Title

Essays in Bayesian Econometrics and Game Theory

Author

Gao, Jieyu

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Essays in Bayesian Econometrics and Game Theory

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Economics

by

Jieyu Gao

Dissertation Committee:
Associate Professor Ivan Jeliazkov, Chair
Professor Yingying Dong
Associate Professor Jiawei Chen
Professor Emeritus David Brownstone

2024

DEDICATION

To my parents, sister, and fiancé, whose love and support have fueled my journey.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am deeply grateful to my advisor, Ivan Jeliazkov, for his support, guidance, and dedication. Joining his mentorship during the third year of my Ph.D. was a pivotal moment, and his expertise in Bayesian econometrics has profoundly influenced my academic trajectory. Amidst the challenges of the pandemic, I am thankful for his patience and understanding, which sustained me during difficult times. I extend my thanks for his invaluable contributions to Chapter 1 of this dissertation, where his mentorship significantly enhanced the quality of the work and supported my academic growth.

I extend my appreciation to my committee members, Yingying Dong, Jiawei Chen, and David Brownstone. Professor Dong's causal inference class sparked my interest in researching causal treatment models, laying the groundwork for my academic pursuits. Professor Chen's industrial organization course provided essential knowledge, propelling me to initiate my first research project in the industrial organization context. Professor Brownstone's expertise in empirical analysis has been instrumental in shaping my understanding of real-world applications. Their constructive feedback and comments have consistently enhanced the quality of my research, and their encouragement motivated me to strive for excellence.

I am thankful to Michael Choi for his help and support during my first research project, which forms the basis of Chapter 3 of this dissertation. His expertise and guidance in game theory offered me insights and inspiration for embarking on this endeavor. I appreciate his contributions to my academic journey.

I would like to thank my colleagues and friends for their companionship and support. Their encouragement and willingness to engage in discussions about my research ideas have been invaluable to me. Beyond academia, their presence has been a source of comfort and motivation. I am truly blessed to have them by my side.

Lastly, my deepest thanks go to my parents, Fenglin and Yingmin, my sister, Jiejun, and my fiancé, James, for their unwavering love, support, and encouragement. Their presence and belief in my abilities have been a source of strength and motivation, particularly during challenging times. I am forever grateful for their love and support.

VITA

Jieyu Gao

EDUCATION

Doctor of Philosophy in Economics **2024**
University of California, Irvine *Irvine, CA*

Master of Arts in Economics **2019**
University of California, Irvine *Irvine, CA*

Bachelor of Science in Economics **2016**
Purdue University *West Lafayette, IN*

Bachelor of Science in Applied Statistics **2016**
Purdue University *West Lafayette, IN*

TEACHING EXPERIENCE

Teaching Assistant **2019–2023**
University of California, Irvine *Irvine, CA*

FIELDS OF STUDY

Econometrics, Industrial Organization

ABSTRACT OF THE DISSERTATION

Essays in Bayesian Econometrics and Game Theory

By

Jieyu Gao

Doctor of Philosophy in Economics

University of California, Irvine, 2024

Associate Professor Ivan Jeliazkov, Chair

This dissertation comprises three chapters that expand upon Bayesian econometrics and game theory. Chapter 1 integrates heteroskedasticity into various models including regression discontinuity design, the Rubin causal (Roy-type) model, propensity score matching, and inverse probability weighting within a Bayesian framework. The practical adequacy of our modeling techniques is assessed through formal Bayesian model comparisons. Simulation studies evaluate the impact of neglected heteroskedasticity and the efficacy of our proposed methods, while their practical applicability is examined through three applications. Specifically, we investigate the impact of academic probation on subsequent academic performance, the effect of Medigap on healthcare expenditures, and the influence of COVID-19 vaccination on mental well-being. These applications highlight the consequences of misspecification and underscore the importance of addressing risks associated with omitted heteroskedasticity. Chapter 2 introduces a Bayesian treatment model that incorporates self-selection and multiple outcomes. We discuss marginal likelihood estimation for formal model comparison and validate our approach through simulation results. We then apply the model to two datasets, examining the influence of insurance on healthcare utilization. In particular, we analyze the impact of Medigap policies on healthcare expenditure using 1987 National Medical Expenditure Survey (NMES) data and the effect of types of private insurance on healthcare utilization using 1996 Medical Expenditure Panel Survey (MEPS) data. Our analysis indi-

cates weak evidence supporting selection bias in both applications. In Chapter 3, we delve into the analysis of a price signaling game that integrates consumer learning. The chapter delineates the perfect Bayesian equilibrium for sellers and employs the undefeated equilibrium refinement to determine optimal choices. Furthermore, our analysis encompasses comparative statics, examining how variations in the probability of type revelation, initial customer review scores, and the quality of both high and low-quality seller products impact the expected return for sellers. Collectively, these chapters offer methodological advancements in Bayesian econometrics and game theory, and provide insights into real-world phenomena across various industries.

Chapter 1

The Impact of Heteroskedasticity in Observational Studies of Causal Effects

There is a large and rapidly growing causal inference literature, yet little is known about the impact of heteroskedasticity in popular causal settings. In observational studies where the presence of heteroskedasticity can not be ruled out with certainty, its effects on treatment assignment and response generation must be studied not only because they can be of interest in their own right, but also because omitted heteroskedasticity can interact with nonlinearities in each case and impact the bias and consistency properties of estimators which can not be corrected by standard error adjustments. Our approach is Bayesian and involves modeling whose practical adequacy is addressed through model comparisons. We extend the methodology underlying well-known settings such as the regression discontinuity designs, the Rubin causal (Roy-type) model, propensity score matching, and inverse probability weighting. Key features of that methodology include flexible modeling, the development of customized computationally efficient estimation algorithms, the ability to recover various

functions of the treatment parameters, and improved efficiency of estimation. Simulation studies demonstrate the effects of omitted heteroskedasticity and gauge the adequacy of our proposed modeling and estimation methods, while their practical applicability is studied in three applications. In particular, we examine the effect of academic probation on subsequent academic performance, the influence of Medigap on healthcare expenditures, and the impact of COVID-19 vaccination on mental well-being. These applications illustrate the consequences of misspecification and provide strong evidence that the dangers of omitted heteroskedasticity should not be ignored.

1.1 Introduction

The formulation of an identification framework through which the observed outcomes for the treated and untreated units can be compared plays a crucial role in observational studies of causal effects. This importance is underscored by challenges arising from non-random treatment assignments, unobserved confounders, or the inherent missingness of counterfactual outcomes at the unit level. A variety of parametric, semi-parametric, and nonparametric approaches have been proposed in the literature that have dealt with the effects of continuous, binary, and categorical treatments in non-experimental settings in both classical and Bayesian contexts. Regression discontinuity, Rubin causal (a.k.a. Roy-type) models, and matching estimators, among others, have been proposed and implemented in applications. Classical approaches have tended to center around estimators that are robust to potential misspecifications of the data-generating process (DGP), which is often not explicitly stated, and inference is asymptotic. The Bayesian methods adopted in this paper, consider the DGP directly and explicitly, leading to finite-sample inferences; possible misspecification is handled by allowing for flexible modeling and conducting formal model comparisons and specification searches.

The Bayesian literature on treatment effect estimation encompasses a diverse array of models. One strand of this literature has focused on estimating treatment effects in continuous and discrete (binary, ordinal, and count) instrumental variable models (Koop and Tobias, 2004; Mintz, Currim and Jeliazkov, 2013; Li and Tobias, 2014; Vossmeyer, 2014*a*), settings with sequential outcomes Munkin (2011), as well as models with nonparametric endogeneity (Kline and Tobias, 2008; Chib, Greenberg and Jeliazkov, 2009). Chan and Tobias (2015) propose methods for analyzing models with imperfect instruments which are not necessarily excluded from the primary regression equation of interest. Moreover, models embodying both endogeneity and sample selection have been presented in Chib, Greenberg and Jeliazkov (2009), Vossmeyer (2014*b*), and Vossmeyer (2016) (see also van Hasselt, 2014).

Work has also been done within the broader potential outcomes framework for causal analysis offered by the Rubin causal model Rubin (1974, 1977, 1978, 2004, 2005), often referred to as a Roy-type model following the work of Roy (1951). Bayesian research in cross-sectional settings encompasses both continuous and discrete outcome variables, while considering binary or categorical treatments (see, e.g., Munkin and Trivedi, 1999; Chib and Hamilton, 2000; Munkin, 2003; Munkin and Trivedi, 2003; Deb, Munkin and Trivedi, 2006; Li and Tobias, 2008, 2011). Extensions to longitudinal settings have been addressed in Chib and Hamilton (2002) and Jacobi, Wagner and Frühwirth-Schnatter (2016). Estimation has been approached both by explicitly simulating the counterfactuals from their joint distribution with the observed outcomes (Li, Poirier and Tobias, 2004) and by solely involving the observed outcomes Chib (2007). Heckman, Lopes and Piatek (2014) proposed a method to model the joint distribution of potential outcomes by introducing a latent factor into the analysis.

Following the seminal work of Rosenbaum and Rubin (1983) and Rosenbaum (1987), substantial attention has also been directed toward the development and application of methods centered around the conditional probability of receiving treatment, known as the propen-

sity score (Dehejia and Wahba, 1999; Imai and van Dyk, 2004; Brand and Halaby, 2006; Zhao, 2008; Caliendo and Kopeinig, 2008; An, 2010; Zhao, van Dyk and Imai, 2020; Chaudhuri and Howley, 2022; Chesnaye et al., 2022; Duan et al., 2023). A recent review of these methodologies is offered in Rosenbaum and Rubin (2022). Propensity score matching (PSM) and inverse probability of treatment weighting (IPTW) estimators have found application across a broad range of settings. The framework is elegant and theoretically powerful; yet, in practice, results from its implementation have often been mixed. Because propensity score misspecification can compromise the efficacy of PSM and IPTW methods, model uncertainty is a pivotal challenge that warrants formal consideration in empirical practice.

There has also been a recent surge in interest in causal analysis within the regression discontinuity design (RDD) framework. The RDD approach, introduced in Thistlethwaite and Campbell (1960) aims to address causal inference in a quasi-experimental setting where treatment assignments are based on an ancillary variable crossing a known cutoff, with a discontinuous treatment assignment rule at this cutoff point. Sharp RDD has a strict rule for treatment assignment based on the cutoff, whereas in fuzzy RDD the assignment is probabilistic. There are many different applications and extensions in the literature (see Hahn, Todd and Van der Klaauw, 2001; Calonico, Cattaneo and Titiunik, 2014*a*,*b*; Cattaneo, Frandsen and Titiunik, 2015; Dong, 2015; Dong and Lewbel, 2015; Fletcher and Tokmouline, 2018; Dong, 2019; Wright, 2020; Dong, Lee and Gou, 2023, among others), yet Bayesian analysis has been relatively recent (Chib and Jacobi, 2016; Branson et al., 2019; Geneletti et al., 2019; Chib, Greenberg and Simoni, 2023). RDD methods continue to evolve rapidly, offering new perspectives and analytical tools in the study of causal relationships.

Despite the large and rapidly expanding body of causal methodology, the ramifications of heteroskedasticity in many popular treatment models remain poorly understood. One recent exception is the work of Ferman and Pinto (2019) who show that the presence of heteroskedasticity can severely impede the performance of standard methods, even in straightforward

linear specifications such as difference-in-differences, especially when confronted with small data sets. In non-linear contexts, including those mentioned earlier, the detrimental effects of heteroskedasticity in treatment assignment and response generation are expected to be amplified by any non-linearity and affect not only the efficiency, but also the bias and consistency properties of traditional estimators. This challenge underscores the need for a deeper study of the effects of heteroskedasticity, the development of new methodologies, and their careful implementation in applications.

In this chapter, we pursue these objectives by integrating heteroskedasticity into models within RDD, the Rubin causal (Roy-type) model, and the PSM and IPTW estimation frameworks. In each setting, we present customized Markov chain Monte Carlo (MCMC) simulation algorithms that are used for the estimation of the model parameters, including the treatment effects, as well as for estimating marginal likelihoods for the purpose of model comparisons. Marginal likelihood estimation is approached by calling upon existing techniques, when those are available, and by developing novel computationally efficient approaches when needed, e.g., in multi-block samplers that would otherwise be computationally costly. Furthermore, in each case we conduct targeted simulation studies in order to illustrate the effects of heteroskedasticity and demonstrate the performance of the estimation and model comparison algorithms. Finally, we employ the techniques in several applications to gauge their practical relevance. In particular, we study the effect of academic probation on subsequent academic performance, the influence of Medigap on healthcare expenditures, and the impact of COVID-19 vaccination on mental well-being in the UK.

The rest of the chapter is organized as follows. In Section 1.2, we build upon and extend existing Bayesian methods for the regression discontinuity designs in which we couple nonparametric modeling of the running variable with a model for heteroskedasticity. In Section 1.3, we develop modeling and estimation techniques for the analysis of a heteroskedastic Rubin causal model, while Section 1.4 focuses on PSM and IPTW estimation under heteroskedas-

ticity. Each section presents MCMC estimation algorithms, assesses the performance and impact of heteroskedasticity through simulations, and applies them in practical scenarios. Section 1.5 offers concluding remarks.

1.2 Regression Discontinuity Design

This section considers heteroskedastic variants of the RDD framework and provides the necessary estimation and model comparison techniques. Key parts of the methodology, e.g., the sampling of heteroskedasticity parameters and the approach for estimating the marginal likelihood, will continue to play a pivotal role in subsequent sections. Nonparametric functions are employed to safeguard against misspecification while improving efficiency by capitalizing on the entirety of available data (cf. Branson et al., 2019; Chib, Greenberg and Simoni, 2023), as opposed to limiting the analysis to only a subset of observations near the cutoff point. Techniques are developed for both continuous and binary outcomes.

In the sharp RD setting, the treatment $T_i \in \{0, 1\}$ for unit $i = 1, \ldots, n$ is determined by a running variable w_i and a known cutoff w^* such that $T_i = \mathbb{1}\{w_i \geq w^*\}$. The potential outcomes of unit i are continuous and are denoted by y_{i0} and y_{i1} for $T_i = 0$ and $T_i = 1$, respectively, and are assumed to be generated by the additive specification

$$y_{ij} = g_j(w_i) + x_i'\beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N\left(0, \sigma_{ij}^2\right), \quad \ln(\sigma_{ij}^2) = z_i'\gamma_j, \quad \text{for} \quad j \in \{0, 1\}, \tag{1.1}$$

where we observe $y_i = (1 - T_i)\, y_{i0} + T_i y_{i1}$. Heterogeneity is allowed to depend on some collection of variables z_i that could include, but is not necessarily limited to, the variables in $\{x_i, w_i, T_i\}$ and their interactions. Extensions beyond normality can be pursued through scale mixtures of normals (Andrews and Mallows, 1974) or Dirichlet process mixtures (Ferguson, 1973; Antoniak, 1974).

The model specified in equation (1.1) is one of structural change between the treated and untreated samples; thus, intuitively, estimation can simply be performed separately within each sub-sample. However, estimation has typically been performed under the assumption $\beta_0 = \beta_1$, which emphasizes the discontinuity in the running variable and, if confirmed by the data, makes inference more precise (the assumption will be examined in the application in Section 1.2.3). In the literature, the RD average treatment effect (RD ATE) is defined as

$$\begin{aligned} \Delta_{SRD} &\equiv \lim_{w \downarrow w^{*+}} E\left(Y_1 | w, x_i\right) - \lim_{w \uparrow w^{*-}} E\left(Y_0 | w, x_i\right) \\ &= \lim_{w \downarrow w^{*+}} E\left(g_1(w) + x_i'\beta_1\right) - \lim_{w \uparrow w^{*-}} E\left(g_0(w) + x_i'\beta_0\right) , \end{aligned} \tag{1.2}$$

which, in the special case when $\beta_1 = \beta_0$, leads to

$$\Delta_{SRD} = \lim_{w \downarrow w^{*+}} g_1(w) - \lim_{w \uparrow w^{*-}} g_0(w) . \tag{1.3}$$

The function $g_j(\cdot)$ plays a crucial role in this setting and is modeled nonparametrically with only local penalties for smoothness in order to mitigate the potential for undue influence of values of w far from w^* on the estimated values of $g_j(\cdot)$ close to w^* (Gelman and Imbens, 2019). Flexible functional modeling can be implemented through a variety of approaches including B-splines, regression splines, natural splines, truncated power series, or wavelets, among others (for a review, see, e.g., Ruppert, Wand and Carroll, 2003; Ahamada and Flachaire, 2010). Here we focus on flexible modeling through Gaussian random fields Poirier (1973); Shiller (1984); Williams (1998); Fahrmeir and Lang (2001); Koop and Poirier (2004); Koop, Poirier and Tobias (2005); Rue and Held (2005); Chib and Jeliazkov (2006); Chan and Jeliazkov (2009*a*); Jeliazkov (2013); Branson et al. (2019) because it allows for smoothing at every observed value of the running variable (instead of a coarser set of knots), while retaining key desirable computational properties. Extensions of the model in (1.1) to an additive nonparametric mean structure for the covariates x_i can be implemented as in Koop, Poirier and Tobias (2005), Panagiotelis and Smith (2008), or Jeliazkov (2013). Chib and

Greenberg (2010) combine nonparametric mean and error distribution modeling.

To facilitate the derivations, arrange the data $\{T_i, y_i, w_i, x_i\}_{i=1}^n$ so that observations $1 \leq i \leq n_0$ belong to the control group, while the n_1 units $i \in [n_0+1, n]$ pertain to the treated group. Let $y_0 \equiv (y_1, \ldots, y_{n_0})'$, $w_0 \equiv (w_1, \ldots, w_{n_0})'$, $y_1 \equiv (y_{n_0+1}, \ldots, y_n)'$, and $w_1 \equiv (w_{n_0+1}, \ldots, w_n)'$. To define $g_j(\cdot)$, denote the *unique ordered values* of w_j as v_j, i.e., $v_0 = (w_{0,\min}, \ldots, w^*)' = (v_{01}, \ldots, v_{0m_0})'$, $v_1 = (w^*, \ldots, w_{1,\max})' \equiv (v_{11}, \ldots, v_{1m_1})'$, with m_j being the number of elements in v_j. Note that the cutoff w^* appears in both v_0 and v_1 to enable computation of $g_0(w^*)$ and $g_1(w^*)$.

The function evaluations $g_j = \left(g\left(v_{j1}\right), \ldots, g\left(v_{jm_j}\right)\right)' \equiv \left(g_{j1}, \ldots, g_{jm_j}\right)'$, $j \in \{0,1\}$, are modeled as a second order Markov process

$$g_{jl} = \left(1 + \frac{h_{jl}}{h_{jl-1}}\right) g_{jl-1} - \frac{h_{jl}}{h_{jl-1}} g_{jl-2} + \mu_{jl},$$

where $h_{jl} \equiv v_{jl} - v_{jl-1}$, $\mu_{jl} \sim N\left(0, \tau_j^2 h_{jl}\right)$ and the process is initialized at

$$\begin{pmatrix} g_{j1} \\ g_{j2} \end{pmatrix} | \tau_j^2 \sim N\left(\begin{pmatrix} g_{j10} \\ g_{j20} \end{pmatrix}, \tau_j^2 G_{j0}\right),$$

where G_{j0} is a 2×2 symmetric positive definite matrix and τ^2 is a smoothness parameter whose magnitude determines the penalty to deviations from a locally linear relationship. Letting

$$H_j = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ \frac{h_{j3}}{h_{j2}} & -(1+\frac{h_{j3}}{h_{j2}}) & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & \frac{h_{jm}}{h_{jm-1}} & -(1+\frac{h_{jm}}{h_{jm-1}}) & 1 \end{pmatrix}, \quad \Sigma_j = \begin{pmatrix} G_{j0} & & & \\ & h_{j3} & & \\ & & \ddots & \\ & & & h_{jm} \end{pmatrix},$$

we obtain the joint distribution $g_j|\tau_j^2 \sim N\left(g_{j0}, \tau_j^2 K_j^{-1}\right)$, where $g_{j0} = H_j^{-1}\left(g_{j10}, g_{j20}, 0, \ldots, 0\right)'$ and $K_j = H_j' \Sigma_j^{-1} H_j$. Of key importance is the fact that K is a banded and operations involving it are of order $\mathcal{O}(n)$ (Fahrmeir and Lang, 2001; Chib and Jeliazkov, 2006). With this definition of g_j, $j \in \{0, 1\}$, stacking the model in (1.1), we can write

$$y_j = Q_j g_j + X_j \beta_j + \varepsilon_j, \quad \varepsilon_j \sim N\left(0, \Omega_j\right), \quad \Omega_j = \text{diag}\left(\{\sigma_{ij}^2\}_{i=1}^{n_j}\right), \quad \ln(\sigma_{ij}^2) = z_i' \gamma_j,$$

where Q_j is a $n \times m$ incidence matrices with entries $Q_j\left(i, k\right) = 1$ if $w_{ji} = v_{jk}$, and 0 otherwise. The model is completed by the prior distributions

$$g_j|\tau_j^2 \sim N\left(g_{j0}, \tau_j^2 K_j^{-1}\right), \quad \tau_j^2 \sim IG\left(\kappa_{j0}/2, d_{j0}/2\right), \quad \beta_j \sim N\left(b_{j0}, B_{j0}\right), \quad \gamma_j \sim N\left(\gamma_{j0}, \Gamma_{j0}\right), \tag{1.4}$$

which, combined with the sampling density

$$f\left(y|g_0, g_1, \tau_0^2, \tau_1^2, \beta_0, \beta_1, \gamma_0, \gamma_1\right) = f_N\left(y_0|Q_0 g_0 + X_0 \beta_0, \Omega_0\right) f_N\left(y_1|Q_1 g_1 + X_1 \beta_1, \Omega_1\right),$$

lead to a joint posterior distribution that can be sampled by the MCMC algorithm presented in Algorithm 1. Precision-based algorithms are used for sampling g_j, whereas efficient simulation of γ_j is obtained by a Metropolis-Hastings (MH) step with proposal density based on iteratively reweighted least squares (Chan et al., 2006; Gu et al., 2009; Gamerman, 1997; Nott and Leonte, 2004). This approach is considerably faster than conventional tailoring by optimization at every MCMC step and is obtained by constructing a Student's t proposal density $q\left(\gamma_j|\hat{\gamma}_j, V_j\right) = f_{T_\nu}\left(\gamma_j|\hat{\gamma}_j, V_j\right)$ with ν degrees of freedom, where $e_{ij} = (y_{ij} - g_j(w_i) - x_i' \beta_j)$ and

$$\begin{aligned} &\eta_{ij} = z_i' \gamma_j + (e_{ij}^2 - \sigma_{ij}^2)/\sigma_{ij}^2, \quad \eta_j = \left(\eta_{1j}, \ldots, \eta_{n_j}\right)' \\ &V_j = \left(\Gamma_{0j}^{-1} + \frac{1}{2} Z_j' Z_j\right)^{-1}, \quad \hat{\gamma}_j = V_j \left(\Gamma_{j0}^{-1} \gamma_{j0} + \frac{1}{2} Z_j' \eta_j\right). \end{aligned} \tag{1.5}$$

The treatment effects in equations (1.2) and (1.3) can be computed by averaging over the

output of the MCMC sampler and, if needed, the empirical distribution of the covariates $\{x_i, w_i, z_i\}$ in the neighborhood of w^*. The homoskedastic model results in the special case when $z_i = 1$. The overall approach is also easily adaptable to binary outcomes y_i as discussed next.

