Additional_Samp,Control_Genoud=list())
```

- **Required**: Previous_Fit, Additional_Samp, Control_Genoud
  - Note that in particular the Control_Genoud function requires the Domain element. We have demonstrated the use of this parameter in the main paper (section 3.2).

```
FitInfer(Design_Object,Boot_Start,Boot_End,Psi_new,N,Location,
         Additional_Samp,Control_Genoud=list())
```

- **Required**: `Design_Object`, `Boot_Start`, `Boot_End`, `Psi_new`, `N`, `Additional_Samp`, `Control_Genoud`
- **Optional**: `Location`

References

B. Ankenman, B. L. Nelson, and J. Staum. Stochastic kriging for simulation metamodeling. In *2008 Winter Simulation Conference*, pages 362–370. IEEE, 2008.

E. Arjas. Causal inference from observational data: A Bayesian predictive approach. In C. Berzuini, P. Dawid, and L. Bernardinell, editors, *Causality: Statistical Perspectives and Applications*, chapter 7. John Wiley & Sons, Chichester, West Sussex, 2012.

E. Arjas and O. Saarela. Optimal dynamic regimes: Presenting a case for predictive inference. *The International Journal of Biostatistics*, 6(2), 2010.

W. Artman. *SMARTbayesR: Bayesian set of best dynamic treatment regimes and sample size in SMARTs for Binary Outcomes*, 2021. URL https://CRAN.R-project.org/package=SMARTbayesR. R package version 2.0.0.

H. Bang and J. M. Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–972, 2005.

H. Bang and J. M. Robins. Correction to “doubly robust estimation in missing data and causal inference models” by H. Bang and J. M. Robins; 61(4), 962–972, 2005. *Biometrics*, 64(2):650–650, 2008.

J. M. Bernardo and A. F. Smith. *Bayesian Theory*, volume 405. John Wiley & Sons, Toronto, 2009.

P. G. Bissiri, C. C. Holmes, and S. G. Walker. A general framework for updating belief distributions. *Journal of the Royal Statistical Society: Series B*, 78(5):1103–1130, 2016.

D. Blackwell and J. B. MacQueen. Ferguson distributions via Pólya urn schemes. *The Annals of Statistics*, 1(2):353–355, 1973.

L. E. Cain, J. M. Robins, E. Lanoy, R. Logan, D. Costagliola, and M. A. Hernán. When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data. *The International Journal of Biostatistics*, 6(2), 2010.

B. Chakraborty and E. E. M. Moodie. *Statistical methods for dynamic treatment regimes*. Springer, New York, 2013.

B. Chakraborty, S. A. Murphy, and V. Strecher. Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical Methods in Medical Research*, 19(3):317–343, 2010.

Y. Chen, Y. Liu, D. Zeng, and Y. Wang. *DTRlearn2: Statistical learning methods for optimizing dynamic treatment regimes*, 2020. URL https://CRAN.R-project.org/package=DTRlearn2. R package version 1.1.

H. A. Chipman, E. I. George, and R. E. McCulloch. Bart: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1):266–298, 2010.

C. Currin, T. Mitchell, M. Morris, and D. Ylvisaker. Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association*, 86(416):953–963, 1991.

A. P. Dawid. Causal inference without counterfactuals. *Journal of the American Statistical Association*, 95(450):407–424, 2000.

B. De Finetti. Sul significato soggettivo della probabilità. *Fundamenta Mathematicae*, 17: 298–329, 1931.

T. S. Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230, 1973.

A. I. Forrester, A. J. Keane, and N. W. Bressloff. Design and analysis of "noisy" computer experiments. *AIAA Journal*, 44(10):2331–2339, 2006.

P. I. Frazier and J. Wang. Bayesian optimization for materials design. In T. Lookman, F. J. Alexander, and K. Rajan, editors, *Information Science for Materials Discovery and Design*, pages 45–75. Springer, New York, 2016.

P. I. Frazier, J. Xie, and S. E. Chick. Value of information methods for pairwise sampling with correlations. In *Proceedings of the 2011 Winter Simulation Conference (WSC)*, pages 3974–3986. IEEE, 2011.