Algorithm 1 (Semi-parametric Sharp RDD)

---

(1) Sample $\left[g_j|y_j, \beta_j, \tau_j^2, \gamma_j\right] \sim N\left(\hat{g}_j, \hat{G}_j\right)$, where $\hat{g}_j = \hat{G}_j\left(\frac{1}{\tau_j^2}K_j g_{j0} + Q_j'\Omega_j^{-1}\left(y_j - X_j\beta_j\right)\right)$ and $\hat{G}_j = \left(\frac{K_j}{\tau_j^2} + Q_j'\Omega_j^{-1}Q_j\right)^{-1}$, $j = 0, 1$.

(2) Sample $\left[\beta_j|y_j, g_j, \gamma_j\right] \sim N\left(\hat{\beta}_j, \hat{B}_j\right)$, where $\hat{\beta}_j = \hat{B}_j\left(B_{j0}^{-1}b_{j0} + X_j'\Omega_j^{-1}\left(y_j - Q_j g_j\right)\right)$ and $\hat{B}_j = \left(B_{j0}^{-1} + X_j'\Omega_j^{-1}X_j\right)^{-1}$, $j = 0, 1$. If $\beta_0 = \beta_1$, sample $\left[\beta|y, g_0, g_1, \gamma_0, \gamma_1\right] \sim N\left(\hat{\beta}, \hat{B}\right)$, where $\hat{\beta} = \hat{B}\left(B_0^{-1}b_0 + X'\Omega^{-1}\left(y - Qg\right)\right)$, $\hat{B} = \left(B_0^{-1} + X'\Omega^{-1}X\right)^{-1}$, $y = (y_0', y_1')'$, $g = (g_0', g_1')'$, $X = \begin{pmatrix} X_0 \\ X_1 \end{pmatrix}$, and $\Omega = \begin{pmatrix} \Omega_0 & 0 \\ 0 & \Omega_1 \end{pmatrix}$.

(3) Sample $\left[\tau_j^2|g_j\right] \sim IG\left(\frac{\kappa_{j0}+m_j}{2}, \frac{d_{j0}+(g_j-g_{j0})'K(g_j-g_{j0})}{2}\right)$, $j = 0, 1$.

(4) Sample $\left[\gamma_j|y_j, g_j, \beta_j\right]$, $j = 0, 1$, using an MH step by drawing a proposed $\gamma_j^\dagger \sim q\left(\gamma_j|\hat{\gamma}_j, V_j\right)$, where $\hat{\gamma}_j$ and V_j are computed in (1.5) using the current value of γ_j. Also use $\gamma_j^\dagger$ in equation (1.5) to produce $\hat{\gamma}_j^\dagger$. Accept the proposed $\gamma_j^\dagger$ with probability

$$\alpha(\gamma_j, \gamma_j^\dagger|y_j, g_j, \beta_j) = \min\left\{1, \frac{f(y_j|g_j, \beta_j, \gamma_j^\dagger)\pi(\gamma_j^\dagger|\gamma_{j0}, \Gamma_{j0})}{f(y_j|g_j, \beta_j, \gamma_j)\pi(\gamma_j|\gamma_{j0}, \Gamma_{j0})}\frac{q(\gamma_j|\hat{\gamma}_j^\dagger, V_j)}{q(\gamma_j^\dagger|\hat{\gamma}_j, V_j)}\right\},$$

otherwise repeat the current value γ_j in the next MCMC iteration.

To handle binary outcomes $y_i \in \{0, 1\}$, we use data augmentation (Albert and Chib, 1993) and introduce the latent variables y_i^* such that $y_i = \mathbb{1}\{y^* \geq 0\}$ and

$$y_{ij}^* = g_j(w_i) + x_i'\beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N\left(0, \sigma_{ij}^2\right), \quad \ln(\sigma_{ij}^2) = z_i'\gamma_j, \quad \text{for} \quad j \in \{0, 1\}. \tag{1.6}$$

For identification purposes, the vector z_i, which plays a role in determining the variance, does not include a constant term (Gu et al., 2009). The variance in the homoskedastic version of

the model is fixed at 1 and is not estimated.

The complete data likelihood can be expressed as

$$
\begin{aligned}
& f\left(y, y^* | g_0, g_1, \beta_0, \beta_1, \gamma_0, \gamma_1, \tau_0^2, \tau_1^2\right) \\
= & \prod_{i:T_i=0} \Bigg(\left(f_N\left(y_i^* | g_0(w_i) + x_i'\beta_0, \sigma_{i0}^2\right) \mathbb{1}\{y_i^* \geq 0\}\right)^{y_i} \\
& \qquad \times \left(f_N\left(y_i^* | g_0(w_i) + x_i'\beta_0, \sigma_{i0}^2\right) \mathbb{1}\{y_i^* < 0\}\right)^{1-y_i} \Bigg) \\
\times & \prod_{i:T_i=1} \Bigg(\left(f_N\left(y_i^* | g_1(w_i) + x_i'\beta_1, \sigma_{i1}^2\right) \mathbb{1}\{y_i^* \geq 0\}\right)^{y_i} \\
& \qquad \times \left(f_N\left(y_i^* | g_1(w_i) + x_i'\beta_1, \sigma_{i1}^2\right) \mathbb{1}\{y_i^* < 0\}\right)^{1-y_i} \Bigg),
\end{aligned}
$$

which, combined with the priors in (1.4) produces the joint posterior that is sampled in Algorithm 2. Note that Algorithms 1 and 2 are closely related, but the latter makes use of the suitably generated latent $\{y_i^*\}$ instead of the observed $\{y_i\}$; both algorithms also trivially handle homoskedasticity. Finally, for computing RD ATE in equations (1.2)-(1.3) in the case of a binary outcome, it is helpful to recognize that $E\left(Y_j | w, x_i, z_i\right)$ is given by $\Phi\left((g_j(w) + x_i'\beta_j)/\sqrt{\exp\left(z_i'\gamma_j\right)}\right)$, which is averaged over the MCMC draws and covariates $\{x_i, w_i, z_i\}$ for w_i in a neighborhood of w^*. Accounting for both covariate variability and estimation uncertainty is critical in this context because of the inherent nonlinearity of the estimand (see, e.g., Verlinda, 2006; Jeliazkov and Vossmeyer, 2018).

The heteroskedastic framework can also be extended to the fuzzy RD context. This extension involves the introduction of an unobserved discrete confounding variable s which categorizes individuals into three distinct types: (1) compliers (denoted as $\mathcal{C}$) who take the treatment when assigned, i.e., $T_i = \mathbb{1}\{w_i \geq w^*\}$, (2) never-takers (denoted as $\mathcal{N}$) who consistently abstain from treatment, and (3) always-takers (denoted as $\mathcal{A}$) who always opt for the treatment. The presence of defiers, individuals whose behavior is opposite that of compliers, is ruled out. Chib, Greenberg and Simoni (2023) show that the Bayesian fuzzy RD

Algorithm 2 (Semi-parametric Sharp RDD with Binary Outcomes)

---

(1) Sample $\left[g_j|y_j^*,\beta_j,\tau_j^2,\gamma_j\right] \sim N\left(\hat{g}_j,\hat{G}_j\right)$, where $\hat{G}_j = \left(\frac{K_j}{\tau_j^2}+Q_j'\Omega_j^{-1}Q_j\right)^{-1}$ and $\hat{g}_j = \hat{G}_j\left(\frac{1}{\tau_j^2}K_jg_{j0}+Q_j'\Omega_j^{-1}\left(y_j^*-X_j\beta_j\right)\right)$, $j=0,1$.

(2) Sample $\left[\beta_j|y_j^*,g_j,\gamma_j\right] \sim N\left(\hat{\beta}_j,\hat{B}_j\right)$, where $\hat{\beta}_j = \hat{B}_j\left(B_{j0}^{-1}b_{j0}+X_j'\Omega_j^{-1}\left(y_j^*-Q_jg_j\right)\right)$ and $\hat{B}_j = \left(B_{j0}^{-1}+X_j'\Omega_j^{-1}X_j\right)^{-1}$, $j=0,1$. If $\beta_0=\beta_1$, sample $[\beta|y^*,g_0,g_1,\gamma_0,\gamma_1] \sim N\left(\hat{\beta},\hat{B}\right)$, where $\hat{\beta} = \hat{B}\left(B_0^{-1}b_0+X'\Omega^{-1}\left(y^*-Qg\right)\right)$, $\hat{B} = \left(B_0^{-1}+X'\Omega^{-1}X\right)^{-1}$, $y^* = \left(y_0^{*\prime},y_1^{*\prime}\right)'$, $g=(g_0',g_1')'$, $X = \begin{pmatrix}X_0\\X_1\end{pmatrix}$, and $\Omega = \begin{pmatrix}\Omega_0 & 0\\0 & \Omega_1\end{pmatrix}$.

(3) Sample $\left[\tau_j^2|g_j\right] \sim IG\left(\frac{\kappa_{j0}+m_j}{2},\frac{d_{j0}+(g_j-g_{j0})'K(g_j-g_{j0})}{2}\right)$, $j=0,1$.

(4) Sample $\left[\gamma_j|y_j^*,g_j,\beta_j\right]$, $j=0,1$, using an MH step by drawing a proposed $\gamma_j^\dagger \sim q\left(\gamma_j|\hat{\gamma}_j,V_j\right)$, where $e_i = y_i^*-g_j(w_i)-x_i'\beta_j$ and $\hat{\gamma}_j$ and V_j are computed in (1.5) using the current value of γ_j and y_j^*. Also use $\gamma_j^\dagger$ in equation (1.5) to produce $\hat{\gamma}_j^\dagger$. Accept the proposed $\gamma_j^\dagger$ with probability

$$\alpha = \min\left\{1,\frac{f(y_j^*|g_j,\beta_j,\gamma_j^\dagger)\pi(\gamma_j^\dagger|\gamma_{j0},\Gamma_{j0})q(\gamma_j|\hat{\gamma}_j^\dagger,V_j)}{f(y_j^*|g_j,\beta_j,\gamma_j)\pi(\gamma_j|\gamma_{j0},\Gamma_{j0})q(\gamma_j^\dagger|\hat{\gamma}_j,V_j)}\right\}.$$

otherwise repeat the current value γ_j in the next MCMC iteration.

(5) Sample $\left[y_{ij}^*|y_{ij},g_j,\beta_j,\gamma_j\right] \sim TN_{\mathcal{B}_i}\left(g_j\left(\omega_i\right)+x_i'\beta_j,\exp\left(z_i'\gamma_j\right)\right)$, $i=1,\ldots,n$, $j=0,1$, where $\mathcal{B}_i=(-\infty,0]$ if $y_i=0$, and $\mathcal{B}_i=(0,\infty)$ if $y_i=1$.

---

model can then be viewed as a mixture model over the latent categories which form a basis for category-specific modeling of functions, parameters, and heteroskedasticity relationships. The heteroskedastic model and related estimation algorithm details are presented in A.2, but our study reveals an important caveat that must be kept in mind when working with this class of models. In particular, it is well known that the likelihood function of mixture models can be ill-behaved even in simple cases, and can exhibit multimodality and singularities, and estimation algorithms can exhibit misclassification and label switching when the clusters are not well-separated (Celeux, 1998; Frühwirth-Schnatter, 2006, 2011). While some regulariza-

tion in the Bayesian context can be achieved through the prior, and larger samples can be beneficial, much care is needed in practical applications because the severity of the aforementioned problems is rarely known a priori. In fact, our simulations reveal cases in which the estimates recover the true data-generating process very well, but also point to instances where misclassification and label switching can lead to erroneous parameter and treatment effect estimates. The effect of heteroskedasticity can also be ambiguous, since lower variances can serve to increase cluster separation and improve identification, and vice versa. While the methods proposed in Frühwirth-Schnatter (2001) and Frühwirth-Schnatter and Kaufmann (2008) could be employed in this context, their applicability in high-dimensional contexts with multiple non-parametric functions has not been studied. Therefore, we urge practitioners to thoroughly review their findings before drawing definitive conclusions in fuzzy RD designs.

1.2.1 Bayesian Model Comparison and Marginal Likelihood Estimation

In the presence of multiple competing models, each reflecting alternative hypotheses about the data y, Bayesian model comparison provides a systematic approach for addressing model uncertainty. Specifically, by Bayes' formula, the posterior probability of model $\mathcal{M}_s$ can be expressed as

$$P(\mathcal{M}_s|y) \propto P(\mathcal{M}_s)m(y|\mathcal{M}_s),$$

where $P(\mathcal{M}_s)$ represents the prior probability of model $\mathcal{M}_s$ and $m(y|\mathcal{M}_s)$ denotes the marginal likelihood $m(y|\mathcal{M}_s) = \int f(y|\theta_s, \mathcal{M}_s)\pi(\theta_s|\mathcal{M}_s)\, d\theta_s$, where $f(y|\theta_s, \mathcal{M}_s)$ is the likelihood function and $\pi(\theta_s|\mathcal{M}_s)$ is the prior density on the parameters θ_s in model $\mathcal{M}_s$. An important approach for estimating the marginal likelihood was introduced by (Chib, 1995)

based on the recognition that

$$m(y|\mathcal{M}_s) = \frac{f(y|\theta_s, \mathcal{M}_s)\, \pi(\theta_s|\mathcal{M}_s)}{\pi(\theta_s|y, \mathcal{M}_s)}, \tag{1.7}$$

which holds for any θ_s in the parameter space. The terms in the numerator of equation (1.7) are often available directly, so the primary challenge is in estimating the posterior density $\pi(\theta_s|y, \mathcal{M}_s)$ in the denominator of equation (1.7). In practice, the right-hand side of equation (1.7) is evaluated at some appropriate point θ_s^*, typically taken to be the posterior mean or mode.

Marginal likelihoods, and their ratios known as Bayes factors (Kass and Raftery, 1995), serve as a cornerstone for implementing Bayesian model comparison. The approach exhibits a number of desirable properties. For instance, it provides finite-sample model probabilities that can be used for model averaging or model choice, and does not demand that competing models be nested, enhancing its applicability across diverse model structures. In addition, it exhibits appealing asymptotic behavior, giving rise to the well-known information criterion proposed by Schwarz (1978). An often underappreciated aspect of marginal likelihoods is that they provide a measure of sequential out-of-sample predictive fit, which can be seen by writing

$$\begin{aligned} m(y|\mathcal{M}_s) &= \prod_{i=1}^{n} m(y_i|\{y_j\}_{j<i}, \mathcal{M}_s) \\ &= \prod_{i=1}^{n} \int f(y_i|\{y_j\}_{j<i}, \theta_s, \mathcal{M}_s)\pi(\theta_s|\{y_j\}_{j<i}, \mathcal{M}_s)\, d\theta_s. \end{aligned}$$

Thus, model adequacy, as indicated by the marginal likelihood, corresponds to cumulative out-of-sample predictive performance. This assessment involves evaluating the fit of y_i based on the posterior density, utilizing data $\{y_j\}_{j<i}$ up to the ith data point. Unlike in-sample measures conditioned on the entire dataset y or split-sample comparisons sensitive to sample selection, the marginal likelihood remains unaffected by permutations in the order of data.

Thus, model adequacy, as captured by the marginal likelihood, corresponds to its cumulative out-of-sample predictive record where the fit of y_i is measured with respect to the posterior density using data $\{y_j\}_{j<i}$ up to the ith data point. This is in sharp contrast to in-sample measures of fit that condition on the entire data set y, or split-sample comparisons in which the outcome may depend on the choice of estimation and comparison samples. In contrast, the marginal likelihood is invariant to permutations in the indices of the data, so that the same $m(y|\mathcal{M}_s)$ will be obtained if the data are arbitrarily rearranged.

To simplify the notation in the remainder of our discussion, we suppress the model indicator $\mathcal{M}_s$ and focus on the case where the parameter vector consists of several blocks $\theta = (\theta_1', \ldots, \theta_B')'$. To handle this case, the posterior density in the denominator of (1.7), evaluated at the point θ^*, can be decomposed as

$$\pi(\theta^*|y) = \pi(\theta_1^*|y)\, \pi(\theta_2^*|y, \theta_1^*) \cdots \pi(\theta_B^*|y, \theta_1^*, \ldots, \theta_{B-1}^*),$$

where individual components $\pi(\theta_b^*|y, \{\theta_s^*\}_{(s<b)})$ on the right-hand side can be evaluated as

$$\pi(\theta_b^*|y, \{\theta_s^*\}_{(s<b)}) = E\left\{\pi\left(\theta_b^*|y, \{\theta_s^*\}_{(s<b)}, \{\theta_s\}_{(s>b)}\right)\right\} \tag{1.8}$$

when the full-conditional density $\pi(\theta_b^*|y, \{\theta_s^*\}_{(s<b)}, \{\theta_s\}_{(s>b)})$ is known (Chib, 1995), or as

$$\pi(\theta_b^*|y, \{\theta_s^*\}_{(s<b)}) = \frac{E\left\{\alpha\left(\theta_b, \theta_b^*|y, \{\theta_s^*\}_{(s<b)}, \{\theta_s\}_{(s>b)}\right)\, q\left(\theta_b, \theta_b^*|y, \{\theta_s^*\}_{(s<b)}, \{\theta_s\}_{(s>b)}\right)\right\}}{E\left\{\alpha\left(\theta_b^*, \theta_b|y, \{\theta_s^*\}_{(s<b)}, \{\theta_s\}_{(s>b)}\right)\right\}} \tag{1.9}$$

when the full-conditional density is non-standard and sampling requires the MH algorithm (Chib and Jeliazkov, 2001). The expectation in equation (1.8) is with respect to the distribution $\pi\left(\{\theta_s\}_{(s>b)}|y, \{\theta_s^*\}_{(s<b)}\right)$, whereas the expectations in the numerator and denominator of equation (1.9) are evaluated using the distributions $\pi\left(\{\theta_s\}_{(s\geq b)}|y, \{\theta_s^*\}_{(s<b)}\right)$ and $q\left(\theta_b^*, \theta_b|y, \{\theta_s^*\}_{(s<b)}, \{\theta_s\}_{(s>b)}\right)\pi\left(\{\theta_s\}_{(s>b)}|y, \{\theta_s^*\}_{(s<b)}\right)$, respectively. Estimation of the

marginal likelihood could, therefore, become computationally intensive as it requires additional simulation of $\{\theta_s\}_{(s\geq b)}$ in reduced runs where $\{\theta_s^*\}_{(s<b)}$ are held fixed.

To deal with this problem and improve computational efficiency, we group all parameter blocks that are sampled from known densities into the set $\psi = \{\psi_1, \ldots, \psi_R\}$, whereas latent data and parameters that are sampled from non-standard distributions are denoted by ξ. We propose an estimation of the joint ordinate of the blocks in ψ based on the invariance condition of Markov chains (Ritter and Tanner, 1992; Jeliazkov and Lee, 2010) as

$$\pi\left(\psi^*|y\right) = E\left\{K\left(\psi, \psi^*|y, \xi\right)\right\}, \tag{1.10}$$

where $K(\cdot)$ represents the Gibbs transition kernel

$$K\left(\psi, \psi^*|y\right) = \prod_{r=1}^{R} \pi\left(\psi_r^*|y, \{\psi_s^*\}_{(s<r)}, \{\psi_s\}_{(s>r)}, \xi\right),$$

with draws of $(\psi, \xi) \sim \pi(\psi, \xi|y)$ obtained in the main MCMC run. This method avoids the computation of the ordinates for $\{\psi_1, \ldots, \psi_R\}$ individually, which obviates the need for reduced runs for those densities. In the sharp RDD case, $\psi = \{\beta_0, \beta_1, g_0, g_1, \tau_0^2, \tau_1^2\}$, while $\xi = \gamma$ with $\gamma = \{\gamma_0, \gamma_1\}$ for continuous outcomes and $\xi = \{\gamma, \{y_{i0}^*\}, \{y_{i1}^*\}\}$ when outcomes are binary. The marginal likelihood in the sharp RDD case with continuous outcomes can be succinctly expressed as

$$\hat{m}\left(y\right) = \frac{f\left(y|\psi^*, \gamma^*\right)\pi\left(\psi^*, \gamma^*\right)}{\pi\left(\psi^*|y\right)\pi\left(\gamma^*|y, \psi^*\right)},$$

where $\pi(\psi^*|y)$ can be estimated using equation (1.10) with draws from the main MCMC run, and $\pi(\gamma^*|y, \psi^*)$ is obtained by equation (1.9), which requires a single reduced run. The binary data case is handled analogously, but includes integration over the latent $\{y_{ij}^*\}$ (Chib and Jeliazkov, 2001).

In addition to the RDD case, the methodology introduced here will be applied to models addressed later in this paper; these implementations will involve either straightforward simplification or direct adaptation of the techniques presented in this section.

1.2.2 Simulation Study

In this section, we conduct targeted simulations to assess the influence of ignored heteroskedasticity, evaluate the performance of the MCMC algorithm, and examine the effectiveness of the proposed model comparison approach. We simulate the data from the model in (1.1) for three sample sizes, $n = 500, 5000, 50000$. We let $g_0(w) = 1 - \sin(w+1) + (w+1)^2$, $g_1(w) = -1 - \sin(w) + w^2$,where w is uniformly generated from an evenly spaced grid within the range of -1 to 1. The parameters β_j and γ_j exhibit different values in these samples. Specifically, for the sample where $n = 500$, the parameter values are as follows: $\beta_0 = (0.48, -1.30, 0.15, -0.18, 0.36)'$, $\beta_1 = (0.12, -0.71, 2.52, 0.02, 0.33)'$, $\gamma_0 = (-0.71, 0.27, -0.14)'$, and $\gamma_1 = (-0.81, 0.37, 1.32)'$. For the sample where $n = 5000$, the values are: $\beta_0 = (0.88, 2.10, -0.70, -0.00, -1.00)'$, $\beta_1 = (-0.08, 0.29, 0.42, -1.80, 0.19)'$, $\gamma_0 = (0.10, 0.14, -0.82)'$, and $\gamma_1 = (0.44, 0.08, 0.71)'$. Lastly, for the sample where $n = 50000$, the parameter values are: $\beta_0 = (-0.04, -1.99, 0.08, 0.55, 0.65)'$, $\gamma_0 = (-1.01, -0.51, -0.82)'$, $\beta_1 = (-0.76, 0.20, -1.24, 2.36, -1.24)$, $\gamma_1 = (-0.52, 0.60, 0.88)'$. The covariates x_i are generated from independent standard normal distributions of dimension 5. The covariates z_i consist of a constant term and the first columns of x_i and w_i.

We report means, standard deviations, and 95% credible intervals of the posterior distribution for the treatment effect in each model. Additionally, we present marginal likelihood estimates to facilitate model comparisons. We also report the RD ATE estimates, standard errors, and 95% confidence intervals provided by RDRobust (Calonico et al., 2017).

The marginal likelihood and the estimated treatment effect are presented in Table 1.1. The

heteroskedastic model is supported by the data in all scenarios. As the sample size increases, both homoskedastic and heteroskedastic models yield point estimates that approach the true treatment effect. However, as the sample size increases, the evidence in favor of the heteroskedastic model grows based on the marginal likelihood estimates reported in Table 1.1. The Bayesian models provide more precise estimates than RDRobust, because the latter relies only on data around the cutoff (sample sizes are reported in Table 1.1). Sensitivity analysis was performed to study the impact of the prior distributions on the parameter estimates and their variability; the results, presented in Table A.1 in A.1, show no impact on the point estimates and demonstrate that the reduction in variability is not an artifact of the priors.

Table 1.1: RD ATE with Continuous Outcomes, $\beta_1 \neq \beta_0$

n	**Model**	Δ_{SRD}	$\hat{\Delta}_{SRD}$	**SD**	**95% CI**	$\ln(\mathbf{m}(\mathbf{y}))$	$(\mathbf{n_0}, \mathbf{n_1})$
500	Homosk.	-2.24	-2.36	0.25	$(-2.87, -1.88)$	-765.22	$(230, 270)$
	Heterosk.	-2.24	-2.26	0.18	$(-2.61, -1.92)$	-617.77	$(230, 270)$
	RDRobust	-2.24	-2.63	0.68	$(-4.18, -1.07)$		$(49, 68)$
5000	Homosk.	-2.22	-2.04	0.13	$(-2.29, -1.79)$	-8711.93	$(2480, 2520)$
	Heterosk.	-2.22	-2.10	0.11	$(-2.31, -1.89)$	-7929.93	$(2480, 2520)$
	RDRobust	-2.22	-2.19	0.21	$(-2.71, -1.72)$		$(955, 994)$
50000	Homosk.	-2.21	-2.20	0.04	$(-2.28, -2.12)$	-68715.40	$(24873, 25127)$
	Heterosk.	-2.21	-2.20	0.03	$(-2.26, -2.14)$	-59044.64	$(24873, 25127)$
	RDRobust	-2.21	-2.06	0.07	$(-2.20, -1.88)$		$(7983, 8124)$

SD: Standard deviation for the Bayesian methods; Standard Error for RDRobust.
CI: Credible Interval for the Bayesian methods; Confidence Interval for RDRobust.

A second set of simulations was conducted under the restriction $\beta_0 = \beta_1$, with all other parameters sampled as before. The results, presented in Table 1.2, support the aforementioned conclusions, namely that the heteroskedastic model yields more efficient estimates RDRobust and the homoskedastic specification.

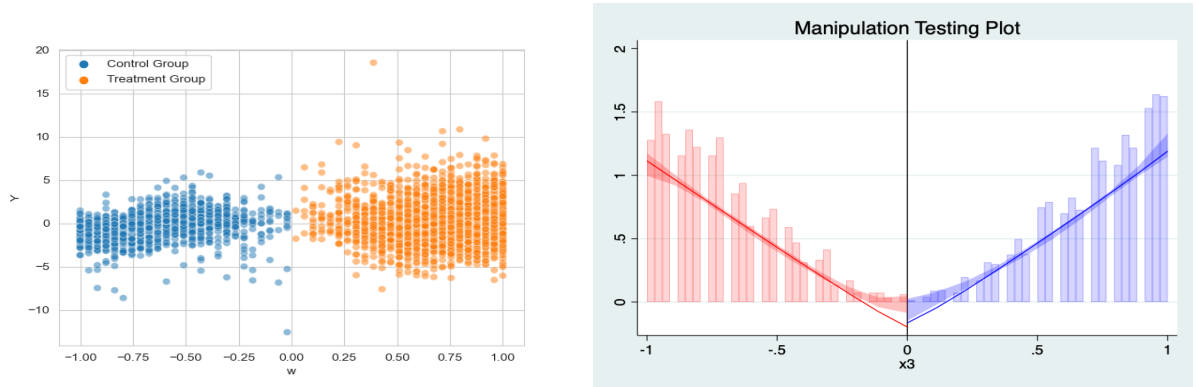
To demonstrate the pitfalls of only employing data within a small band around w^*, we

Table 1.2: RD ATE with Continuous Outcome Variable, $\beta_1 = \beta_0$

n	**Model**	Δ_{SRD}	$\hat{\Delta}_{SRD}$	**SD**	**95% CI**	$\ln(\mathbf{m}(\mathbf{y}))$	$(\mathbf{n_0}, \mathbf{n_1})$
500	Homosk.	-2.16	-2.31	0.228	$(-2.76, -1.87)$	-784.62	$(238, 262)$
	Heterosk.	-2.16	-2.40	0.176	$(-2.74, -2.05)$	-680.02	$(238, 262)$
	RDRobust	-2.16	-2.39	0.225	$(-2.86, -1.82)$		$(72, 80)$
5000	Homosk.	-2.16	-2.10	0.113	$(-2.32, -1.88)$	-7811.57	$(2529, 2471)$
	Heterosk.	-2.16	-2.11	0.110	$(-2.33, -1.90)$	-7621.34	$(2529, 2471)$
	RDRobust	-2.16	-1.92	0.160	$(-2.22, -1.50)$		$(2424, 2576)$
50000	Homosk.	-2.16	-2.15	0.039	$(-2.22, -2.07)$	-54592.15	$(24815, 25185)$
	Heterosk.	-2.16	-2.16	0.029	$(-2.21, -2.10)$	-48880.91	$(24815, 25185)$
	RDRobust	-2.16	-2.15	0.032	$(-2.21, -2.06)$		$(7429, 7479)$

present a study with a sample size $n = 5000$ from the model in (1.1), where $g_0(w) = \sin(w) + \exp\left(-20(w+0.5)^2\right)$, $g_1(w) = 1.2 - \sin(w) - \exp\left(-20(w-0.5)^2\right)$, $\gamma_0 = (-2, 2, 1)'$, $\gamma_1 = (-2, 2, -1)'$, $\beta_0 = -0.37$, and $\beta_1 = -2.13$. In this case, the cutoff point w^* is in a low-density region of w presented in Figure 1.1. The covariate x_i is sampled from a standard normal distribution, and the covariates z_i consist of a constant term, $||w_i| - 1.5|$, and x_i. In this scenario, the paucity of observations around the cutoff is compounded by more pronounced heteroskedasticity in that region. The generated data passed the density test (McCrary, 2008) with p-value 0.4655.

Figure 1.1: Data and Running Variable Density with Sparse Data around w^*



(a) Data

(b) Running Variable Density

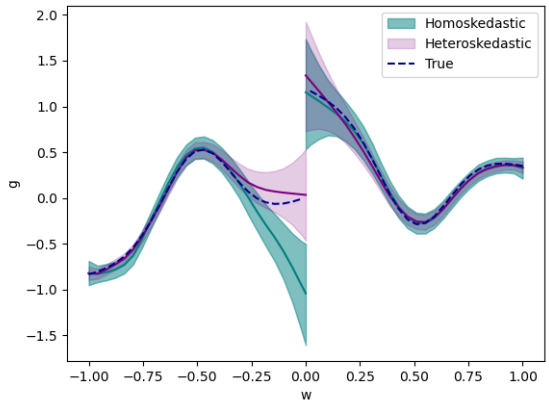
The estimated RD ATE is provided in Table 1.3. The estimated parameter $\hat{g}_0$ and $\hat{g}_1$

can be found in Figure 1.2. In this context, the estimates from the homoskedastic model can be adversely affected by the outliers near the cutoff point, which ultimately resulted in dramatically distorted RD ATE estimates. On the other hand, the figure shows that the heteroskedastic model can estimate the true function well, owing to the fact that the data points are properly weighted. In this scenario, RDRobust yielded a notably wide 95% confidence interval, primarily due to the drastically smaller number of data points near the cutoff. We take this as a warning about the importance of accounting for the behavior of both w and the error variances around the cutoff w^* in determining the merits of alternative estimators.

Table 1.3: RD ATE with Sparse Data around w^*

Model	Δ_{SRD}	$\hat{\Delta}_{SRD}$	**SD**	**95% CI**	$\ln(\mathbf{m}(\mathbf{y}))$	$(\mathbf{n_0}, \mathbf{n_1})$
Homosk.	1.20	2.21	0.42	$(1.38, 3.02)$	-7642.76	$(2489, 2511)$
Heterosk.	1.20	1.32	0.40	$(0.53, 2.09)$	-6066.56	$(2489, 2511)$
RDRobust	1.20	2.10	0.87	$(-0.15, 4.12)$		$(202, 190)$

Figure 1.2: Estimated Functions $\hat{g}_j$ in Sample with Sparse Data around w^*

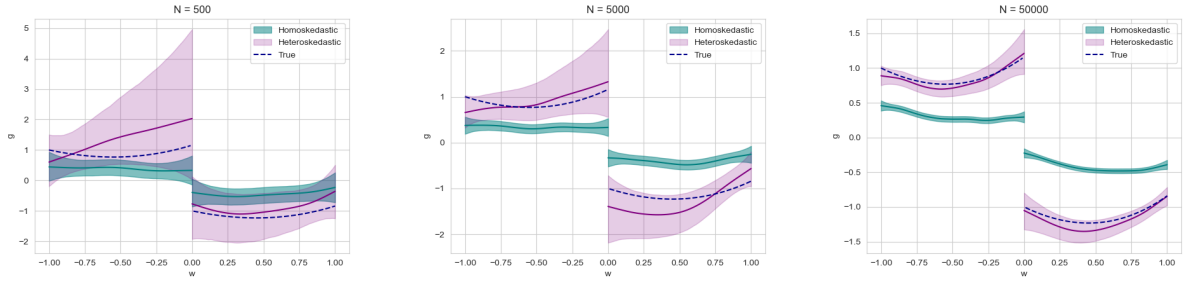


Finally, we simulate data for settings with binary outcomes using (1.6) with $g_0(w) = 1 - \sin(w+1) + (w+1)^2$, $g_1(w) = -1 - \sin(w) + w^2$, $\gamma_0 = \gamma_1 = 2$, and w is uniformly sampled from a grid within the range of -1 to 1. The remaining parameters for the scenario with $n = 500$ are as follows: $\beta_0 = (-0.10, 1.66, -0.70)'$, and $\beta_1 = (-1.19, 1.32, 0.79)$. In the case of $n =$

5000, the corresponding values are: $\beta_0 = (1.74, 0.16, 0.02)'$, and $\beta_1 = (1.15, -0.60, -1.87)$. Lastly, for the scenario with $n = 50000$, the parameter values are $\beta_0 = (1.53, -0.13, 0.06)'$, and $\beta_1 = (0.24, 2.00, -0.81)'$. The covariates x_i are sampled from independent standard normal distributions with a dimension of 3, and the covariate z_i consists of $||w_i| - 1.5|$.

Estimates of the nonparametric functions $\hat{g}_0$ and $\hat{g}_1$ are depicted in Figure 1.3. Table 1.4 presents the estimated RD ATE and log-marginal likelihoods. As the sample size increases, the heteroskedastic model provides a closer approximation, whereas both the homoskedastic model and RDRobust exhibit inconsistencies while also significantly understating the estimation variability. The impact of ignoring heteroskedasticity is amplified when the nonlinear features of the model become more prominent. With larger samples, the evidence in support of the heteroskedastic specification grows stronger as demonstrated by the marginal likelihood results in Table 1.4.

Figure 1.3: Estimated Functions $\hat{g}_j$ with Binary Outcomes



(a) $n = 500$ (b) $n = 5000$ (c) $n = 50000$

1.2.3 Application: The Effect of Academic Probation on Student Performance

Academic probation is commonly used as a catalyst to motivate students and improve effort levels. Fletcher and Tokmouline (2018) and Wright (2020) employed RDD to evaluate the impact of academic probation on academic performance. We use public data from the Texas

Table 1.4: RD ATE with Binary Outcomes

n	**Model**	Δ_{SRD}	$\hat{\Delta}_{SRD}$	**SD**	**95% CI**	$\ln(\mathbf{m}(\mathbf{y}))$	$(\mathbf{n_0}, \mathbf{n_1})$
500	Homosk.	-0.23	-0.19	0.10	$(-0.39, -0.01)$	-306.18	$(262, 238)$
	Heterosk.	-0.23	-0.24	0.10	$(-0.43, -0.04)$	-301.29	$(262, 238)$
	RDRobust	-0.23	-0.23	0.13	$(-0.54, 0.08)$		$(109, 91)$
5000	Homosk.	-0.26	-0.22	0.04	$(-0.31, -0.14)$	-2811.87	$(2529, 2471)$
	Heterosk.	-0.26	-0.30	0.05	$(-0.39, -0.20)$	-2770.39	$(2529, 2471)$
	RDRobust	-0.26	-0.22	0.05	$(-0.34, -0.12)$		$(941, 932)$
50000	Homosk.	-0.26	-0.17	0.02	$(-0.20, -0.13)$	-27844.76	$(24815, 25185)$
	Heterosk.	-0.26	-0.26	0.02	$(-0.30, -0.22)$	-27470.87	$(24815, 25185)$
	RDRobust	-0.26	-0.18	0.02	$(-0.22, -0.14)$		$(6949, 7155)$

Higher Education Opportunity Project (THEOP) to study the impact of academic probation on students' academic performance. We performed analysis employing both homoskedastic and heteroskedastic models, using model comparison techniques to assess their practical relevance in this context.

The treated group consists of students who received academic probation at the end of their first semester. The outcome variables are the students' GPA in two subsequent semesters as well as their graduation status. We use the longitudinal administrative data from the University of Texas, Austin, for students admitted from 1991 through 2000. Students are placed on probation (the treatment) if their cumulative GPA falls below 2.0; such students must raise their GPA above the threshold or face dismissal from the university. Covariates in this setting include the student's gender, citizenship, race, standardized SAT score, high school decile, an indicator of private high school attendance, and an indicator if the student has a major in the first semester. Summary statistics for the data are presented in Tables 1.5, 1.6 and 1.7.

One key underlying assumption for the sharp RD design is that the students near the threshold can not manipulate their GPA. Following McCrary (2008), a density test of the running variable was performed, resulting in a t-statistic of -1.00 and a corresponding p-value of

Table 1.5: Summary Statistics, Second Semester GPA Data

	Control ($n_0 = 37980$)		Treated ($n_1 = 5789$)	
Variable	**Mean**	**SD**	**Mean**	**SD**
Female	0.52	0.50	0.39	0.49
Non US Citizen	0.00	0.05	0.00	0.04
Minority	0.03	0.18	0.07	0.25
SAT	0.08	1.00	−0.40	0.91
Second Decile	0.26	0.44	0.31	0.46
Third Decile	0.12	0.32	0.21	0.41
Fourth Decile or Below	0.08	0.28	0.22	0.42
Private High School	0.05	0.22	0.05	0.21
Has Major	0.75	0.43	0.70	0.46
Second semester term GPA	3.00	0.76	2.03	0.89
First semester term GPA	3.21	0.55	1.34	0.50

Table 1.6: Summary Statistics, Third Semester GPA Data

	Control ($n_0 = 36738$)		Treated ($n_1 = 4366$)	
Variable	**Mean**	**SD**	**Mean**	**SD**
Female	0.52	0.50	0.39	0.49
Non US Citizen	0.00	0.05	0.00	0.04
Minority	0.03	0.18	0.07	0.25
SAT	0.07	1.00	−0.40	0.92
Second Decile	0.26	0.44	0.31	0.46
Third Decile	0.12	0.32	0.21	0.41
Fourth Decile or Below	0.08	0.27	0.23	0.42
Private High School	0.05	0.22	0.05	0.22
Has Major	0.76	0.43	0.69	0.46
Third semester term GPA	2.95	0.81	2.17	0.89
First semester term GPA	3.21	0.54	1.41	0.45

0.32. Therefore, there is no evidence to suggest that students manipulate their GPA to avoid the treatment. Additionally, since the GPA data is rounded to the nearest tenth, it is possible that some students with GPA of $[1.95, 2.05)$ are misclassified. To address this issue, students who have a first semester cumulative GPA of exactly 2.0 are eliminated from the sample (Fletcher and Tokmouline, 2018), although their covariates X and Z can be retained for evaluating the averages that define treatment effects.

Table 1.7: Summary Statistics, Graduation Data

	Control ($n_0 = 38525$)		**Treated ($n_1 = 6494$)**	
Variable	**Mean**	**SD**	**Mean**	**SD**
Female	0.52	0.50	0.39	0.49
Non US Citizen	0.00	0.05	0.00	0.04
Minority	0.03	0.18	0.07	0.25
SAT	0.08	1.00	−0.38	0.92
Second Decile	0.26	0.44	0.31	0.46
Third Decile	0.12	0.32	0.21	0.41
Fourth Decile or Below	0.08	0.28	0.23	0.42
Private High School	0.05	0.22	0.05	0.21
Has Major	0.75	0.43	0.69	0.46
4-Year Graduation	0.53	0.50	0.13	0.33
Graduation	0.80	0.40	0.33	0.47

Our analysis centers on the effect of heteroskedasticity and sidesteps potential complications that may arise due to sample selection or endogeneity related to the decision to stay in school (see, e.g., Dong, 2019). Such complications are unlikely to be important at short time horizons, such as the second and third semesters, where we see a strong impact on GPA, but could be relevant in the longer run where our results are inconclusive. In particular, the estimated functions $\hat{g}_0$ and $\hat{g}_1$ are represented in Figure 1.4, whereas estimates of the RD ATE and the marginal likelihoods are provided in Table 1.8. Notably, academic probation is practically relevant with a considerable positive effect on subsequent semester GPAs in all the Bayesian models, while the estimated impact from RDRobust is not always of the expected sign, and lacks statistical significance. The impact of academic probation on graduation rates is indeterminate according to the results of all models. Additionally, the marginal likelihoods suggest that the heteroskedastic model, with the constraint that $\beta_0 = \beta_1$, is preferred in all scenarios.