M. Gasparini. Exact multivariate Bayesian bootstrap distributions of moments. *The Annals of Statistics*, pages 762–768, 1995.

S. Ghosal. The Dirichlet process, related priors and posterior asymptotics. In N. L. Hjort, C. Holmes, P. Müller, and S. G. Walker, editors, *Bayesian Nonparametrics*, chapter 2. Cambridge University Press, New York, 2010.

S. Ghosal and A. van der Vaart. *Fundamentals of nonparametric Bayesian inference*, volume 44. Cambridge University Press, Cambridge, United Kingdom, 2017.

P. W. Goldberg, C. K. Williams, and C. M. Bishop. Regression with input-dependent noise: A Gaussian process treatment. *Advances in Neural Information Processing Systems*, 10: 493–499, 1997.

S. Greenland. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*, 14(3):300–306, 2003.

S. Greenland, J. Pearl, and J. M. Robins. Causal diagrams for epidemiologic research. *Epidemiology*, 10(1):37–48, 1999.

L. Gunter, J. Zhu, and S. A. Murphy. Variable selection for optimal decision making. In *Proceedings of the 11th Conference on Artificial Intelligence in Medicine*. AIME, 2007.

S. M. Hammer, D. A. Katzenstein, M. D. Hughes, H. Gundacker, R. T. Schooley, R. H. Haubrich, W. K. Henry, M. M. Lederman, J. P. Phair, M. Niu, et al. A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine*, 335(15): 1081–1090, 1996.

M. Henmi and S. Eguchi. A paradox concerning nuisance parameters and projected estimating functions. *Biometrika*, 91(4):929–941, 2004.

M. A. Hernán and J. M. Robins. *Causal inference: What if*. Chapman & Hall/CRC, Boca Raton, 2020.

E. Hewitt and L. J. Savage. Symmetric measures on cartesian products. *Transactions of the American Mathematical Society*, 80(2):470–501, 1955.

P. W. Holland. Statistics and causal inference. *Journal of the American statistical Association*, 81(396):945–960, 1986.

S. T. Holloway, E. B. Laber, K. A. Linn, B. Zhang, M. Davidian, and A. A. Tsiatis. *DynTxRegime: Methods for estimating optimal dynamic treatment regimes*, 2020. URL https://CRAN.R-project.org/package=DynTxRegime. R package version 4.9.

L. Hu, J. Hogan, A. Mwangi, and A. Siika. Modeling the causal effect of treatment initiation time on survival: Application to HIV/TB co-infection. *Biometrics*, 74(2):703–713, 2018.

D. Huang, T. T. Allen, W. I. Notz, and N. Zeng. Global optimization of stochastic black-box systems via sequential kriging meta-models. *Journal of Global Optimization*, 34(3): 441–466, 2006.

B. A. Johnson and A. A. Tsiatis. Estimating mean response as a function of treatment duration in an observational study, where duration may be informatively censored. *Biometrics*, 60(2):315–323, 2004.

B. A. Johnson and A. A. Tsiatis. Semiparametric inference in observational duration-response studies, with duration possibly right-censored. *Biometrika*, 92(3):605–618, 2005.

M. E. Johnson, L. M. Moore, and D. Ylvisaker. Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, 26(2):131–148, 1990.

D. R. Jones, M. Schonlau, and W. J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):455–492, 1998.

J. D. Kang and J. L. Schafer. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22 (4):523–539, 2007.

R. A. Kaslow, D. G. Ostrow, R. Detels, J. P. Phair, B. F. Polk, C. R. RINALDO Jr, and M. A. C. Study. The multicenter AIDS cohort study: rationale, organization, and selected characteristics of the participants. *American Journal of Epidemiology*, 126(2):310–318, 1987.

K. Kersting, C. Plagemann, P. Pfaff, and W. Burgard. Most likely heteroscedastic Gaussian process regression. In *Proceedings of the 24th International Conference on Machine Learning*, pages 393–400, 2007.

M. B. Klein, K. N. Althoff, Y. Jing, B. Lau, M. Kitahata, V. Lo Re III, G. D. Kirk, M. Hull, H. N. Kim, and G. Sebastiani. Risk of end-stage liver disease in HIV-viral hepatitis coinfected persons in North America from the early to modern antiretroviral therapy eras. *Clinical Infectious Diseases*, 63(9):1160–1167, 2016.

E. F. Krakow, M. Hemmer, T. Wang, B. Logan, M. Arora, S. Spellman, D. Couriel, A. Alousi, J. Pidala, M. Last, S. Lachance, and E. E. M. Moodie. Tools for the precision medicine era: How to develop highly personalized treatment recommendations from cohort and registry data using Q-learning. *American Journal of Epidemiology*, 186(2):160–172, 2017.

D. G. Krige. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, 52(6):119–139, 1951.

M. G. Kundu. *LongCART: Recursive partitioning for longitudinal data and right censored data using baseline covariates*, 2021. URL https://CRAN.R-project.org/package=LongCART. R package version 3.1.

H. J. Kushner. A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. *Journal of Fluids Engineering*, 86:97–106, 1964.

J. Lee, P. F. Thall, Y. Ji, and P. Müller. Bayesian dose-finding in two treatment cycles based on the joint utility of efficacy and toxicity. *Journal of the American Statistical Association*, 110(510):711–722, 2015.

D. J. Lizotte. *Practical Bayesian Optimization*. PhD thesis, University of Alberta, Edmonton, AB, Canada, 2008.

M. Locatelli. Bayesian algorithms for one-dimensional global optimization. *Journal of Global Optimization*, 10(1):57–76, 1997.

J. L. Loeppky, J. Sacks, and W. J. Welch. Choosing the sample size of a computer experiment: A practical guide. *Technometrics*, 51(4):366–376, 2009.

B. R. Logan, R. Sparapani, R. E. McCulloch, and P. W. Laud. Decision making and uncertainty quantification for individualized treatments using Bayesian additive regression trees. *Statistical Methods in Medical Research*, 28(4):1079–1093, 2019.

Y. Luo, D. A. Stephens, D. J. Graham, and E. J. McCoy. Bayesian doubly robust causal inference via loss functions. 2022. URL https://arxiv.org/abs/2103.04086.

S. P. Lyddon, C. C. Holmes, and S. G. Walker. General Bayesian updating and the loss-likelihood bootstrap. *Biometrika*, 106(2):465–478, 2019.

S. N. MacEachern. Dependent nonparametric processes. In *ASA proceedings of the section on Bayesian statistical science*, volume 1, pages 50–55. Alexandria, Virginia. American Statistical Association, 1999.

J. Macías et al. Antiretroviral therapy based on protease inhibitors as a protective factor against liver fibrosis progression in patients with chronic hepatitis C. *Antiviral Therapy*, 11(7):839, 2006.

M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.

W. R. Mebane, Jr. and J. S. Sekhon. Genetic optimization using derivatives: The rgenoud package for R. *Journal of Statistical Software*, 42(11):1–26, 2011. URL http://www.jstatsoft.org/v42/i11/.

R. Mitra and P. Müller. *Nonparametric Bayesian inference in biostatistics*. Springer, New York, 2015.

W. Mo, Z. Qi, and Y. Liu. Learning optimal distributionally robust individualized treatment rules. *Journal of the American Statistical Association*, 116(534):659–674, 2021.

E. E. M. Moodie and T. S. Richardson. Estimating optimal dynamic regimes: Correcting bias under the null. *Scandinavian Journal of Statistics*, 37(1):126–146, 2010.

E. E. M. Moodie, B. Chakraborty, and M. S. Kramer. Q-learning for estimating optimal dynamic treatment rules from observational data. *Canadian Journal of Statistics*, 40(4): 629–645, 2012.

P. Müller, F. A. Quintana, A. Jara, and T. Hanson. *Bayesian nonparametric data analysis.* Springer, New York, 2015.

S. A. Murphy. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):331–355, 2003.

S. A. Murphy. An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine*, 24(10):1455–1481, 2005a.

S. A. Murphy. A generalization error for Q-learning. *Journal of Machine Learning Research*, 6:1073–1097, 2005b.

S. A. Murphy, M. J. van der Laan, and J. M. Robins. Marginal mean models for dynamic regimes. *Journal of the American Statistical Association*, 96(456):1410–1423, 2001.