Table 1.8: Impact of Academic Probation on Student Performance

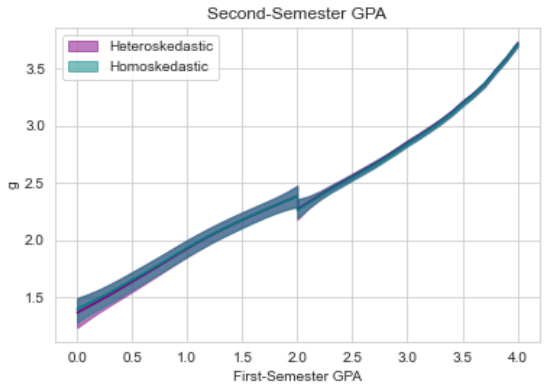
Outcome	**Model**	$\hat{\Delta}_{SRD}$	**SD**	**95% CI**	$\ln(\mathbf{m}(\mathbf{y}))$	$(\mathbf{n_0}, \mathbf{n_1})$
GPA, 2nd Semester	Homosk.	0.17	0.05	$(0.08, 0.27)$	-42714.08	$(37980, 5789)$
	Homosk., $\beta_0 = \beta_1$	0.17	0.05	$(0.07, 0.27)$	-42677.36	$(37980, 5789)$
	Heterosk.	0.17	0.06	$(0.06, 0.28)$	-41248.43	$(37980, 5789)$
	Heterosk., $\beta_0 = \beta_1$	0.17	0.06	$(0.06, 0.28)$	-41211.59	$(37980, 5789)$
	RDRobust	-0.14	0.10	$(-0.45, 0.26)$		$(1937, 3030)$
GPA, 3rd Semester	Homosk.	0.16	0.06	$(0.05, 0.27)$	-44410.90	$(36738, 4366)$
	Homosk., $\beta_0 = \beta_1$	0.14	0.06	$(0.03, 0.26)$	-44383.80	$(36738, 4366)$
	Heterosk.	0.14	0.06	$(0.02, 0.26)$	-43617.56	$(36738, 4366)$
	Heterosk., $\beta_0 = \beta_1$	0.12	0.04	$(0.01, 0.24)$	-43592.42	$(36738, 4366)$
	RDRobust	-0.14	0.11	$(-0.43, 0.34)$		$(1671, 2824)$
4-Year Grad.	Homosk.	-0.02	0.02	$(-0.06, 0.03)$	-27659.71	$(38525, 6494)$
	Homosk., $\beta_0 = \beta_1$	-0.03	0.02	$(-0.07, 0.02)$	-27631.28	$(38525, 6494)$
	Heterosk.	-0.03	0.02	$(-0.07, 0.02)$	-27642.73	$(38525, 6494)$
	Heterosk., $\beta_0 = \beta_1$	-0.03	0.02	$(-0.07, 0.01)$	-27619.22	$(38525, 6494)$
	RDRobust	-0.03	0.05	$(-0.30, 0.08)$		$(2022, 3138)$
Grad.	Homosk.	0.02	0.02	$(-0.03, 0.07)$	-21773.71	$(38525, 6494)$
	Homosk., $\beta_0 = \beta_1$	0.01	0.02	$(-0.03, 0.06)$	-21752.81	$(38525, 6494)$
	Heterosk.	0.01	0.02	$(-0.04, 0.05)$	-21754.73	$(38525, 6494)$
	Heterosk., $\beta_0 = \beta_1$	0.01	0.02	$(-0.04, 0.06)$	-21749.15	$(38525, 6494)$
	RDRobust	-0.01	0.06	$(-0.30, 0.14)$		$(2022, 3138)$

1.3 Rubin Causal (Roy-Type) Model

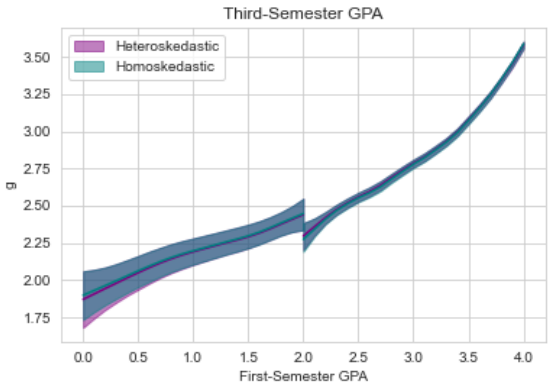
In this section, we introduce a potential outcome framework (Roy, 1951; Rubin, 1974, 1977, 1978, 2004, 2005) with self-selection for estimating the treatment effect, following the approach outlined in Chib (2007). We assume that there are two potential outcome variables y_0 and y_1 for the treated and untreated states. The binary treatment status T is determined by a latent variable T^* and $T = \mathbb{1}\{T^* > 0\}$. We also assume that there is heteroskedasticity in both the treatment assignment and the outcome. The model can be represented as

$$y_i = X_i\beta + \epsilon_i, \epsilon_i \sim N\left(0, \Omega_i\right),$$

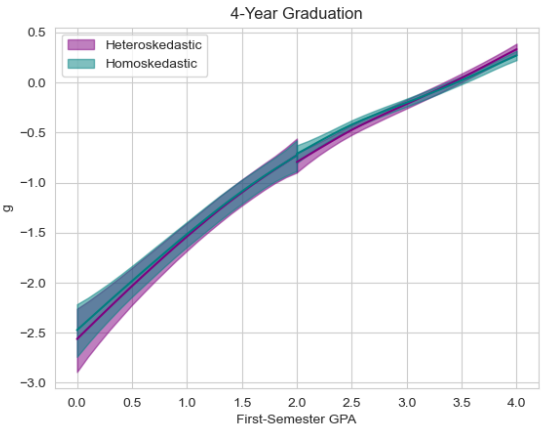
Figure 1.4: Academic Probation: Estimated Nonparametric Functions



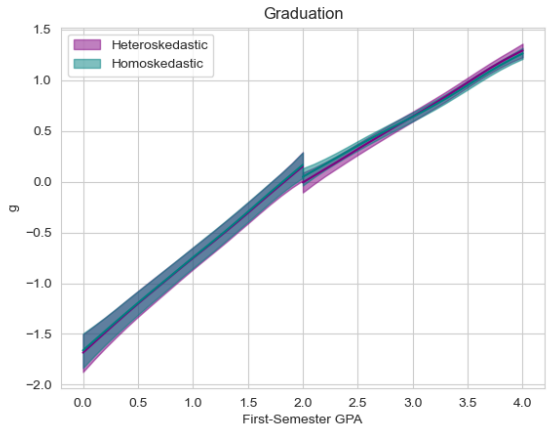
(a) Second-Semester GPA



(b) Third-Semester GPA



(c) 4-Year Graduation Rate



(d) Graduation Rate

where

$$y_i = \begin{pmatrix} T_i^* \\ y_{i0} \\ y_{i1} \end{pmatrix}, \quad X_i = \begin{pmatrix} x_{iT}' & 0 & 0 \\ 0 & x_{i0}' & 0 \\ 0 & 0 & x_{i1}' \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_T \\ \beta_0 \\ \beta_1 \end{pmatrix}, \text{ and } \varepsilon = \begin{pmatrix} \varepsilon_{iT} \\ \varepsilon_{i0} \\ \varepsilon_{i1} \end{pmatrix}.$$

Let N_j denote the set $\{i : T_i = j\}$ and n_j to denote the cardinality of N_j, $j \in \{0, 1\}$. The covariance matrix Ω_i is given by

$$\Omega_i = \begin{pmatrix} \omega_{iTT} & \omega_{i0T} & \omega_{i1T} \\ \omega_{i0d} & \omega_{i00} & \omega_{i01} \\ \omega_{i1d} & \omega_{i01} & \omega_{i11} \end{pmatrix},$$

but due to the missing counterfactuals, ω_{i01} is not identified. To deal with this feature of the model, estimation can proceed either by augmenting the MCMC sampler with the missing counterfactuals (Li, Poirier and Tobias, 2004) or by solely involving the observed outcomes Chib (2007). We pursue the latter approach, but develop an algorithm, which, relative to the approach of Chib (2007), simplifies estimation by not involving MH steps in the homoskedastic case. To prepare the groundwork for our subsequent discussion, we introduce the following notation

$$\Omega_{i0} = \begin{pmatrix} \omega_{iTT} & \omega_{iT0} \\ \omega_{i0T} & \omega_{i00} \end{pmatrix}, \quad \Omega_{i1} = \begin{pmatrix} \omega_{iTT} & \omega_{iT1} \\ \omega_{i1T} & \omega_{i11} \end{pmatrix}, \quad J_0 = \begin{pmatrix} I & 0 & 0 \\ 0 & 0 & I \end{pmatrix}, \quad J_1 = \begin{pmatrix} 0 & I & 0 \\ 0 & 0 & I \end{pmatrix},$$
$$\tilde{X}_{i0} = \begin{pmatrix} x'_{iT} & 0 \\ 0 & x'_{i0} \end{pmatrix}, \quad \tilde{X}_{i1} = \begin{pmatrix} x'_{iT} & 0 \\ 0 & x'_{i1} \end{pmatrix}, \quad \tilde{y}_{i0} = \begin{pmatrix} T_i^* \\ y_{i0} \end{pmatrix}, \quad \tilde{y}_{i1} = \begin{pmatrix} T_i^* \\ y_{i1} \end{pmatrix}.$$

Thus we have $J_0\beta = (\beta'_T, \beta'_0)'$ and $J_1\beta = (\beta'_T, \beta'_1)'$. The complete data density is given by

$$f(y_0, y_1, T^*|\beta_0, \beta_1, \Omega_0, \Omega_1) = \left[\prod_{i \in N_0} f(\tilde{y}_{i0}|\beta_0, \Omega_{i0}) \, \mathbb{1}\{T_i^* \leq 0\}\right] \times \left[\prod_{i \in N_1} f(\tilde{y}_{i1}|\beta_1, \Omega_{i1}) \, \mathbb{1}\{T_i^* > 0\}\right].$$

In the homoskedastic model, we impose the normalization $\omega_{TT} = 1$ for identification purposes

and define the following quantities

$$\Omega_0 = \begin{pmatrix} 1 & \omega_{T0} \\ \omega_{0T} & \omega_{00} \end{pmatrix}, \qquad \Omega_1 = \begin{pmatrix} 1 & \omega_{T1} \\ \omega_{1T} & \omega_{11} \end{pmatrix},$$

$$\Omega_{22\cdot 1} = \omega_{11} - \omega_{1T}\omega_{T1}, \quad \Omega_{22\cdot 0} = \omega_{00} - \omega_{0T}\omega_{T0}.$$

Because of the unit restrictions in Ω_0 and Ω_1, we work directly with the quantities $\Omega_{22\cdot j}$ and w_{jT}, from which Ω_0 and Ω_1 can be recovered (Dreze and Richard, 1983; Munkin and Trivedi, 2003; Chib, Greenberg and Jeliazkov, 2009; Vossmeyer, 2016), and let $\Omega_{22\cdot j} \sim IG\left(r_j/2, R_j/2\right)$ and $\omega_{jT}|\Omega_{jj\cdot 2} \sim N\left(q_j, \Omega_{22\cdot j}\right)$, $j = 0, 1$, and $\beta \sim N\left(b_0, B_0\right)$. Algorithm 3 presents a Gibbs sampler for this homoskedastic model that is based on simulation from fully tractable densities and does not require any MH steps.

To extend the model to the case of multivariate heteroskedasticity, allowing for changing variances in both treatment assignment and the potential outcomes, we decompose the covariance matrices (Chan and Jeliazkov, 2009*b*) as

$$\Omega_{i0} = L_0 G_{i0} L_0', \quad \Omega_{i1} = L_1 G_{i1} L_1',$$

where, for $j = 0, 1$,

$$L_j \equiv \begin{pmatrix} 1 & 0 \\ a_{jT} & 1 \end{pmatrix}, \quad G_{ij} \equiv \begin{pmatrix} \lambda_{iT} & 0 \\ 0 & \lambda_{ij} \end{pmatrix}.$$

The model can be rewritten as

$$\begin{pmatrix} T_i^* \\ y_{ij} \end{pmatrix} = \begin{pmatrix} x_{iT}' & \mathbf{0} \\ \mathbf{0} & x_{ij}' \end{pmatrix} \begin{pmatrix} \beta_T \\ \beta_j \end{pmatrix} + L_j \begin{pmatrix} \psi_{iT} \\ \psi_{ij} \end{pmatrix}, \text{where } \begin{pmatrix} \psi_{iT} \\ \psi_{ij} \end{pmatrix} \sim N\left(\mathbf{0}, G_{ij}\right), \tag{1.11}$$

with heterogeneity allowed to depend on observables through $\lambda_{ij} = \exp(z_{ij}'\gamma_j)$, $\lambda_{iT} =$

Algorithm 3 (Rubin Causal (Roy-Type) Model with Homoskedasticity)

---

(1) Sample $\beta \sim N\left(\hat{b}, \hat{B}\right)$, where $\hat{b} = \hat{B}\left(B_0^{-1}b_0 + \sum_{i\in N_0} J_0'\tilde{X}_{i0}'\Omega_0^{-1}\tilde{y}_{i0} + \sum_{i\in N_1} J_1'\tilde{X}_{i1}'\Omega_1^{-1}\tilde{y}_{i1}\right)$, and $\hat{B} = \left(B_0^{-1} + \sum_{i\in N_0} J_0'\tilde{X}_{i0}'\Omega_0^{-1}\tilde{X}_{i0}J_0 + \sum_{i\in N_1} J_1'\tilde{X}_{i1}'\Omega_1^{-1}\tilde{X}_{i1}J_1\right)^{-1}$.

(2) Sample $T_i^* \sim TN_{\mathcal{B}_i}\left(\mu_{iTj}, \hat{\omega}_{TT}\right)$, where $\mathcal{B}_i = (-\infty, 0]$ if $i \in N_0$, $\mathcal{B}_i = (0, \infty)$ if $i \in N_1$, $\mu_{iTj} = x_{iT}'\beta_T + \omega_{jT}\omega_{jj}^{-1}\left(y_{ij} - x_{ij}'\beta_j\right)$, and $\hat{\omega}_{22} = 1 - \omega_{jT}\omega_{jj}^{-1}\omega_{jT}$, $j = 0, 1$.

(3) For $j = 0, 1$, $\pi\left(\omega_{jT}|\Omega_{22\cdot j}, \beta, y_i, z_i\right) = f_N\left(\omega_{jT}|q_t, \Omega_{22\cdot j}\right)\prod_{i\in N_j} f_N\left(y_{ij}|\mu_{ij|2}, \Omega_{22\cdot j}\right)$, where $\mu_{ij|2} = x_{ij}'\beta_j + \omega_{jT}\left(T_i^* - x_{iT}'\beta_T\right)$. Thus the posterior distribution for ω_{jT} is $\omega_{jT} \sim N\left(\hat{q}_j, \hat{\omega}_{22\cdot j}\right)$, where $\hat{\omega}_{22\cdot j} = \left(\Omega_{22\cdot j}^{-1} + \left(\sum_{i=1}^{n_j}\left(\varepsilon_{iT}^2\Omega_{22\cdot j}^{-1}\right)\right)^{-1}\right)^{-1}$, and $\hat{q}_j = \hat{\omega}_{22\cdot j}\left(\Omega_{22\cdot j}^{-1}q_t + \Omega_{22\cdot j}^{-1}\sum_{i=1}^{n_j}\varepsilon_{iT}\varepsilon_{ij}\right)$, where $\varepsilon_{ij} \equiv y_{ij} - x_{ij}'\beta_j$ and $\varepsilon_{iT} \equiv T_i^* - x_{iT}'\beta_T$.

(4) For $j = 0, 1$,

$$\pi\left(\Omega_{22\cdot j}|\omega_{jT}, \beta, y_i, T^*\right) = \pi\left(\Omega_{22\cdot j}\right) f_N\left(\omega_{jT}|q_t, \Omega_{22\cdot j}\right)\prod_{i\in N_j} f_N\left(y_{ij}|\mu_{ij|2}, \Omega_{22\cdot j}\right).$$

The posterior distribution is as follows: $\Omega_{22\cdot j} \sim IG\left(\frac{\hat{r}_j}{2}, \frac{\hat{R}_j}{2}\right)$, where $\hat{r}_j = r_j + 1 + n_j$, and $\hat{R}_j = R_j + \left(\omega_{jT} - q_t\right)^2 + \sum_{i\in N_j}\left(\varepsilon_{ij} - \omega_{jT}\varepsilon_{iT}\right)^2$.

$\exp(z_{iT}'\gamma_T)$. The prior distributions are specified as

$$\beta \sim N\left(b_0, B_0\right), \quad \gamma_j \sim N\left(\gamma_{0j}, \Gamma_{0j}\right), \quad \gamma_T \sim N\left(\gamma_{0T}, \Gamma_{0T}\right), \quad a_{jd} \sim N\left(a_{0j}, A_{0j}\right),$$

and the estimation algorithm is detailed in Algorithm 4. In our model, the average treatment effect (ATE) and the average treatment effect on the treated (ATT) are defined as

$$ATE = E(Y_1 - Y_0) = E(x_i'\beta_1 - x_i'\beta_0), \quad ATT = E(Y_1 - Y_0|D = 1) = E(x_{i1}'\beta_1 - x_{i1}'\beta_0),$$

and can be estimated using the MCMC output.

The marginal likelihood can be estimated using the approach introduced in Section 1.2.1. We

Algorithm 4 (Rubin Causal (Roy-Type) Model with Heteroskedasticity)

---

(1) Sample $\beta \sim N\left(\hat{b}, \hat{B}\right)$, where $\hat{b} = \hat{B}\left(B_0^{-1}b_0 + \sum_{i\in N_0} J_0'\tilde{X}_{i0}'\Omega_0^{-1}\tilde{y}_{i0} + \sum_{i\in N_1} J_1'\tilde{X}_{i1}'\Omega_1^{-1}\tilde{y}_{i1}\right)$, and $\hat{B} = \left(B_0^{-1} + \sum_{i\in N_0} J_0'\tilde{X}_{i0}'\Omega_0^{-1}\tilde{X}_{i0}J_0 + \sum_{i\in N_1} J_1'\tilde{X}_{i1}'\Omega_1^{-1}\tilde{X}_{i1}J_1\right)^{-1}$.

(2) Sample $T_i^* \sim TN\left(\mu_{iTj}, \hat{\omega}_{iTT}\right)$, where $T_i^* \in (-\infty, 0)$ if $i \in N_0$, $T_i^* \in [0, \infty)$ if $i \in N_1$, $\mu_{iTj} = x_{iT}'\beta_T + \omega_{ijT}\omega_{ijj}^{-1}\left(y_{ij} - x_{ij}'\beta_j\right)$, and $\hat{\omega}_{iTT} = \omega_{iTT} - \omega_{ijT}\omega_{ijj}^{-1}\omega_{ijT}$, $j = 0, 1$.

(3) Sample $a_{jT} \sim N\left(\hat{a}_j, \hat{A}_j\right)$, $j = 0, 1$, where $\hat{A}_j = \left(A_{0j}^{-1} + \sum_{i=1}^{n_j}\psi_{iT}'\lambda_{ij}^{-1}\psi_{iT}\right)^{-1}$ and $\hat{a}_j = \hat{A}_j\left(A_{0j}^{-1}a_{0j} + \sum_{i=1}^{n_j}\psi_{iT}'\lambda_{ij}^{-1}u_{ij}\right)$, where $u_{ij} \equiv y_{ij} - x_{ij}'\beta_j$.

(4) Sample $[\gamma_T|T^*, \beta_T, a_{0T}, a_{1T}]$ using an MH step by drawing a proposal value $\gamma_T^\dagger \sim q\left(\gamma_T|\hat{\gamma}_T, V_T\right)$, where $e_i = T_i^* - x_{iT}'\beta_T$ and $\hat{\gamma}_T$ and V_T are defined similarly to (1.5) using the current value of γ_j and T_i^*. Also use $\gamma_T^\dagger$ in equation (1.5) to produce $\hat{\gamma}_j^\dagger$ and accept the proposed $\gamma_T^\dagger$ with probability

$$\alpha = \min\left\{1, \frac{f\left(T_i^*|a_{0T}, a_{1T}, \beta_T, \gamma_T^\dagger\right)\pi\left(\gamma_T^\dagger|\gamma_{T0}, \Gamma_{T0}\right)q\left(\gamma_T|\hat{\gamma}_T^\dagger, V_T\right)}{f(T_i^*|a_{0T}, a_{1T}, \beta_T, \gamma_T)\pi\left(\gamma_T|\gamma_{T0}, \Gamma_{T0}\right)q\left(\gamma_T^\dagger|\hat{\gamma}_T, V_T\right)}\right\}.$$

otherwise the current value γ_T is repeated in the next MCMC iteration.

(5) Let $e_{ij} = \left(y_{ij} - x_{ij}'\beta_j - a_{jT}\psi_{iT}\right)$, $\eta_{ij} = z_{ij}'\gamma_j + \frac{e_{ij}^2 - \omega_{ijj}}{\omega_{ijj}}$, and $\eta_j = \left(\eta_{1j}, \ldots, \eta_{n_jj}\right)'$ and sample $[\gamma_j|a_{jT}, \beta_j]$ similarly to Step (4).

---

compute the posterior ordinate of parameters with known densities using equation (1.10), and with non-standard densities using equation (1.9). Future extensions of this framework to cases where a system of outcomes, possibly involving endogeneity, incidental truncation, multiple selection mechanisms, or unknown covariate functions in the treated and untreated states can be pursued along the lines of Chib, Greenberg and Jeliazkov (2009) and Vossmeyer (2016).

1.3.1 Simulation Study

In this section, we performed a simulation study with the aim of achieving the same goals as outlined in Section 1.2.2. We generate simulated data from the model in (1.11) for three different sample sizes: $n = 500$, 5000, and 50000. We let $a_{0T} = -0.2$, $a_{1T} = 0.2$. The parameter values vary depending on the sample size. For the sample with $n = 500$, the parameters are $\gamma_0 = (1.04, 0.15)'$, $\gamma_1 = (-1.70, 0.15)'$, $\gamma_d = (0.99, 1.30)'$, and $\beta = (-0.20, 0.79, 0.18, 1.08, 0.38, -1.21, 1.82)'$. For $n = 5000$, the parameter values are $\gamma_0 = (1.13, 0.14)'$, $\gamma_1 = (-0.62, 0.78)'$, $\gamma_d = (-0.44, 0.92)'$, and $\beta = (0.58, 0.33, -0.01, -0.76, -0.28, -1.26, -0.69)'$. For the sample where $n = 50000$, the values are $\gamma_0 = (-0.05, -0.51)'$, $\gamma_1 = (0.06, 0.71)'$, $\gamma_d = (0.84, -0.60)'$, and $\beta = (-0.62, 0.94, 0.20, 1.59, 0.45, -0.12, -0.82)'$. The covariates x_{i0} comprise a constant term and a variable generated from a standard normal distribution with a dimension of 2. The covariates x_{i1} mirrors x_{i0}, while x_{iT} includes x_{i0} and a new variable generated from a standard normal distribution with a dimension of 3.