T. A. Murray, Y. Yuan, and P. F. Thall. A Bayesian machine learning approach for optimizing dynamic treatment regimes. *Journal of the American Statistical Association*, 113(523): 1255–1267, 2018.

M. A. Newton and A. E. Raftery. Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society: Series B*, 56(1):3–26, 1994.

A. O'Hagan, M. C. Kennedy, and J. E. Oakley. Uncertainty analysis and other inference tools for complex computer codes. In *Bayesian Statistics 6: Proceedings of the Sixth Valencia International Meeting*, pages 503–524. Oxford University Press, 1999.

L. Orellana, A. Rotnitzky, and J. M. Robins. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part I: Main content. *The International Journal of Biostatistics*, 6(2), 2010a.

L. Orellana, A. Rotnitzky, and J. M. Robins. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part II: Proof of results. *The International Journal of Biostatistics*, 6(2), 2010b.

A. O'Hagan. Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety*, 91(10-11):1290–1300, 2006.

J.-S. Park and J. Baek. Efficient computation of maximum likelihood estimators in a spatial linear model with power exponential covariogram. *Computers & Geosciences*, 27(1):1–7, 2001.

M. L. Petersen, K. E. Porter, S. Gruber, Y. Wang, and M. J. van der Laan. Diagnosing and responding to violations in the positivity assumption. *Statistical Methods in Medical Research*, 21(1):31–54, 2012.

V. Picheny and D. Ginsbourger. Noisy kriging-based optimization methods: a unified implementation within the diceoptim package. *Computational Statistics & Data Analysis*, 71:1035–1053, 2014.

V. Picheny, T. Wagner, and D. Ginsbourger. A benchmark of kriging-based infill criteria for noisy optimization. *Structural and Multidisciplinary Optimization*, 48(3):607–626, 2013.

T. Pourmohamad and H. K. Lee. *Bayesian optimization with application to computer experiments*. Springer, Cham, 2021.

N. Quadrianto, K. Kersting, and Z. Xu. *Gaussian Process*, pages 535–548. Springer US, Boston, MA, 2017. ISBN 978-1-4899-7687-1. doi: 10.1007/978-1-4899-7687-1_108. URL https://doi.org/10.1007/978-1-4899-7687-1_108.

J. M. Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9):1393–1512, 1986.

J. M. Robins. Analytic methods for estimating HIV-treatment and cofactor effects. In D. G. Ostrow and R. C. Kessler, editors, *Methodological Issues in AIDS Behavioral Research*, pages 213–287. Springer, 1993.

J. M. Robins. Optimal structural nested models for optimal sequential decisions. In *Proceedings of the second Seattle Symposium in Biostatistics*, pages 189–326. Springer, 2004.

J. M. Robins, M. A. Hernan, and B. Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, 2000.

J. M. Robins, L. Orellana, and A. Rotnitzky. Estimation and extrapolation of optimal treatment and testing strategies. *Statistics in Medicine*, 27(23):4678–4721, 2008.

D. Rodriguez Duque, D. A. Stephens, and E. E. M. Moodie. Estimation of optimal dynamic treatment regimes using Gaussian processes. 2022a. URL https://arxiv.org/abs/2105.12259.

D. Rodriguez Duque, D. A. Stephens, E. E. M. Moodie, and M. B. Klein. Semiparametric Bayesian inference for dynamic treatment regimes via dynamic regime marginal structural models. *Biostatistics*, 2022b. URL https://doi.org/10.1093/biostatistics/kxac007.

M. Rosenblum and M. J. van der Laan. Targeted maximum likelihood estimation of the parameter of a marginal structural model. *The International Journal of Biostatistics*, 6 (2), 2010.

O. Roustant, D. Ginsbourger, and Y. Deville. Dicekriging, diceoptim: Two R packages for the analysis of computer experiments by kriging-based metamodelling and optimization. *Journal of Statistical Software*, 51(1):1–55, 2012.

D. B. Rubin. Bias reduction using Mahalanobis-metric matching. *Biometrics*, 36(2):293–298, 1980.

D. B. Rubin. The Bayesian bootstrap. *The Annals of Statistics*, 9(1):130–134, 1981.

O. Saarela and E. Arjas. A method for Bayesian monotonic multiple regression. *Scandinavian Journal of Statistics*, 38(3):499–513, 2011.

O. Saarela, E. Arjas, D. A. Stephens, and E. E. M. Moodie. Predictive Bayesian inference and dynamic treatment regimes. *Biometrical Journal*, 57(6):941–958, 2015a.

O. Saarela, D. A. Stephens, E. E. M. Moodie, and M. B. Klein. On Bayesian estimation of marginal structural models. *Biometrics*, 71(2):279–288, 2015b.

O. Saarela, L. R. Belzile, and D. A. Stephens. A Bayesian view of doubly robust causal inference. *Biometrika*, 103(3):667–681, 2016.

J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4(4):409–423, 1989.

T. J. Santner, B. J. Williams, W. Notz, and B. J. Williams. *The design and analysis of computer experiments*. Springer, New York, second edition, 2018.

D. O. Scharfstein, A. Rotnitzky, and J. M. Robins. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94 (448):1096–1120, 1999.

E. F. Schisterman, S. R. Cole, and R. W. Platt. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology*, 20(4):488–495, 2009.

J. Schulz and E. E. M. Moodie. Doubly robust estimation of optimal dosing strategies. *Journal of the American Statistical Association*, 116(533):256–268, 2021.

J. Sethuraman. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4(2):639–650, 1994.

J. Q. Shi and T. Choi. *Gaussian process regression analysis for functional data*. CRC Press, New York, 2011.

S. M. Shortreed and E. E. M. Moodie. Estimating the optimal dynamic antipsychotic treatment regime: evidence from the sequential multiple-assignment randomized clinical antipsychotic trials of intervention and effectiveness schizophrenia study. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(4):577–599, 2012.

G. Simoneau, E. E. M. Moodie, J. S. Nijjar, R. W. Platt, and SERA Inception Cohort Investigators. Estimating optimal dynamic treatment regimes with survival outcomes. *Journal of the American Statistical Association*, 115(531):1531–1539, 2020.

D. A. Stephens. G-estimation for dynamic treatment regimes in the longitudinal setting. In M. R. Kosorok and E. E. M. Moodie, editors, *Adaptive Treatment Strategies in Practice Planning Trials and Analyzing Data for Precision Medicine*, chapter 7. John Wiley & Sons, Philadelphia, 2015.

D. A. Stephens, W. S. Nobre, E. E. M. Moodie, and A. M. Schmidt. Causal inference under mis-specification: adjustment based on the propensity score. 2022. URL https://arxiv.org/abs/2201.12831.

R. K. Sterling et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology*, 43(6):1317–1325, 2006.

Y. W. Teh. *Dirichlet Process*, pages 361–370. Springer US, Boston, MA, 2017. ISBN 978-1-4899-7687-1. doi: 10.1007/978-1-4899-7687-1_219. URL https://doi.org/10.1007/978-1-4899-7687-1_219.

A. A. Tsiatis. *Semiparametric theory and missing data.* Springer, New York, 2007.

A. A. Tsiatis, M. Davidian, S. T. Holloway, and E. B. Laber. *Dynamic treatment regimes: Statistical methods for precision medicine.* Chapman and Hall/CRC, New York, 2019.

M. J. van der Laan and M. L. Petersen. Causal effect models for realistic individualized treatment and intention to treat rules. *The International Journal of Biostatistics*, 3(1), 2007.

A. S. Wahed and P. F. Thall. Evaluating joint effects of induction–salvage treatment regimes on overall survival in acute leukaemia. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62(1):67–83, 2013.

S. G. Walker. Bayesian nonparametric methods: motivation and ideas. In N. L. Hjort, C. Holmes, P. Müller, and S. G. Walker, editors, *Bayesian Nonparametrics*, chapter 1. Cambridge University Press, New York, 2010.