We report mean, standard deviations, and 95% credible intervals of the posterior distribution for the estimated treatment effects in each model. The estimated ATE and ATT are summarized in Table 1.9. In all cases, the heteroskedastic model is supported in all scenarios based on the marginal likelihood results. These findings highlight that ignoring heteroskedasticity leads to inconsistent and biased ATE and ATT estimates.

1.3.2 Application: The Effect of Medigap on Healthcare Expenditure

In this application, we consider the influence of private health insurance on healthcare expenditures of the elderly using the Medical Expenditure Panel Survey (MEPS). For individuals aged 65 and above, Medicare provides coverage, but some seniors opt to purchase private

Table 1.9: Treatment Effect Results Using Rubin Causal (Roy-Type) Model

n	**Model**		**True**	**Mean**	**SD**	**95% CI**	$\ln(\mathbf{m}(\mathbf{y}))$
500	Heterosk.	**ATE**	-2.36	-2.30	0.19	$(-2.68, -1.93)$	-798.15
		ATT	-1.50	-1.47	0.25	$(-1.96, -0.97)$	
	Homosk.	**ATE**	-2.36	-1.54	0.29	$(-2.11, -0.98)$	-853.74
		ATT	-1.50	-0.82	0.36	$(-1.53, -0.11)$	
5000	Heterosk.	**ATE**	-0.49	-0.47	0.15	$(-0.75, -0.16)$	-7073.77
		ATT	-0.57	-0.54	0.15	$(-0.83, -0.21)$	
	Homosk.	**ATE**	-0.49	-1.48	0.34	$(-2.17, -0.81)$	-8108.51
		ATT	-0.57	-1.56	0.36	$(-2.29, -0.86)$	
50000	Heterosk.	**ATE**	-1.70	-1.67	0.03	$(-1.74, -1.61)$	-77644.71
		ATT	-2.68	-2.66	0.03	$(-2.72, -2.60)$	
	Homosk.	**ATE**	-1.70	-3.18	0.04	$(-3.26, -3.10)$	-90039.30
		ATT	-2.68	-3.33	0.03	$(-3.39, -3.27)$	

insurance known as Medigap to supplement their Medicare benefits. Medigap policies typically offer enhanced coverage compared to the basic Medicare policy, and individuals often choose them in anticipation of reducing out-of-pocket healthcare costs.

We partition the data into two distinct subsets. One sample spans the years 2018 to 2019 prior to the COVID-19 pandemic, and the other comprises survey data from 2020. This division accounts for the potential impact of the pandemic on individuals' behavior. In our study, we assess the impact of acquiring Medigap policies on out-of-pocket healthcare expenditures, employing both heteroskedastic and homoskedastic models.

We incorporate self-perceived health status variables, the number of chronic conditions, location, and various demographic variables as covariates that influence healthcare expenditures. We assume that family income only affects the purchase of private insurance, and does not alter health care utilization directly. Variable definitions and summary statistics are presented in Table 1.10. In the regression, age is standardized, and due to excessive right skew, expenditure (in thousands of dollars) is stabilized using the square root transformation (Amaratunga and Cabrera, 2001). Additionally, we consider models where the variance of

treatment assignment depends on family income, and the variance of healthcare expenditure depends on age and the number of chronic conditions. In this application, heteroskedasticity remains our primary focus, although modeling could be generalized to explicitly model the choice of specific Medigap plans based on their anticipated healthcare expenditures and accommodate the potential endogeneity of the Medicaid variable.

Table 1.10: Variable Definitions and Summary Statistics, Medigap Policies Data

		2020 ($n = 5019$)		2018-19 ($n = 10226$)	
Variables	**Description**	**Mean**	**SD**	**Mean**	**SD**
AGE	Age	73.52	6.24	73.65	6.42
FAMINC	Family income (as % of poverty line)	4.11	3.88	4.19	3.90
NUM_VISIT	# of office-based provider visits	10.11	14.53	12.19	16.63
NUM_CHRON	# of chronic conditions	3.93	2.26	3.82	2.25
EXCHLTH	1 if self-perceived health is excellent	0.16	0.37	0.17	0.38
POORHLTH	1 if self-perceived health is poor	0.04	0.19	0.05	0.22
EXCMHLTH	1 if self-perceived mental health is excellent	0.28	0.45	0.30	0.46
POORMHLTH	1 if self-perceived mental health is poor	0.02	0.14	0.02	0.15
EMPLOYED	1 if employed	0.20	0.40	0.19	0.39
PRIVATE	1 if has private insurance	0.43	0.50	0.46	0.50
NORTHEAST	1 if lives in northeastern U.S.	0.18	0.38	0.17	0.38
MIDWEST	1 if lives in midwestern U.S.	0.21	0.41	0.21	0.41
WEST	1 if lives in western U.S.	0.24	0.43	0.24	0.43
MALE	1 if male	0.44	0.50	0.45	0.50
BLACK	1 if African American	0.12	0.33	0.13	0.33
MARRIED	1 if married	0.50	0.50	0.53	0.50
COLLEGE	1 if has a college degree	0.33	0.47	0.30	0.46
MEDICAID	1 if covered by Medicaid	0.14	0.35	0.14	0.34
ANYLIM	1 if has a condition which limits daily living activities	0.48	0.50	0.47	0.50
EXPENDITURE	Total amount paid by self or family	1434	6413	1496	4391

The estimated ATE and ATT are presented in Table 1.11. Both models suggest a negative impact of Medigap on healthcare expenditure. The marginal likelihood results recommend the heteroskedastic model for both samples. The detailed coefficient estimates are not in-

cluded in this dissertation due to space limitations but can be provided upon request.

Table 1.11: Impact of Medigap Policies on Healthcare Expenditure

Year	**Model**		**Mean**	**SD**	**95% CI**	$\ln(\mathbf{m}(\mathbf{y}))$
2018, 2019	Homosk.	ATE	−0.15	0.02	(−0.19, −0.11)	−15073.28
		ATT	0.03	0.02	(−0.01, 0.07)	
	Heterosk.	ATE	−0.19	0.03	(−0.24, −0.13)	−9742.19
		ATT	−0.05	0.03	(−0.11, 0.00)	
2020	Homosk.	ATE	−0.24	0.03	(−0.30, −0.18)	−7655.27
		ATT	−0.06	0.03	(−0.12, −0.01)	
	Heterosk.	ATE	−0.19	0.04	(−0.26, −0.12)	−4941.65
		ATT	−0.05	0.04	(−0.12, 0.02)	

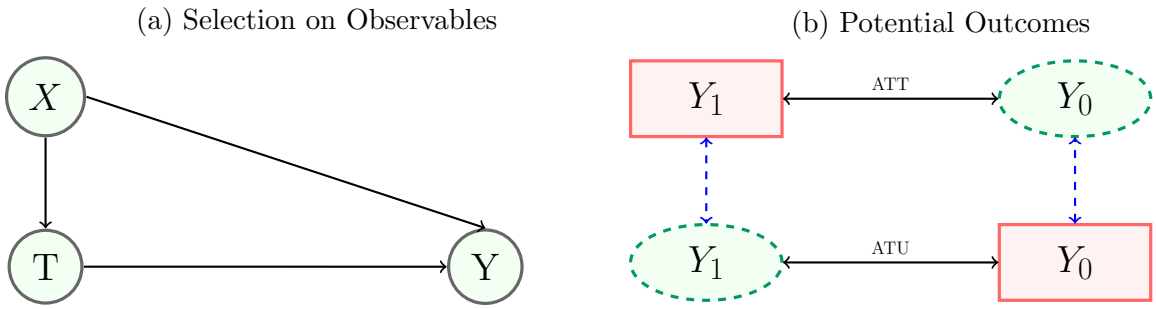
1.4 Propensity Score

Let $p(x) \equiv \Pr(T = 1|x)$ denote the propensity score, which represents the conditional probability of assignment to the treatment given the covariates x (Rosenbaum and Rubin, 1983). Two popular methods that utilize propensity scores to deal with selection bias are propensity score matching and inverse probability of treatment weighting (IPTW). The key ideas behind these models are captured in Figure 1.5. Specifically, Figure 1.5a depicts the key assumption of selection on observables, whereas Figure 1.5b demonstrates that the fundamental problem of estimating treatment effects is caused by the missing counterfactuals. In practice, we have the observed treated and untreated outcomes denoted by the rectangles in Figure 1.5b, whereas the dashed ovals are the unobserved counterfactuals. The idea behind matching observations on the basis of the propensity score is to generate sub-samples from the observed treated and untreated groups that are comparable to one another as a means of uncovering the unobserved counterfactuals and estimating the desired treatment effect.

In this section, we introduce a model to estimate the propensity score with heteroskedasticity. Then we discuss the impacts of ignored heteroskedasticity in two settings: propensity score

matching and IPTW. Both models are employed to assess the treatment effect of COVID-19 vaccination on mental well-being.

Figure 1.5: Selection on Observables and Missing Counterfactuals



Propensity score matching (Rosenbaum and Rubin, 1983) is a popular method for estimating causal treatment effects. The approach is instrumental in mitigating selection bias by leveraging the propensity score as a balancing score that effectively enables the creation of comparable control and treatment groups.

The approach is valid when $x \perp\!\!\!\perp T|p(x)$. To see this, note that by the definition of the propensity score, we have that $f\left(T|p\left(x\right),x\right) = f\left(T|p\left(x\right)\right)$, whereby $f\left(x|p\left(x\right),T\right) = \frac{f(T|p(x),x)f(x|p(x))}{f(T|p(x))} = f\left(x|p\left(x\right)\right)$. In this sense, conditioning on the propensity score generates "balanced" samples of treated and untreated units with similar characteristics x. Crucially, however, proper specification of the propensity score is required for the theory to hold, so that the search for a $p(x)$ that is supported by the data serves as the motivation for our study, especially as it relates to possibly omitted heteroskedasticity.

Researchers employ inverse probability of treatment weighting (IPTW) to counteract non-randomization challenges in observational studies (Rosenbaum, 1987). Successful application of this model necessitates accurate specification of the propensity score (Chesnaye et al., 2022), thereby highlighting the essential inclusion of heteroskedasticity in propensity score estimation. We assign weights to individual observations by taking the inverse of the probability associated with their respective actual treatment status. In other words, we can

calculate the average treatment effect as

$$\hat{ATE} = \frac{1}{n}\sum_{i=1}^{n}\frac{T_i Y_i}{p(x_i)} - \frac{1}{n}\sum_{i=1}^{n}\frac{(1-T_i)Y_i}{1-p(x_i)}.$$

To study this issue, we employ a heteroskedastic model for the propensity score. Owing to the nonlinearity of the setting, Jensen's inequality implies that erroneously omitting heteroskedasticity will impact the bias and consistency properties of estimators and can not be dealt with by simply adjusting the standard errors. For $i = 1, \ldots, n$, the heteroskedastic probit model is specified as

$$T_i = \mathbb{1}\{T_i^* > 0\} = \mathbb{1}\{x_i'\beta + \nu_i > 0\}, \qquad \nu_i \sim N\left(0, \sigma_i^2\right). \tag{1.12}$$

We impose the constraint that the variance of ν_i equals 1 in the homoskedastic model for identification purpose. In the heteroskedastic model, we assume that $var(\nu_i) = \exp(z_i'\gamma)$ and for identification z_i does not include a constant term. We specify the prior distributions $\beta \sim N(b_0, B_0)$ and $\gamma \sim N(\gamma_0, \Gamma_0)$. Algorithms 5 and 6 provide details on the propensity score estimation following the frameworks provided Albert and Chib (1993) and Gu et al. (2009), which also allow for possible extensions to settings with heterogeneity or categorical treatments.

Algorithm 5 (Bayesian Propensity Score Estimation with Homoskedasticity)

---

(1) Sample $\beta \sim N\left(\hat{b}, \hat{B}\right)$, where $\hat{b} = \hat{B}\left(B_0^{-1}b_0 + \sum_{i\in N} x_i T_i^*\right)$, and $\hat{B} = \left(B_0^{-1} + \sum_{i\in N} x_i x_i'\right)^{-1}$.

(2) Sample $T_i^* \sim TN_{\mathcal{B}_i}(x_i'\beta, 1)$, where $\mathcal{B}_i = (-\infty, 0]$ if $T_i = 0$, and $\mathcal{B}_i = (0, \infty)$ if $T_i = 1$.

An estimate of the marginal likelihood for the homoskedastic and heteroskedastic models is obtained as a straightforward special case of the approach presented in Section 1.2.1.

Algorithm 6 (Bayesian Propensity Score Estimation with Heteroskedasticity)

---

(1) Sample $\beta \sim N\left(\hat{b}, \hat{B}\right)$, where $\hat{b} = \hat{B}\left(B_0^{-1}b_0 + \sum_i x_i \exp(z_i'\gamma)^{-1}T_i^*\right)$, and $\hat{B} = \left(B_0^{-1} + \sum_i x_i \exp(z_i'\gamma)^{-1}x_i'\right)^{-1}$.

(2) Sample $T_i^* \sim TN_{\mathcal{B}_i}\left(x_i'\beta, \exp(z_i'\gamma)\right)$, where $\mathcal{B}_i = (-\infty, 0]$ if $T_i = 0$, and $\mathcal{B}_i = (0, \infty)$ if $T_i = 1$.

(3) Sample $[\gamma|\beta, T^*]$ using an MH step by drawing a proposal value $\gamma^\dagger \sim q\left(\gamma|\hat{\gamma}_T, V_T\right)$, where $e_i = T_i^* - x_i'\beta$ and $\hat{\gamma}$ and V are defined similarly to (1.5) using the current value of γ_j and T_i^*. Also use $\gamma_T^\dagger$ in equation (1.5) to produce $\hat{\gamma}_j^\dagger$ and accept the proposed $\gamma_T^\dagger$ with probability

$$\alpha = \min\left\{1, \frac{f\left(T_i^*|\beta, \gamma_T^\dagger\right)\pi\left(\gamma^\dagger|\gamma_0, \Gamma_0\right)q\left(\gamma|\hat{\gamma}^\dagger, V\right)}{f(T_i^*|\beta, \gamma)\pi\left(\gamma|\gamma_0, \Gamma_0\right)q\left(\gamma^\dagger|\hat{\gamma}, V\right)}\right\}.$$

otherwise the current value γ is repeated in the next MCMC iteration.

1.4.1 Simulation Study

In this section, we performed simulation studies to test the effectiveness of the MCMC and marginal likelihood algorithms of the propensity score estimation, and to study the impact of ignored heteroskedasticity in both the propensity score matching and IPTW settings.

1.4.1.1 Simulation: Propensity Score Matching

We illustrate that neglected heteroskedasticity can lead to the emergence of imbalanced samples. The specification of a model with heteroskedasticity is one step in addressing misspecification in addition to other possible approaches that can be taken, such as considering the problem of variable selection or misspecification of the mean function.

The simulation study is based on the data in Dehejia and Wahba (1999), which comes

from the National Supported Work Demonstration (NSW) and the panel study of income dynamics (PSID). The treatment T is the NSW participation. We believe that the variables age (age), years of education ($educ$), high school degree status ($nodegree$), race ($Black$ or $Hispanic$), marital status ($married$), real earnings in 1975 ($RE75$), and real earnings in 1974 ($RE94$) will affect the outcome variable of interest. There are 185 observations in the treatment group and 2490 observations in the control group. We assume that treatment assignment is generated as

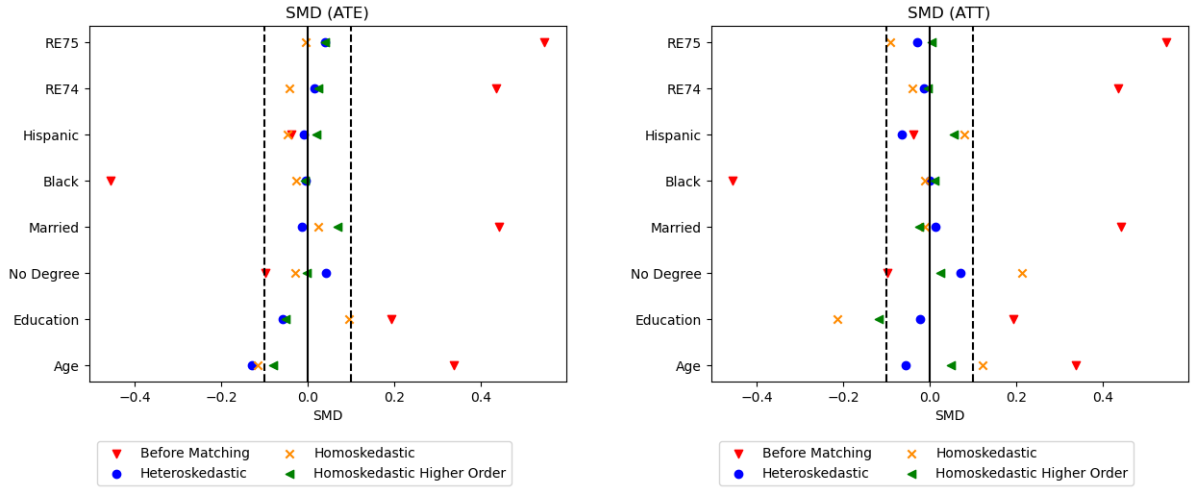
$$T_i = \{-2 - 0.17age_i - 0.001educ_i + 0.3744nodegree_i - 0.9630married_i + 1.2285black_i + 1.219hispanic_i - 0.000005RE74_i - 0.0001RE74_i + \nu_i > 0\}, \quad \nu_i \sim N(0, age_i).$$

The standardized mean difference (SMD) calculated as $\frac{\bar{X}_T - \bar{X}_C}{\sqrt{(S_T^2 + S_C^2)/2}}$, where $\bar{X}_T$ and $\bar{X}_C$ are the sample averages, and S_T and S_C are the standard deviations for the treatment and control groups, respectively, is often used as a balance measure (Rosenbaum and Rubin, 1985; Thoemmes, 2012), with SMD exceeding 0.1 being considered as a sign of imbalance (Zhang et al., 2019). We employ nearest neighbor matching with replacement with a radius of 0.2 times the standard deviation of the estimated propensity score (Austin, 2011; Chaudhuri and Howley, 2022).

Empirical researchers typically proceed by incorporating higher-order and interaction terms to improve the balance of the matched samples if it failed in the beginning (Dehejia and Wahba, 1999; Caliendo and Kopeinig, 2008). Thus, we consider three models in this section: the correctly specified heteroskedastic model, the homoskedastic model with all covariates, and the extended homoskedastic model incorporating all covariates, age squared (age^2), and interactions between age and other covariates. Figure 1.6 shows the SMD before and after matching. Before matching, the covariates are imbalanced. The heteroskedastic model effectively enhances balance within both the ATE and ATT samples. In this example, the homoskedastic model falls short of achieving balance in the ATT estimation sample.

However, by including higher order and interaction terms, the balance is improved.

Figure 1.6: Balance Score Results (SMD)



(a) SMD (ATE)

(b) SMD (ATT)

1.4.1.2 Simulation: Inverse Probability of Treatment Weighting

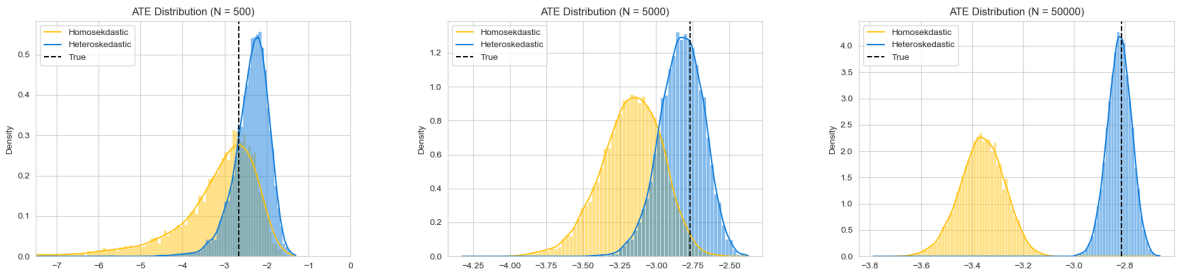
In this section, we illustrate that neglected heteroskedasticity can lead to biased and inconsistent treatment effect estimates. We generate simulated data from the model described in (1.12), where $\beta = (0.9, 1.2, -1.2)'$ and $\gamma = 0.6$. The covariates x_i include a constant term and two variables sampled from independent truncated normal distributions with a range of $(1, +\infty)$, having a mean of 1 and a variance of 1. The covariate z_i comprises the second column of x_i.

Table 1.12 summarizes the estimated ATE, and Figure 1.7 depicts the histogram of the estimated ATE. Marginal likelihood results consistently favor the heteroskedastic model in all scenarios. These findings demonstrated that the ignored heteroskedasticity can lead to biased and inconsistent ATE estimates. When the sample size increases, the heteroskedastic model accurately represents the ATE, whereas the homoskedastic model estimates fail to capture the true effect.