S. G. Walker. Bayesian inference with misspecified models. *Journal of Statistical Planning and Inference*, 143(10):1621–1633, 2013.

M. P. Wallace and E. E. M. Moodie. Doubly-robust dynamic treatment regimen estimation via weighted least squares. *Biometrics*, 71(3):636–644, 2015.

M. P. Wallace, E. E. M. Moodie, and D. A. Stephens. Model validation and selection for personalized medicine using dynamic-weighted ordinary least squares. *Statistical Methods in Medical Research*, 26(4):1641–1653, 2017.

M. P. Wallace, E. E. M. Moodie, D. A. Stephens, G. Simoneau, and J. Schulz. *DTRreg: DTR estimation and inference via g-estimation, dynamic WOLS, Q-learning, and dynamic weighted survival modeling (DWSurv)*, 2020. URL https://CRAN.R-project.org/package=DTRreg. R package version 1.7.

C. Wang. *Gaussian process regression with heteroscedastic residuals and fast MCMC methods.* PhD thesis, University of Toronto, Toronto, ON, Canada, 2014.

C. K. Williams and C. E. Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press, Cambridge, MA, 2006.

Y. Xiao, M. Abrahamowicz, and E. E. M. Moodie. Accuracy of conventional and marginal structural Cox model estimators: a simulation study. *The International Journal of Biostatistics*, 6(2), 2010.

Y. Xu, P. Müller, A. S. Wahed, and P. F. Thall. Bayesian nonparametric estimation for dynamic treatment regimes with sequential transition times. *Journal of the American Statistical Association*, 111(515):921–950, 2016.

J. Yin, S. H. Ng, and K. M. Ng. Kriging metamodel with modified nugget-effect: The heteroscedastic variance case. *Computers & Industrial Engineering*, 61(3):760–777, 2011.

J. Young, V. Lo Re III, H. Kim, T. R. Sterling, K. Althoff, K. Gebo, M. J. Gill, M. A. Horberg, A. M. Mayor, R. D. Moore, M. J. Silverberg, and M. B. Klein. Do contemporary antiretrovirals increase the risk of end-stage liver disease? Signals from patients starting therapy in the NA-ACCORD. *Pharmacoepidemiology and Drug Safety*, 31(2):214–224, 2021.

T. Zajonc. Bayesian inference for dynamic treatment regimes: Mobility, equity, and efficiency in student tracking. *Journal of the American Statistical Association*, 107(497):80–92, 2012.

B. Zhang, A. A. Tsiatis, M. Davidian, M. Zhang, and E. Laber. Estimating optimal treatment regimes from a classification perspective. *Stat*, 1(1):103–114, 2012.

B. Zhang, A. A. Tsiatis, E. B. Laber, and M. Davidian. Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika*, 100(3):681–694, 2013.

Q.-H. Zhang and Y.-Q. Ni. Improved most likely heteroscedastic Gaussian process regression via Bayesian residual moment estimator. *IEEE Transactions on Signal Processing*, 68: 3450–3460, 2020.

Y. Zhao, M. R. Kosorok, and D. Zeng. Reinforcement learning design for cancer clinical trials. *Statistics in Medicine*, 28(26):3294–3315, 2009.

Y. Zhao, D. Zeng, J. Rush, and M. R. Kosorok. Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107 (499):1106–1118, 2012.

Y.-Q. Zhao, D. Zeng, E. B. Laber, and M. R. Kosorok. New statistical learning methods for estimating optimal dynamic treatment regimes. *Journal of the American Statistical Association*, 110(510):583–598, 2015.

X. Zhou, N. Mayer-Hamblett, U. Khan, and M. R. Kosorok. Residual weighted learning for estimating individualized treatment rules. *Journal of the American Statistical Association*, 112(517):169–187, 2017.

Y. Zhu, R. A. Hubbard, J. Chubak, J. Roy, and N. Mitra. Core concepts in pharmacoepidemiology: Violations of the positivity assumption in the causal analysis of observational data: Consequences and statistical approaches. *Pharmacoepidemiology and Drug Safety*, 30(11):1471–1485, 2021.