Table 1.12: Average Treatment Effect Results Using IPTW

n	Model	True	Mean	SD	95% CI	ln(m(y))
500	Heterosk.	−2.66	−2.36	0.46	(−3.35, −1.70)	−247.68
	Homosk.	−2.66	−3.24	1.21	(−6.25, −1.96)	−252.71
5000	Heterosk.	−2.77	−2.83	0.15	(−3.14, −2.56)	−2421.02
	Homosk.	−2.77	−3.17	0.21	(−3.61, −2.80)	−2434.46
50000	Heterosk.	−2.81	−2.82	0.05	(−2.92, −2.73)	−23005.06
	Homosk.	−2.81	−3.36	0.09	(−3.54, −3.20)	−23116.63

Figure 1.7: Treatment Effect Distribution Using IPTW



(a) n = 500 (b) n = 5000 (c) n = 50000

1.4.2 Application: The Effect of COVID-19 Vaccination on Mental Well-Being

Chaudhuri and Howley (2022) evaluate the impact of COVID-19 vaccination on mental health. The treatment variable is an indicator of whether subject i received any dose of a COVID-19 vaccine. This is a sample of waves 7 and 8 of the COVID-19 survey by the UK Household Longitudinal Study (University of Essex, Institute for Social and Economic Research, 2021). This survey includes the vaccination, demographic and mental health information of $21,985$ survey participants. The outcome variable in this study is assessed using the GHQ-12 questionnaire, which is designed to evaluate an individual's mental health condition through a series of 12 questions. Each question in the GHQ-12 is rated on a four-point scale. The resulting GHQ scores can range from 0 to 36. In the context of this

particular sample, we follow Chaudhuri and Howley (2022) and reverse the GHQ scores to improve the interpretability, such that the score is directly proportional to the level of mental well-being in the evaluated individuals. Summary statistics of key variables are presented in Table 1.13, and analysis by propensity score matching and IPTW methods are presented in Sections 1.4.2.1 and 1.4.2.2.

Table 1.13: Summary Statistics, COVID-19 Vaccination Data

	Control ($n_0 = 12423$)		**Treated** ($n_1 = 9562$)	
Variables	**Mean**	**SD**	**Mean**	**SD**
GHQ-12	23.14	6.15	24.04	5.64
Age	49.19	15.64	61.67	13.84
Born in UK	0.87	0.34	0.90	0.30
Clinically Vulnerable	0.34	0.47	0.56	0.50
Male	0.42	0.49	0.41	0.49
Key Worker	0.25	0.43	0.25	0.43
Couple	0.69	0.46	0.73	0.44
Willingness to Take Vaccine	0.91	0.29	0.95	0.22

1.4.2.1 Application: Propensity Score Matching

We estimated treatment effects using four models. The first one is a heteroskedastic model with variance $var(\nu_i) = \exp(z_i'\gamma)$, while the second model is more parsimonious with variance $var(\nu_i) = age_i$. The third model is a homoskedastic model with all the covariates, and the fourth model incorporates age_i^2 as well as interaction terms between age and the other covariates.

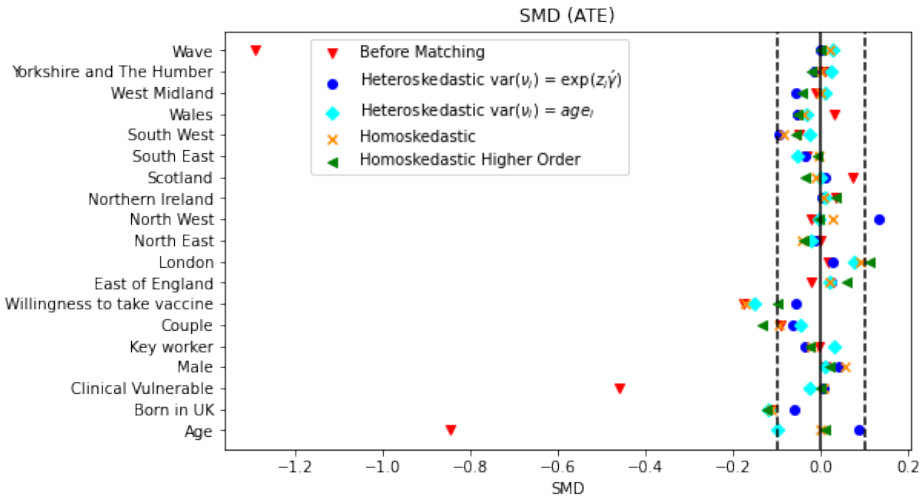
The estimated impact of COVID-19 vaccination on mental well-being is presented in Table 1.14. The marginal likelihood results suggest that the heteroskedastic model with $var(\nu_i) = \exp(z_i'\gamma)$ fits the data best. These results suggest that COVID-19 vaccination is expected to improve mental health. Figure 1.8 and 1.9 show the SMD before and after matching. The figures show that for the ATE samples, all the models can improve the balance. For the

Table 1.14: Impact of COVID-19 Vaccination on Mental Health (PSM)

	ATE			ATT			
Model	**Mean**	**SD**	**95% CI**	**Mean**	**SD**	**95% CI**	$\ln(\mathbf{m}(\mathbf{y}))$
Heterosk. ($\exp(z_i'\gamma)$)	1.20	0.22	$(0.76, 1.64)$	2.59	0.47	$(1.68, 3.52)$	-7687.86
Heterosk. (age_i^2)	0.52	0.12	$(0.30, 0.77)$	1.36	0.22	$(0.92, 1.80)$	-9337.68
Homosk.	0.39	0.12	$(0.15, 0.61)$	1.15	0.22	$(0.72, 1.60)$	-8753.30
Homosk. (Higher-order)	0.36	0.13	$(0.13, 0.64)$	0.94	0.23	$(0.51, 1.46)$	-8030.50

ATT samples, the heteroskedastic model with $var(\nu_i) = \exp(z_i'\gamma)$ performs better than the alternatives.

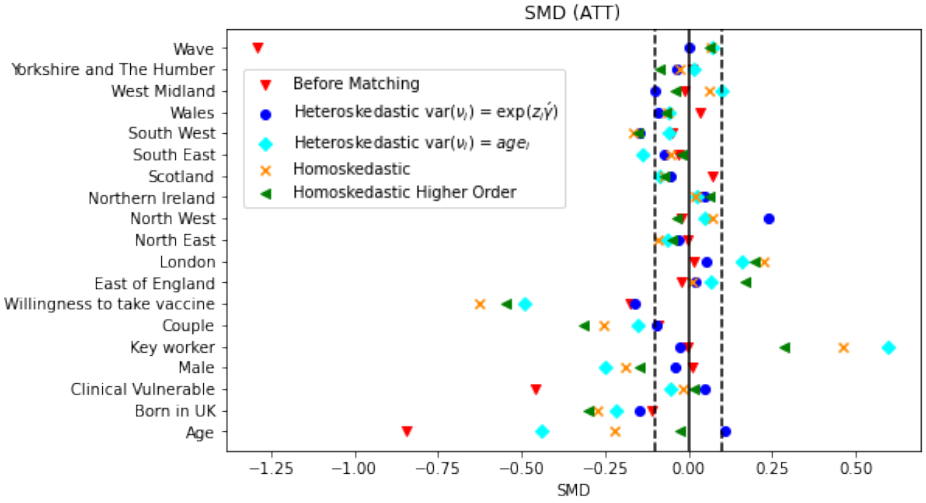
Figure 1.8: COVID-19 Vaccination SMD (ATE Sample)



1.4.2.2 Application: Inverse Probability of Treatment Weighting

We computed the Average Treatment Effect (ATE) utilizing both a heteroskedastic model and a homoskedastic model to assess the impact of COVID-19 vaccination on mental well-being, as outlined in Table 1.15. The marginal likelihood strongly supports the heteroskedastic model, and its estimate closely aligns with the findings from propensity score matching.

Figure 1.9: COVID-19 Vaccination SMD (ATT Sample)



However, the ATE distribution's standard deviation, significantly larger than that in the propensity score matching section, renders the impact of COVID-19 vaccination less precise than the propensity score matching method. The table also demonstrates a well-known problem with inverse probability estimators, namely their large variance because there is no guarantee that the inverse probability must be bounded. Moreover, the estimated treatment effect in the homoskedastic model lacks interpretability, because it lies entirely outside the range of the mental health score $[0, 36]$. This underscores the importance of heteroskedasticity in this context.

Table 1.15: Impact of COVID-19 Vaccination on Mental Health (IPTW)

Model	**Mean**	**SD**	**95% CI**	$\ln(\mathbf{m}(\mathbf{y}))$
Heterosk.	3.45	1.85	$(-0.33, 6.98)$	-7633.95
Homosk.	54.23	8.34	$(40.25, 72.39)$	-8753.28

1.5 Conclusion

This chapter has studied the impact of heteroskedasticity in regression discontinuity designs, potential outcome models, and propensity score matching. Because of the nonlinearities in these contexts, the question of whether heteroskedasticity is present has to be addressed directly, as it can lead to bias and inconsistency with consequences can not be handled by correcting the standard errors. In our Bayesian context, we treat the presence of heteroskedasticity as a question of model uncertainty. On the computational side, we develop new computationally efficient simulation-based estimation algorithms tailored to each setting and discuss their implementation in computing marginal likelihoods to enable formal model comparison. Moreover, we propose an approach for reducing the number of reduced MCMC runs required for marginal likelihood estimation in settings with multiple parameter blocks.

Simulation studies have been provided in order to evaluate the empirical consequences of omitted heteroskedasticity, assess the performance of the proposed estimation algorithms, and validate the proposed model comparison techniques. Our investigation has revealed that when non-linearity is pronounced, ignoring heteroskedasticity can result in biased estimates of treatment effects. We also find that the proposed MCMC methods perform well and can recover the true parameters and models used in generating the data.

To assess the practical applicability and relevance of our methods, this chapter has devoted considerable attention to several applications. In particular, we have explored the impact of academic probation on students' academic performance, the effects of Medigap policies on out-of-pocket healthcare expenditures, and the influence of COVID-19 vaccination on mental well-being. RDD results suggest that academic probation improves subsequent semester GPA, while exhibiting no discernible impact on graduation status. Using a Rubin causal model (Roy-type model) in our second application, we find that Medigap policies are expected to reduce out-of-pocket healthcare expenditures. Finally, results from propensity

score matching and inverse probability of treatment weighting indicate that COVID-19 vaccination improved the mental well-being of vaccine recipients in the UK. Based on model comparisons in each application, we found that heteroskedastic models were favored in all settings. The results emphasize the importance of allowing for heteroskedasticity in observational observational studies of causal effects and demonstrate that the presence of heteroskedasticity can be uncovered through model comparisons. While our analysis has primarily centered on the effects of heteroskedasticity, we believe that other concerns such as sample selection or endogeneity may also be present in many settings. We intend to study their impact, as well as their interactions with heteroskedasticity, on treatment effect estimation in future work.

Chapter 2

Bayesian Analysis of a Self-selection Model with Multiple Outcomes

We develop a Bayesian treatment model that incorporates self-selection and accommodates multiple outcomes. We discuss the estimation of marginal likelihood or formal model comparison. To validate our approach, we test the algorithm and model comparison techniques through simulation results. Subsequently, we employ our model on two datasets, enabling us to analyze the influence of insurance on healthcare utilization. Specifically, we estimate the impact of Medigap policies on healthcare expenditure using the 1987 National Medical Expenditure Survey (NMES) data and the impact of different types of private insurance on healthcare utilization using the 1996 Medical Expenditure Panel Survey (MEPS) data. Our results reveal weak evidence supporting selection bias in both applications.

2.1 Introduction

This chapter introduces a parametric self-selection Bayesian model featuring a binary treatment and two distinct outcome variables—one count and the other continuous. Allowing for endogenous selection, we are interested in studying the treatment impact on the conditional mean of the outcome variables. By leveraging this model, we conducted an empirical investigation into the influence of private insurance on healthcare expenditures and the frequency of physician office visits across two datasets.

Sample selection arises when the observed sample does not accurately represent the broader population of interest, introducing bias into estimators. Heckman (1979) highlighted the consequential inconsistency and bias resulting from ignoring sample selection and introduced Heckman correction as a remedy. In the context of treatment models, we typically only observe treated outcomes for individuals in the treated group and untreated outcomes for those in the untreated group, potentially leading to sample selection bias if selection decisions are non-random. Addressing this challenge, Chib, Greenberg and Jeliazkov (2009) pioneered a Bayesian model capable of analyzing data afflicted by both sample selection and endogeneity issues.

Many papers discussed possible selection concerns in the health insurance industry. Cutler and Zeckhauser (1998) examined evidence suggestive of adverse selection, shedding light on its implications. Keane and Stavrunova (2016) advanced a model aimed at estimating moral hazard and selection dynamics within the Medigap market, revealing a weak adverse selection alongside notable moral hazard effects, especially among individuals with better health conditions. Sapelli and Vial (2003) studied the self-selection using the Chilean physician visits data and hospital days data. They found some evidence supporting the existence of self-selection. Nghiem and Graves (2019) endeavored to estimate the impact of moral hazard and selection biases within Australia's private insurance landscape, unearthing evi-

dence indicative of advantageous selection tendencies, wherein risk-averse individuals exhibit a heightened propensity to procure health insurance coverage.

On the other hand, Reschovsky, Kemper and Tu (2000) posited that there exists minimal disparity in hospital utilization across various insurance plan types. Cardon and Hendel (2001) found that there is no discernible evidence of asymmetric information significantly influencing insurance plan choices. Munkin and Trivedi (2003) found weak evidence of self-selection in their model estimating the impact of health insurance on healthcare utilization.

We expand upon the potential outcome methodology employed in Munkin and Trivedi (2003) by relaxing the assumption of constant treatment effects. This enhancement allows for a more nuanced exploration of treatment effects. Incorporating insights from Bayesian research by Albert and Chib (1993), Chib, Greenberg and Winkelmann (1998) and Chib, Greenberg and Jeliazkov (2009), we use the Markov chain Monte Carlo (MCMC) method to estimate the parameters. Our methodology was applied to two datasets to investigate the influence of public or private health insurance on healthcare expenditure and the frequency of doctor's office visits. Specifically, we analyzed the 1996 Medical Expenditure Panel Survey (MEPS) and the 1987 National Medical Expenditure Survey (MEPS) data to gain insights into these impacts. Additionally, we conducted estimations using three alternative parsimonious models to provide a comprehensive comparative analysis. To facilitate model comparison, we computed the marginal likelihood for each model.

The remainder of the chapter is structured as follows. In Section 2.2, we delineate the model, its corresponding estimation algorithm, and the model comparison technique. In Section 2.3, we present the simulation results. Section 2.4 discusses the empirical findings. Finally, Section 2.5 concludes the paper.

2.2 Model

This section considers a potential outcome framework with self-selection for estimating the treatment effect for multiple outcomes. The binary treatment variable is represented by D_i, where $D_i = 1$ if subject i receives treatment and $D_i = 0$ if subject i is in the control group. For each i in the sample, we denote the potential outcomes for the count and continuous outcome variables as y_{j1i} and y_{j2i} respectively, where $j = 0, 1$ signifies the treatment status. We assume that

$$y_{j1i} \sim \text{Poisson}(\mu_{ji}), \quad j = 0, 1.$$

For each i in the sample, the observed response is expressed as

$$y_{1i} = y_{01i} + (y_{11i} - y_{01i})\, D_i,$$
$$y_{2i} = y_{02i} + (y_{12i} - y_{02i})\, D_i.$$

We assume that $D_i = \mathbb{1}\{d_i^* \geq 0\}$, where d_i^* represents a latent variable that determines the values of D_i. Additionally, we let x_{d1i} to denote the covariates for y_{d1i}, x_{d2i} to denote the covariates for y_{d2i}, where $d = 0, 1$. Lastly, we use x_{di} to denote the covariates for d_i^*.

The model can be represented as

$$g_i = X_i\beta + \varepsilon_i, \quad \varepsilon_i \sim N(0, \Omega), \tag{2.1}$$

where

$$g_i = \left(d_i^*, \ln\left(\mu_{0i}\right), y_{02i}, \ln\left(\mu_{1i}\right), y_{12i}\right)', \quad X_i = \begin{pmatrix} x'_{di} & 0 & 0 & 0 & 0 \\ 0 & x'_{01i} & 0 & 0 & 0 \\ 0 & 0 & x'_{02i} & 0 & 0 \\ 0 & 0 & 0 & x'_{11i} & 0 \\ 0 & 0 & 0 & 0 & x'_{12i} \end{pmatrix},$$

$$\beta = \left(\beta'_d, \beta'_{01}, \beta'_{02}, \beta'_{11}, \beta'_{12}\right)', \quad \varepsilon_i = \left(\varepsilon_{di}, \varepsilon_{01i}, \varepsilon_{02i}, \varepsilon_{11i}, \varepsilon_{12i}\right)'.$$

If $d_i^* \geq 0$, we observe $\left(\ln\left(\mu_{1i}\right), y_{12i}\right)$; otherwise, we observe $\left(\ln\left(\mu_{01}\right), y_{02i}\right)$. We define the vectors $g_{0i} = \left(d_i^*, \ln\left(\mu_{0i}\right), y_{02i}\right)'$ and $g_{1i} = \left(d_i^*, \ln(\mu_{1i}), y_{12i}\right)'$. The covariates matrix is given by

$$X_{1i} = \begin{pmatrix} x'_{di} & 0 & 0 \\ 0 & x'_{11i} & 0 \\ 0 & 0 & x'_{12i} \end{pmatrix}, \quad X_{0i} = \begin{pmatrix} x'_{di} & 0 & 0 \\ 0 & x'_{02i} & 0 \\ 0 & 0 & x'_{12i} \end{pmatrix}.$$

The covariance matrix is represented as

$$\Omega = \begin{pmatrix} 1 & \omega_{d01} & \omega_{d02} & \omega_{d11} & \omega_{d12} \\ \omega_{01d} & \omega_{01} & \omega_{012} & \omega_{0111} & \omega_{0112} \\ \omega_{02d} & \omega_{021} & \omega_{02} & \omega_{0211} & \omega_{0212} \\ \omega_{11d} & \omega_{1101} & \omega_{1102} & \omega_{11} & \omega_{112} \\ \omega_{12d} & \omega_{1201} & \omega_{1202} & \omega_{121} & \omega_{12} \end{pmatrix}.$$

Due to missing data, we cannot identify Ω_{0111}, Ω_{0112}, Ω_{0211}, and Ω_{0212}. To address this identification issue, we assume that $\Omega_d = 1$.

Thus the covariance matrix that can be identified is expressed as

$$\Omega = \begin{pmatrix} 1 & \omega_{d01} & \omega_{d02} & \omega_{d11} & \omega_{d12} \\ \omega_{01d} & \omega_{01} & \omega_{012} & . & . \\ \omega_{02d} & \omega_{021} & \omega_{02} & . & . \\ \omega_{11d} & . & . & \omega_{11} & \omega_{112} \\ \omega_{12d} & . & . & \omega_{121} & \omega_{12} \end{pmatrix}.$$

We introduce the following notation

$$\Omega_0 = \begin{pmatrix} 1 & \omega_{d01} & \omega_{d02} \\ \omega_{01d} & \omega_{01} & \omega_{012} \\ \omega_{02d} & \omega_{021} & \omega_{02} \end{pmatrix} = \begin{pmatrix} 1 & \Omega_{012} \\ \Omega_{021} & \underset{(2\times 2)}{\Omega_{011}} \end{pmatrix}, \quad \Omega_1 = \begin{pmatrix} 1 & \omega_{d11} & \omega_{d12} \\ \omega_{d11} & \omega_{11} & \omega_{112} \\ \omega_{12d} & \omega_{121} & \omega_{12} \end{pmatrix} = \begin{pmatrix} 1 & \Omega_{112} \\ \Omega_{121} & \underset{(2\times 2)}{\Omega_{111}} \end{pmatrix},$$

$$J_0 = \begin{pmatrix} I & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 \end{pmatrix}, \quad J_1 = \begin{pmatrix} I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & I \end{pmatrix},$$

$$\Omega_{22\cdot 0} = \Omega_{011} - \Omega_{021}\Omega_{012}, \quad \Omega_{22\cdot 1} = \Omega_{111} - \Omega_{121}\Omega_{112}.$$

Given the unit restriction, we directly handle the quantities $\Omega_{22\cdot 0}$, $\Omega_{22\cdot 1}$, Ω_{012}, and Ω_{112}, from which we subsequently recover Ω_0 and Ω_1 (Munkin and Trivedi, 2003; Chib, Greenberg and Jeliazkov, 2009; Vossmeyer, 2016).

Thus for i in the control group, we have

$$g_{0i} \propto |\Omega_0|^{-1/2} \exp\left(-\frac{1}{2}\left(g_{0i} - X_{0i}J_0\beta\right)'\Omega_0^{-1}\left(g_{0i} - X_{0i}J_0\beta\right)\right),$$

and for i in the treated group, we have

$$g_{1i} \propto |\Omega_1|^{-1/2} \exp\left(-\frac{1}{2}\left(g_{1i} - X_{1i}J_1\beta\right)'\Omega_1^{-1}\left(g_{1i} - X_{1i}J_1\beta\right)\right).$$

The complete data density function is given by

$$\begin{aligned} f(y_1, y_2, d^*, \ln(\mu)|\beta, \Omega_0, \Omega_1) &= f(\ln(\mu), y_2, d^*|\theta) f(y_1|\ln(\mu)) \\ &= \left[\prod_{i:D_i=1} f_N(g_{i1}|\beta, \Omega_1) f\big(y_{11i}|\ln(\mu_{1i})\big) \mathbb{1}\{d_i^* \geq 0\}\right] \\ &\qquad \times \left[\prod_{i:D_i=0} f(g_{0i}|\beta, \Omega_0) f\big(y_{01i}|\ln(\mu_{0i})\big) \mathbb{1}\{d_i^* < 0\}\right]. \end{aligned}$$

The prior distribution for β is specified as $\beta \sim N(b_0, B_0)$. We let $\Omega_{22\cdot 0} \sim IW(r_0, R_0)$, $\Omega_{22\cdot 1} \sim IW(r_1, R_1)$, $\Omega_{012}|\Omega_{22\cdot 0} \sim N(q_0, \Omega_{22\cdot 0})$, and $\Omega_{112}|\Omega_{22\cdot 1} \sim N(q_1, \Omega_{22\cdot 1})$.

In our model, the average treatment effect (ATE) and the average treatment effect on the treated (ATT) for outcome Y_1 and Y_2 are defined as

$$\begin{aligned} ATE_1 &= E(Y_{11i} - Y_{01i}) = E(x'_{11i}\beta_{11} - x'_{01i}\beta_{01}), \\ ATE_2 &= E(Y_{12i} - Y_{02i}) = E(x'_{12i}\beta_{12} - x'_{02i}\beta_{02}), \\ ATT_1 &= E(Y_{11i} - Y_{01i}|D_i = 1) = E(x'_{11i}\beta_{11} - x'_{11i}\beta_{01}), \\ ATT_2 &= E(Y_{12i} - Y_{02i}|D_i = 1) = E(x'_{12i}\beta_{12} - x'_{12i}\beta_{02}). \end{aligned}$$

These quantities can be estimated using the MCMC output.

2.2.1 Markov chain Monte Carlo (MCMC) Algorithm

In this section, we delve into the Markov chain Monte Carlo (MCMC) simulation algorithm tailored for our model. We derive the conditional distribution for each parameter and provide a comprehensive discussion of the sampling methodology employed.

2.2.1.1 Sampling β

The posterior distribution for β is $N(b, B)$, where

$$b = B\left(B_0^{-1}b_0 + \sum_{i\in N_0} J_0' X_{0i}' \Omega_0^{-1} g_{0i} + \sum_{i\in N_1} J_1' X_{1i}' \Omega_1^{-1} g_{1i}\right),$$

and

$$B = \left(B_0^{-1} + \sum_{i\in N_0} J_0' X_{0i}' \Omega_0^{-1} X_{0i} J_0 + \sum_{i\in N_1} J_1' X_{i1}' \Omega_1^{-1} X_{1i} J_1\right)^{-1}.$$

2.2.1.2 Sampling d_i^*

We sample $d_i^* \sim TN_{\mathcal{B}_i}(\mu_{dji}, \Omega_{dj})$, where $\mathcal{B}_i = (-\infty, 0)$ if $D_i = 0$, and $\mathcal{B}_i = [0, +\infty)$ if $D_i = 1$, $\mu_{dji} = x_{di}'\beta_d + \Omega_{j12}\left(\Omega_{j11}\right)^{-1}\left(\ln(\mu_{ji}) - x_{j1i}'\beta_{j1}, y_{j2i} - x_{j2i}'\beta_{j2}\right)'$, and $\Omega_{dj} = 1 - \Omega_{j12}\left(\Omega_{j11}\right)^{-1}\Omega_{j21}$, $j = 0, 1$.

2.2.1.3 Sampling $\ln(\mu_{it})$

The conditional distribution of $\ln(\mu_{ji})$ given y_{j1i}, y_{j2i}, d_i^* and all the parameters is

$$(\ln(\mu_{ji}) | y_{j2i}, d_i^*, \beta, \Omega_j) \sim N\left(\mu_{1|2,ji}, \sigma_{1|2,j}\right),$$

where $\mu_{1|2,ji} = x'_{j1i}\beta_{j1} + \tilde{\Omega}_{j12}\left(\tilde{\Omega}_{j22}\right)^{-1}\left(d_i^* - x'_{di}\beta_d, y_{j2i} - x'_{j2i}\beta_{j2}\right)'$, $\tilde{\Omega}_{j12} = (\omega_{j1d}, \omega_{j12})$, $\sigma_{1|2,j} = \omega_{j1} - \tilde{\Omega}_{j12}(\tilde{\Omega}_{j22})^{-1}\tilde{\Omega}_{j21}$, and $\tilde{\Omega}_{j22} = \begin{pmatrix} 1 & \omega_{dj2} \\ \omega_{j2d} & \omega_{j2} \end{pmatrix}$, $j = 0, 1$. Thus the posterior distribution for $\ln(\mu_{it})$ is specified as

$$\begin{aligned}\pi\left(\ln(\mu_{ji})|y_{j1i}, y_{j2i}, d_i^*, \beta, \Omega_j\right) &\propto f_N\left(\mu_{1|2,ji}, \sigma_{1|2,j}\right) f\left(y_{j1i}|\ln\left(\mu_{ji}\right)\right) \\ &= f_N\left(\mu_{1|2,ji}, \sigma_{1|2,j}\right)\frac{\mu^{y_{j1i}}e^{-\mu_{ji}}}{y_{j1i}!}.\end{aligned}$$

We use a random walk Metropolis-Hasting algorithm to sample the $\ln\left(\mu_{ji}\right)$. We use $\ln\left(\mu_{ji}\right)^*$ to denote the proposed value. The proposed density is denoted as $q\left(\ln\left(\mu_{ji}\right), \ln\left(\mu_{ji}\right)^*\right) = f_N\left(\ln(\mu_{ji})^*|\ln\left(\mu_{ji}\right), \left(\left(\sigma_{1|2,j}\right)^{-1} + y_{j1i}^{-1}\right)^{-1}\right)$. The acceptance rate is thus defined as

$$\begin{aligned}\alpha\left(\ln\left(\mu_{ji}\right), \ln\left(\mu_{ji}\right)^*\right) &= \min\left\{\frac{\pi\left(\ln\left(\mu_{ji}\right)^*|y_{j1i}, y_{j2i}, \beta, d_i^*, \Omega_j\right) q\left(\ln\left(\mu_{ji}\right), \ln\left(\mu_{ji}^*\right)\right)}{\pi\left(\ln\left(\mu_{ji}\right)|y_{i1t}, y_{i2t}, \beta, z_i, \Omega_j\right) q\left(\ln\left(\mu_{ji}\right)^*, \ln\left(\mu_{ji}\right)\right)}, 1\right\} \\ &= \min\left\{\frac{\pi\left(\ln\left(\mu_{ji}\right)^*|y_{j1i}, y_{j2i}, \beta, d_i^*, \Omega_j\right)}{\pi\left(\ln\left(\mu_{ji}\right)|y_{j1i}, y_{j2i}, \beta, d_i^*, \Omega_j\right)}, 1\right\}.\end{aligned}$$

We generate a random number p from a uniform distribution $U(0, 1)$. If $p \leq \alpha$, the proposed value is accepted; otherwise, it is rejected.

2.2.1.4 Sampling Ω_{j21}

The conditional distribution of Ω_{j21}, $j = 0, 1$, is given by

$$\begin{aligned}\pi\left(\Omega_{j21}|\Omega_{22\cdot j}, \beta, \ln(\mu_{ji}), \ln(y_{j2i}), d_i^*\right) &\propto \pi\left(\Omega_{j21}|\Omega_{22\cdot j}\right)\pi\left(\left(\ln(\mu_{ji}), y_{j2i}\right)'|d_i^*\right) \\ &= f_N\left(q_j, \Omega_{22\cdot j}\right)\prod_{i:D_i=j} f_N\left(\left(\ln\left(\mu_{ji}\right), y_{j2i}\right)'|\mu_{2|3,ji}, \Omega_{22\cdot j}\right),\end{aligned}$$

where $\mu_{2|3,ji} = \left(x'_{j1i}\beta_{j1}, x'_{j2i}\beta_{j2}\right)' + \Omega_{j21}\left(d_i^* - x'_{di}\beta_d\right)$.

Thus, the posterior distribution for Ω_{j21} is specified as $\Omega_{j21} \sim N\left(\hat{q}_j, \hat{\Sigma}_j\right)$, where $\hat{\Sigma}_j = \left(\epsilon'_{jd}\epsilon_{jd}\left(\Omega_{22\cdot j}\right)^{-1} + \left(\Omega_{22\cdot j}\right)^{-1}\right)^{-1}$, $\hat{q} = \hat{\Sigma}\left(\left(\Omega_{22\cdot j}\right)^{-1}\epsilon'_j\epsilon_{jd} + \left(\Omega_{22\cdot j}\right)^{-1}q_j\right)$, $\epsilon_{jd} = d_i^* - X_d\beta_d$, and $\epsilon_j = \left(\ln\left(\mu_j\right) - X_{j1}\beta_{j1}, y_{j2} - X_{j2}\beta_{j2}\right)$.

2.2.1.5 Sampling $\Omega_{22\cdot j}$

We sample $\Omega_{22\cdot j}$, $j = 0, 1$, from the distribution specified as

$$\begin{aligned}
&\pi\left(\Omega_{22\cdot j}|\Omega_{j21}, \beta, \ln(\mu_{ji}), y_{j2i}, d_i^*\right) \propto \pi\left(\Omega_{22\cdot j}\right)\pi\left(\Omega_{j21}|\Omega_{22\cdot j}\right)\pi\left(\left(\ln\left(\mu_{ji}\right), \ln\left(y_{j2i}\right)\right)'|d_i^*\right) \\
&= \pi\left(\Omega_{22\cdot j}\right) f_N\left(q_j, \Omega_{22\cdot j}\right)\prod_{i:D_i=j} f_N\left(\left(\ln\left(\mu_{ji}\right), y_{j2i}\right)'|\mu_{2|3,ji}, \Omega_{22\cdot j}\right) \\
&\propto \left|\Omega_{22\cdot j}\right|^{\frac{r_j+4+n_j}{2}} \times \exp\Bigg(-\frac{1}{2}\Bigg(tr\Bigg[\Big(R_j + \left(\Omega_{j21} - q_j\right)\left(\Omega_{j21} - q_j\right)' \\
&\qquad\qquad + \sum_{i:D_i=j}\left(\epsilon_{ji} - \Omega_{j21}\epsilon_{jdi}\right)\left(\epsilon_{ji} - \Omega_{j12}\epsilon_{jdi}\right)'\Big)\left(\Omega_{22\cdot j}\right)^{-1}\Bigg]\Bigg)\Bigg),
\end{aligned}$$

where n_0 represents the sample size of the control group, and n_1 denotes the sample size of the treated group. Thus, we sample $\Omega_{22\cdot j}$ from a inverse Wishart distribution, defined as $\Omega_{22\cdot j} \sim IW(\hat{r}_j, \hat{R}_j)$, where $\hat{r}_j = r_j + 1 + n_j$, and $\hat{R}_j = R_j + \left(\Omega_{j21} - q_j\right)\left(\Omega_{j21} - q_j\right)' + \sum_{i:D_i=j}\left(\epsilon_{ji} - \Omega_{j21}\epsilon_{jdi}\right)\left(\epsilon_{ji} - \Omega_{j21}\epsilon_{jdi}\right)'$.

2.2.2 Bayesian Model Comparison

Bayesian model comparison offers a systematic approach for comparing multiple competing models. According to Bayes' formula, the posterior probability for a given model, denoted as $\mathcal{M}_c$, is expressed as

$$\pi\left(\mathcal{M}_c|y\right) \propto \pi\left(\mathcal{M}_c\right) m\left(y|\mathcal{M}_c\right),$$

where $\mathcal{M}_c$ signifies the prior probability of model $\mathcal{M}_c$, while $m\left(y|\mathcal{M}_c\right)$ denotes the marginal likelihood.

According to Chib (1995), the marginal likelihood can be estimated at a specific parameter value θ_c^* in the parameter space, as

$$m\left(y|\mathcal{M}_c\right) = \frac{f\left(y|\mathcal{M}_c, \theta_c^*\right)\pi\left(\theta_c^*|\mathcal{M}_c\right)}{\pi\left(\theta_c^*|\mathcal{M}_c, y\right)}.$$

In our model, $\theta_c = \{\beta, \Omega_{22\cdot 0}, \Omega_{012}, \Omega_{22\cdot 1}, \Omega_{112}\} \equiv \{\psi_1, \psi_2, \psi_3, \psi_4, \psi_5\}$. For simplicity, we will omit the notation $\mathcal{M}_c$ in the subsequent discussions. The likelihood in this model can be decomposed as

$$\begin{aligned} f(y|\theta_c^*) = &\left(\prod_{i:D_i=0} f\left(y_{02i}|\theta_c^*\right) f\left(\ln\left(\mu_{0i}\right), d_i^*|y_{02i}, \theta_c\right) f\left(y_{01i}|\ln\left(\mu_{0i}\right)\right) f\left(D_i|d_i^*\right)\right) \\ &\times \left(\prod_{i:D_i=1} f\left(y_{12i}|\theta_c^*\right) f\left(\ln(\mu_{1i}), d_i^*|y_{12i}, \theta_c\right) f\left(y_{11i}|\ln\left(\mu_{1i}\right)\right) f\left(D_i|d_i^*\right)\right), \end{aligned}$$

where $f\left(y_{02i}|\theta_c^*\right)$ and $f\left(y_{12i}|\theta_c^*\right)$ are straightforward to compute, while the remaining terms can be estimated using the importance sampling method. We employ a Student-t distribution with 5 degrees of freedom as the proposal distribution.

The posterior ordinate $\pi\left(\theta_c^*|y\right)$ can be calculated utilizing the Markov Chain's invariant condition (Ritter and Tanner, 1992; Jeliazkov and Lee, 2010), expressed as

$$\pi\left(\theta_c^*|y\right) = E\left(K\left(\theta_c, \theta_c^*\right)|y, \xi\right),$$

where xi denotes the set of latent data, and $K\left(\cdot\right)$ represents the Gibbs transition kernel defined as

$$K\left(\theta_c, \theta_c^*|y\right) = \pi_{r=1}^5 \pi\left(\psi_r^*|y, \{\psi_s^*\}_{s<r}, \{\psi_s\}_{s>r}, \xi\right)$$

where (ψ, ξ) are obtained in the main MCMC run.

"In this paper, we contrast our primary model, designated as $\mathcal{M}_0$, with several more parsimonious alternatives: $\mathcal{M}_1$, which is the baseline model $\mathcal{M}_0$ with constraints $\Omega_{012} = \Omega_{112} =$

$(0,0)$, M_2, a model featuring a constant treatment effect, and M_3, an extension of M_2 where $\Omega_{12} = (0,0)$ is imposed.

It's important to note that M_2 offers a more parsimonious version of M_0. The specific model details for M_2 are provided in B.1.

2.3 Simulation

In this section, we conduct a simulation study to evaluate both the performance of the MCMC algorithm and the effectiveness of the proposed model comparison approach. We randomly generated $10,000$ observations from the model $\mathcal{M}_0$ using Equation 2.1, and estimated the model using the method introduced in the previous session. The simulation parameters were selected for illustrative purposes, aiming to demonstrate the methods' effectiveness in estimating the true parameters. The parameters are specified as

$$\Omega = \begin{pmatrix} 1 & 0.55 & 0.34 & 0.46 & -0.19 \\ 0.55 & 1.04 & -0.13 & -0.05 & -0.56 \\ 0.34 & -0.13 & 0.32 & 0.30 & 0.17 \\ 0.46 & -0.05 & 0.30 & 0.71 & -0.12 \\ -0.19 & -0.56 & 0.17 & -0.12 & 0.64 \end{pmatrix},$$

$\beta_d = (-0.5, 1, -0.5)'$, $\beta_{01} = (-0.5, 0.5)'$, $\beta_{11} = (0.5, -0.5)'$, $\beta_{02} = (1, -1)'$ and $\beta_{12} = (-1, 1)'$. The covariates are $x_{01i} = (1, \nu_{1i})'$, $x_{11i} = x_{02i} = x_{12i} = x_{01i}$, and $x_{di} = (1, \nu_{1i}, \nu_{2i})'$, where ν_{1i} and ν_{2i} are sampled from two independent standard normal distribution.

We estimated the generated data using $10,000$ iterations with a burn-in period of 1000 iterations. We reported the mean and standard deviation of the posterior distribution for the parameters, along with the 95% credible interval. The 95% credible intervals are derived from

Table 2.1: Simulation Parameters Estimates Using $\mathcal{M}_0$ and $\mathcal{M}_1$

		$\mathcal{M}_0$			$\mathcal{M}_1$		
	True	Mean	SD	95% CI	Mean	SD	95% CI
CONST	-0.5	-0.49	0.02	$(-0.52, -0.46)$	-0.49	0.02	$(0.52, -0.46)$
ν_1	1	0.98	0.02	$(0.94, 1.02)$	0.98	0.02	$(0.94, 1.02)$
ν_2	-0.5	-0.49	0.02	$(-0.52, -0.47)$	-0.49	0.02	$(-0.52, -0.46)$
CONST	-0.5	-0.53	0.03	$(-0.59, -0.46)$	-0.83	0.03	$(-0.89, -0.77)$
ν_1	0.5	0.45	0.03	$(0.38, 0.51)$	0.23	0.03	$(0.18, 0.27)$
CONST	1	0.99	0.01	$(0.97, 1.01)$	0.82	0.01	$(0.81, 0.84)$
ν_1	-1	-1.01	0.01	$(-1.03, -0.99)$	-1.11	0.01	$(-1.13, -1.10)$
CONST	0.5	0.48	0.06	$(0.36, 0.63)$	0.91	0.02	$(0.87, 0.95)$
ν_1	-0.5	-0.49	0.04	$(-0.56, -0.42)$	-0.68	0.02	$(-0.73, -0.64)$
CONST	-1	-1.01	0.05	$(-1.11, -0.93)$	-1.18	0.02	$(-1.22, -1.15)$
ν_1	1	1.01	0.03	$(0.96, 1.06)$	1.08	0.02	$(1.05, 1.11)$
ω_{01}	1.04	1.11	0.06	$(0.99, 1.23)$	0.94	0.05	$(0.85, 1.04)$
ω_{012}	-0.13	-0.13	0.01	$(-0.16, -0.11)$	-0.21	0.01	$(-0.23, -0.18)$
ω_{02}	0.32	0.31	0.01	$(0.30, 0.33)$	0.28	0.00	$(0.27, 0.29)$
ω_{d01}	0.55	0.58	0.05	$(0.48, 0.67)$			
ω_{d02}	0.34	0.32	0.02	$(0.28, 0.35)$			
ω_{11}	0.71	0.68	0.04	$(0.60, 0.75)$	0.59	0.03	$(0.54, 0.64)$
ω_{112}	-0.12	-0.12	0.02	$(-0.16, -0.09)$	-0.09	0.01	$(-0.12, -0.06)$
ω_{12}	0.65	0.64	0.02	$(0.61, 0.67)$	0.63	0.01	$(0.60, 0.65)$
ω_{d11}	0.46	0.43	0.06	$(0.29, 0.55)$			
ω_{d12}	-0.19	-0.17	0.04	$(-0.25, -0.08)$			

Table 2.2: Simulation Treatment Effects Using $\mathcal{M}_0$ and $\mathcal{M}_1$

		$\mathcal{M}_0$			$\mathcal{M}_1$		
	True	Mean	SD	95% CI	Mean	SD	95% CI
ATE_1	1.01	1.02	0.07	$(0.88, 1.17)$	1.74	0.04	$(1.67, 1.82)$
ATT_1	0.33	0.38	0.07	$(0.25, 0.52)$	1.13	0.04	$(1.05, 1.21)$
ATE_2	-2.01	-2.02	0.05	$(2.12, -1.93)$	-2.02	0.02	$(-2.06, -1.98)$
ATT_2	-0.66	-0.65	0.04	$(-0.73, -0.58)$	-0.53	0.02	$(-0.57, -0.50)$
$\ln(m(y))$				-27622.50			-27803.92

the converged empirical estimate distribution by extracting the 2.5% and 97.5% quantiles. The estimated parameter values are presented in Tables 2.1 and 2.3. From the results, it is evident that model $\mathcal{M}_0$ accurately estimates the parameters, whereas all other models exhibit less effective performance. The findings regarding treatment effects are summarized

Table 2.3: Simulation Parameters Estimates Using $\mathcal{M}_2$ and $\mathcal{M}_3$

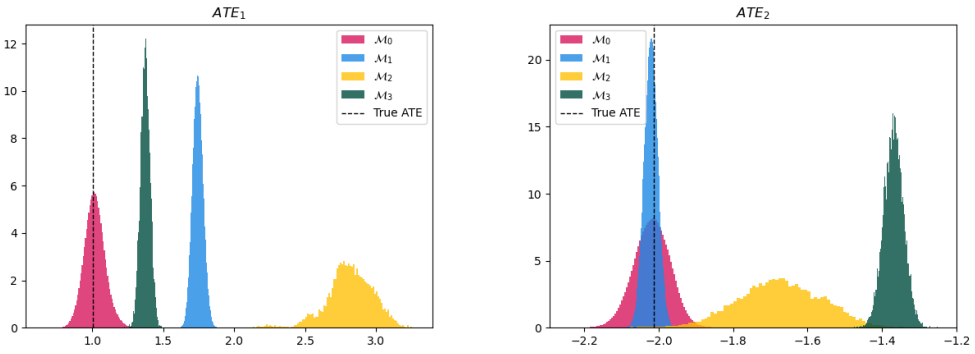
	$\mathcal{M}_2$			$\mathcal{M}_3$		
	Mean	SD	95% CI	Mean	SD	95% CI
CONST	-0.34	0.01	$(-0.37, -0.32)$	-0.34	0.01	$(-0.37, -0.32)$
ν_1	-0.01	0.01	$(-0.03, 0.01)$	-0.01	0.01	$(-0.03, 0.02)$
ν_2	-0.03	0.01	$(-0.05, -0.01)$	-0.03	0.01	$(-0.06, -0.01)$
CONST	-1.43	0.07	$(-1.55, -1.29)$	-0.88	0.02	$(-0.93, -0.83)$
ν_1	-0.14	0.02	$(-0.18, -0.10)$	-0.12	0.02	$(-0.15, -0.08)$
D	2.82	0.16	$(2.47, 3.10)$	1.37	0.03	$(1.31, 1.44)$
CONST	1.25	0.04	$(1.17, 1.33)$	1.13	0.01	$(1.11, 1.16)$
ν_1	-0.33	0.01	$(-0.35, -0.30)$	-0.33	0.01	$(-0.36, -0.31)$
D	-1.67	0.11	$(-1.90, -1.46)$	-1.37	0.03	$(-1.42, -1.32)$
ω_1	1.33	0.10	$(1.14, 1.53)$	0.87	0.03	$(0.82, 0.93)$
ω_{12}	-0.56	0.05	$(-0.66, -0.48)$	-0.46	0.02	$(-0.50, -0.43)$
ω_2	1.24	0.02	$(1.20, 1.29)$	1.22	0.02	$(1.18, 1.25)$
ω_{d1}	-0.85	0.09	$(-1.00, -0.66)$			
ω_{d2}	0.19	0.07	$(-0.06, 0.32)$			

Table 2.4: Simulation Treatment Effects Using $\mathcal{M}_2$ and $\mathcal{M}_3$

		$\mathcal{M}_2$			$\mathcal{M}_3$		
	True	Mean	SD	95% CI	Mean	SD	95% CI
ATE_1	1.01	2.82	0.16	$(2.47, 3.10)$	1.37	0.03	$(1.31, 1.44)$
ATT_1	0.33						
ATE_2	-2.01	-1.67	0.11	$(-1.90, -1.46)$	-1.37	0.03	$(-1.42, -1.32)$
ATT_2	-0.66						
$\ln(m(y))$				-35832.09			-35832.91

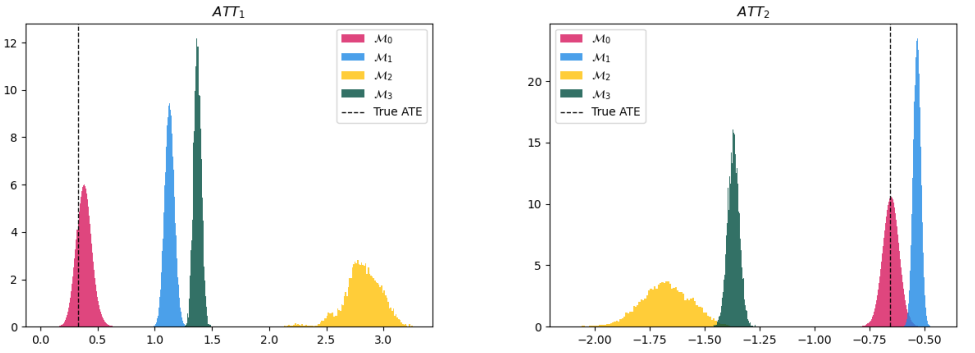
in Tables 2.2 and 2.4. Additionally, we present the posterior distribution of ATE_1, ATE_2, ATT_1, and ATT_2 in Figures 2.1 and 2.2. Model $\mathcal{M}_0$ demonstrates accurate estimations of the treatment effects compared to the other models, which exhibit less effective performance in this regard. Based on the marginal likelihood results, model $\mathcal{M}_0$ emerges as the favored choice.

Figure 2.1: Simulation ATE Histograms



(a) ATE_1 (b) ATE_2

Figure 2.2: Simulation ATT Histograms



(a) ATT_1 (b) ATT_2

2.4 Empirical Application

We apply our model to analyze two datasets to investigate the impact of insurance on healthcare utilization. The first dataset is sourced from the 1987-1988 National Medical Expenditure Survey (NMES), encompassing data on the U.S. elderly population with positive medical expenditures. The second dataset is derived from the 1996 Medical Expenditure Panel Survey (MEPS), comprising non-elderly individuals privately insured with positive medical expenditures. Table 2.5 provides a summary of statistics and variable definitions for both datasets.

Table 2.5: Variable Definition and Summary Statistics

Variable	Data set	NMES		MEPS	
	Number of Observations	3680		5368	
	Definition	Mean	SD	Mean	SD
VIS	# of physician office visits	6.88	6.87	4.88	5.98
EXP	Expenditure on physician office visits	422.8	785.5	499.5	1047.5
EXCH	=1 if self-perceived health is excellent	0.09	0.29	0.32	0.47
POORH	=1 if self-perceived health is poor	0.14	0.34	0.02	0.15
CHRON	# of chronic conditions	2.02	1.42	0.80	1.12
ADL	=1 if has a condition which limits activities of daily living	0.21	0.40		
INJURY	# of injuries which limit activities of daily living since 1996			0.41	0.82
NE	=1 if lives in northeastern U.S.	0.19	0.39	0.21	0.40
MID	=1 if lives in midwestern U.S.	0.26	0.44	0.25	0.43
WEST	=1 if lives in western U.S.	0.19	0.39	0.21	0.41
AGE	age in years (divided by 10)	7.40	0.62	4.14	1.25
BLACK	=1 if is African American	0.10	0.31	0.10	0.30
FEMALE	= 1 if female	0.61	0.49	0.58	0.49
MARR	= 1 if the person is married	0.56	0.50	0.68	0.47
SCH	# of years of education	10.6	3.5	13.32	2.58
FAMINC	Family income in $1,000	25.8	30.1	59.11	39.02
EMPL	=1 if the person is employed	0.10	0.30	0.82	0.38
PRIVATE	=1 if covered by private health insurance	0.80	0.40		
MCAID	=1 if covered by Medicaid	0.09	0.28		
INSUR	=0 if covered by HMO = 1 if FFS			0.51	0.50
SEMP	=1 if self-employed			0.09	0.28
SIZE	size of the company			127.5	177.8
LOC	=1 if multiple locations			0.52	0.50
GOVT	= 1 if the company is governmental			0.18	0.39

2.4.1 Private Insurance

Our first application aims to examine the influence of private insurance on the number of physician office-based visits and the associated office-based total expenditures. We obtained data from $3,680$ observations of elderly Americans from the 1987 National Medical Expenditure Survey (NMES). The definitions of variables and summary statistics can be referenced

in Table 2.5.

Individuals aged 65 and older are typically covered by Medicare, which caters to a broad spectrum of healthcare services. Some individuals opt to supplement this coverage with private insurance, particularly if they have chronic health conditions or poor health status.

Table 2.6: MCMC Parameters Estimates Using $\mathcal{M}_0$ (NMES)

	INS		VIS(0)		EXP(0)		VIS(1)		EXP(1)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
CONST	−0.20	0.34	1.69*	0.42	4.71*	0.60	1.11*	0.23	4.15*	0.31
EXCH	0.07	0.10	0.02	0.15	−0.01	0.20	−0.19*	0.06	−0.18*	0.07
POORH	−0.05	0.07	0.32*	0.08	0.35*	0.12	0.33*	0.05	0.37*	0.07
CHRON	−0.03	0.02	0.14*	0.02	0.12*	0.03	0.14*	0.01	0.13*	0.02
ADL	−0.20*	0.07	0.16*	0.08	0.32*	0.12	0.07	0.04	−0.01	0.06
NE	0.12	0.07	0.07	0.10	0.10	0.13	0.10*	0.04	0.27*	0.06
MID	0.27*	0.07	−0.08	0.10	−0.23	0.15	0.03	0.04	0.03	0.06
WEST	−0.15*	0.07	0.22*	0.10	0.50*	0.14	0.06	0.05	0.27*	0.06
BLACK	−0.80*	0.07	0.09	0.12	0.30	0.24	−0.19*	0.07	−0.45*	0.10
MALE	−0.01	0.06	−0.01	0.08	−0.08	0.11	−0.02	0.03	0.01	0.05
MARR	0.25*	0.06	−0.07	0.09	−0.14	0.13	−0.01	0.04	0.07	0.05
SCH	0.11*	0.01	−0.01	0.02	−0.03	0.03	0.03*	0.01	0.06*	0.01
AGE	−0.01	0.04	−0.11*	0.05	−0.09	0.07	−0.02	0.03	0.00	0.04
EMPL	0.10	0.09	0.08	0.14	−0.04	0.20	−0.02	0.05	−0.01	0.07
MCAID			0.17*	0.08	0.21*	0.10	0.20	0.12	0.20	0.16
FAMINC	0.00*	0.00								
ω_{j1}			0.59*	0.07			0.54*	0.02		
ω_{j12}			0.85*	0.11			0.75*	0.03		
ω_{j2}			1.70*	0.30			1.40*	0.05		
ω_{dj1}			−0.20	0.17			0.15*	0.05		
ω_{dj2}			−0.64	0.39			0.70*	0.07		

We apply our model to assess the influence of private insurance on the number of physician office-based visits and associated expenditures. We incorporate covariates such as self-perceived health status, number of chronic conditions, disability status, geographic location, demographic factors, and insurance variables into the equations determining the number of physician office visits and healthcare expenditures. While there may be heterogeneous effects attributable to various types of private insurance policies, our primary focus is on investi-

gating the overall impact of Medigap plans. We assume that family income influences the purchase of private insurance but does not directly affect healthcare utilization. Additionally, we assume that Medicaid has no impact on the selection equation once family income has been controlled for. We report the posterior mean and standard deviation for the parameter

Table 2.7: MCMC Parameters Estimates Using $\mathcal{M}_1$ (NMES)

	INS		VIS(0)		EXP(0)		VIS(1)		EXP(1)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
CONST	−0.24	0.35	1.81*	0.40	5.09*	0.53	1.24*	0.22	4.64*	0.30
EXCH	0.06	0.10	0.03	0.14	0.02	0.19	−0.19*	0.05	−0.19*	0.07
POORH	−0.06	0.07	0.31*	0.08	0.33*	0.11	0.33*	0.05	0.41*	0.07
CHRON	−0.04	0.02	0.13*	0.02	0.10*	0.03	0.14*	0.01	0.14*	0.02
ADL	−0.24*	0.07	0.14	0.08	0.23*	0.10	0.09*	0.04	0.07	0.06
NE	0.15*	0.07	0.08	0.09	0.16	0.12	0.09	0.04	0.23*	0.06
MID	0.29*	0.07	−0.04	0.09	−0.10	0.13	0.01	0.04	−0.04	0.05
WEST	−0.17*	0.07	0.20*	0.09	0.43*	0.12	0.07	0.05	0.30*	0.06
BLACK	−0.85*	0.08	−0.02	0.07	−0.06	0.10	−0.11	0.07	−0.14	0.09
MALE	−0.04	0.06	−0.01	0.08	−0.09	0.11	−0.02	0.04	0.02	0.05
MARR	0.27*	0.06	−0.03	0.08	0.00	0.10	−0.03	0.04	−0.01	0.05
SCH	0.11*	0.01	0.01	0.01	0.02	0.01	0.02*	0.01	0.03*	0.01
AGE	0.00	0.04	−0.11*	0.05	−0.08	0.07	−0.02	0.03	0.00	0.04
EMPL	0.09	0.10	0.10	0.14	0.04	0.19	−0.02	0.05	−0.02	0.07
MCAID			0.16*	0.07	0.20*	0.10	0.20	0.12	0.23	0.16
FAMINC	0.01*	0.00								
ω_1			0.54*	0.04			0.53*	0.02		
ω_{12}			0.73*	0.05			0.71*	0.02		
ω_2			1.31*	0.07			1.24*	0.03		

estimated. The results are summarized in Table 2.6, 2.7, 2.8, and 2.9. Coefficients significantly different from zero at the 95% confidence level are highlighted using asterisks. It's worth noting that the standard deviation for the control group is larger than that obtained for the treatment group. This discrepancy arises from the fact that only 20% of the sample comprises the control group. Due to the larger standard deviations in the control group, certain coefficients that are not substantially different from zero in the control group appear significant in the treated group. These include variables such as $EXCH$, NE, and SCH in both the doctor's visit and expenditure equations. Conversely, estimates for coefficients of

other variables, such as ADL and $MCAID$, are significant in the control group but not in the treated group.

Table 2.8: MCMC Parameters Estimates Using $\mathcal{M}_2$ (NMES)

	INS		VIS		EXP	
	Mean	SD	Mean	SD	Mean	SD
CONST	−0.23	0.35	1.30*	0.22	4.58*	0.29
EXCH	0.06	0.10	−0.17*	0.05	−0.17*	0.07
POORH	−0.06	0.08	0.32*	0.04	0.38*	0.06
CHRON	−0.04	0.02	0.13*	0.01	0.13*	0.01
ADL	−0.24*	0.07	0.09*	0.04	0.11*	0.05
NE	0.14*	0.07	0.09*	0.04	0.22*	0.05
MID	0.29*	0.07	0.01	0.04	−0.04	0.05
WEST	−0.17*	0.07	0.09*	0.04	0.32*	0.05
BLACK	−0.84*	0.08	−0.09	0.07	−0.12	0.09
MALE	−0.04	0.06	−0.01	0.03	0.00	0.04
MARR	0.26*	0.06	−0.03	0.04	−0.01	0.05
SCH	0.11*	0.01	0.02*	0.01	0.03*	0.01
AGE	0.00	0.04	−0.04	0.03	−0.02	0.03
EMPL	0.08	0.10	−0.02	0.05	−0.02	0.06
MCAID			0.18*	0.06	0.23*	0.08
FAMINC	0.01*	0.00				
PRIVATE			0.11	0.20	0.26	0.23
ω_1	0.53*	0.02				
ω_{12}	0.72*	0.02				
ω_2	1.26*	0.03				
ω_{d1}	0.04	0.11				
ω_{d2}	0.03	0.13				

We estimated the treatment effects and marginal likelihood using the four models, with the results summarized in Tables 2.10 and 2.11. The treatment effects estimated are not significant in the unrestricted models $\mathcal{M}_0$ and $\mathcal{M}_2$, under the endogeneity assumption. However, the treatment effects estimated are significantly positive for the restricted models $\mathcal{M}_1$ and $\mathcal{M}_3$. This suggests that after accounting for the correlation between insurance and utilization, private insurance has no significant impact on healthcare utilization, indicating the presence of selection bias. We observed significant estimates for the covariance components ω_{d11} and ω_{d12} for the treated group in model $\mathcal{M}_0$ to support the statement. However, the

Table 2.9: MCMC Parameters Estimates Using $\mathcal{M}_3$ (NMES)

	INS		VIS		EXP	
	Mean	SD	Mean	SD	Mean	SD
CONST	−0.24	0.36	1.27*	0.20	4.55*	0.26
EXCH	0.06	0.10	−0.17*	0.05	−0.16*	0.07
POORH	−0.06	0.07	0.32*	0.04	0.38*	0.06
CHRON	−0.04	0.02	0.14*	0.01	0.13*	0.01
ADL	−0.24*	0.07	0.10*	0.04	0.11*	0.05
NE	0.14	0.07	0.08*	0.04	0.21*	0.05
MID	0.29*	0.07	0.00	0.04	−0.05	0.05
WEST	−0.17*	0.07	0.09*	0.04	0.33*	0.06
BLACK	−0.85*	0.08	−0.07	0.05	−0.10	0.07
MALE	−0.04	0.06	−0.01	0.03	0.00	0.04
MARR	0.26*	0.06	−0.03	0.03	−0.01	0.04
SCH	0.11*	0.01	0.02*	0.00	0.03*	0.01
AGE	0.00	0.04	−0.04	0.02	−0.02	0.03
EMPL	0.08	0.10	−0.02	0.05	−0.02	0.06
MCAID			0.19*	0.06	0.23*	0.08
FAMINC	0.01*	0.00				
PRIVATE			0.18*	0.04	0.32*	0.06
ω_1	0.53*	0.02				
ω_{12}	0.71*	0.02				
ω_2	1.26*	0.03				

Table 2.10: Impact of Medigap Policies on Healthcare Utilization Using $\mathcal{M}_0$ and $\mathcal{M}_1$ (NMES)

	$\mathcal{M}_0$			$\mathcal{M}_1$		
	Mean	SD	95% CI	Mean	SD	95% CI
ATE_1	0.45	0.26	$(-0.09, 0.97)$	0.20	0.05	$(0.10, 0.29)$
ATT_1	0.49	0.27	$(-0.09, 1.04)$	0.20	0.05	$(0.10, 0.31)$
ATE_2	1.06	0.59	$(-0.23, 1.88)$	0.32	0.06	$(0.20, 0.44)$
ATT_2	1.17	0.63	$(-0.20, 2.04)$	0.32	0.07	$(0.19, 0.45)$
$\ln(m(y))$			−16176.61			−16248.05

estimates for ω_{d01} and ω_{d02} in $\mathcal{M}_0$, as well as the estimates for ω_{d1} and ω_{d2} in model $\mathcal{M}_2$, are not significant. Thus, we remain uncertain about the presence of selection bias. Based on the marginal likelihood results, model $\mathcal{M}_3$ is favored over the other competing models. Notably, this model suggests a positive impact of Medigap policies on healthcare utilization,

Table 2.11: Impact of Medigap Policies on Healthcare Utilization Using $\mathcal{M}_2$ and $\mathcal{M}_3$ (NMES)

	$\mathcal{M}_2$			$\mathcal{M}_3$		
	Mean	SD	95% CI	Mean	SD	95% CI
ATE_1	0.11	0.20	$(-0.28, 0.47)$	0.18	0.04	$(0.10, 0.27)$
ATE_2	0.26	0.23	$(-0.20, 0.68)$	0.32	0.06	$(0.20, 0.43)$
$\ln(m(y))$			-16156.89			-16147.03

indicating the presence of a moral hazard issue.

2.4.2 FFS or HMO

In this application, we compare the impact of selecting different types of private insurance on healthcare utilization using data from the 1996 Medical Expenditure Panel Survey (MEPS). The sample comprises individuals aged between 16 and 65. We are comparing two categories of private insurance: indemnity plans (FFS) and HMO plans. FFS is a payment system where patients pay for services upfront and are later reimbursed by the insurance company. This arrangement can potentially lead to the overuse of healthcare services, as doctors may be incentivized to prescribe more treatments or services. The definitions and summary statistics for the variables are available in Table 2.5.

The treatment variable $INSUR$ equals 1 if the insurance type is indemnity plans (FFS) and 0 if it's HMO. Similar to the previous section, we assume that employment-related variables, such as company size, location, self-employment status, and family income, solely influence the selection of private insurance and do not directly impact healthcare utilization. This assumption is grounded in reality, as many individuals have limited options for health insurance plans, often tied closely to their employment circumstances.

The posterior mean and standard deviation for the four models are summarized in Tables 2.12, 2.13, 2.14, and 2.15. Coefficients significant at the 95% level are highlighted with

Table 2.12: MCMC Parameters Estimates Using $\mathcal{M}_0$ (MEPS)

	INS		VIS(0)		EXP(0)		VIS(1)		EXP(1)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
CONST	0.01	0.12	-0.12	0.14	3.80^*	0.19	0.28^*	0.13	3.38^*	0.17
EXCH	0.10	0.04	-0.28^*	0.04	-0.18^*	0.05	-0.17^*	0.04	0.00	0.05
POORH	-0.04	0.12	0.37^*	0.12	0.43^*	0.17	0.23^*	0.11	0.32	0.16
CHRON	0.00	0.02	0.21^*	0.02	0.22^*	0.02	0.23^*	0.01	0.27^*	0.02
INJURY	0.01	0.02	0.17^*	0.02	0.24^*	0.03	0.12^*	0.02	0.11^*	0.03
BLACK	-0.21^*	0.06	-0.08	0.06	-0.18^*	0.08	-0.16^*	0.07	-0.32^*	0.09
FEMALE	-0.07	0.04	0.39^*	0.04	0.38^*	0.05	0.30^*	0.04	0.33^*	0.05
MARR	0.05	0.04	-0.07	0.04	-0.10^*	0.05	0.00	0.04	0.01	0.06
SCH	0.00	0.01	0.03^*	0.01	0.04^*	0.01	0.02^*	0.01	0.03^*	0.01
EMPL	-0.16^*	0.05	0.03	0.05	0.02	0.07	-0.14^*	0.05	-0.18^*	0.06
AGE	0.08^*	0.02	-0.01	0.02	0.06^*	0.02	0.05^*	0.02	0.14^*	0.02
NE	-0.11^*	0.05	0.14^*	0.05	0.02	0.07	0.09	0.05	0.07	0.07
MID	0.12^*	0.05	-0.10	0.05	-0.25^*	0.07	-0.03	0.04	-0.06	0.06
WEST	-0.40^*	0.05	0.09	0.05	0.13	0.07	-0.09	0.06	-0.29^*	0.07
SIZE	-0.00^*	0.00								
GOVT	0.04	0.04								
LOC	-0.03	0.04								
SEMP	0.05	0.06								
FAMINC	-0.00^*	0.00								
ω_{j1}			0.83^*	0.06			0.59^*	0.03		
ω_{j12}			1.01^*	0.08			0.87^*	0.05		
ω_{j2}			1.60^*	0.10			1.89^*	0.10		
ω_{dj1}			-0.72^*	0.06			0.20^*	0.05		
ω_{dj2}			-0.62^*	0.11			0.97^*	0.07		

asterisks. All four models consistently estimate a significant positive impact of variables $POORH$, $CHRON$, $INJURY$, $FEMALE$, and SCH on healthcare expenditure and the frequency of doctors' office visits, in both the control group and the treated group. Moreover, they estimate a significant negative impact of variable $EXCH$ on healthcare utilization in both groups.

According to Table 2.12, the estimated covariance between insurance selection and healthcare utilization is significantly different from zero in model $\mathcal{M}_0$. Specifically, the covariance is negative in the control group and positive in the treated group. However, the estimated

Table 2.13: MCMC Parameters Estimates Using $\mathcal{M}_1$ (MEPS)

	INS		VIS(0)		EXP(0)		VIS(1)		EXP(1)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
CONST	0.00	0.12	0.47*	0.12	4.30*	0.16	0.45*	0.12	4.14*	0.15
EXCH	0.11*	0.04	−0.23*	0.04	−0.14*	0.05	−0.18*	0.04	−0.07	0.05
POORH	0.01	0.12	0.42*	0.11	0.47*	0.16	0.24*	0.10	0.35*	0.14
CHRON	0.01	0.02	0.22*	0.02	0.22*	0.02	0.23*	0.01	0.27*	0.02
INJURY	0.01	0.02	0.18*	0.02	0.25*	0.03	0.12*	0.02	0.11*	0.03
BLACK	−0.21*	0.06	−0.20*	0.06	−0.28*	0.07	−0.12	0.07	−0.17*	0.08
FEMALE	−0.07	0.04	0.36*	0.04	0.35*	0.05	0.31*	0.04	0.37*	0.05
MARR	0.02	0.04	−0.06	0.04	−0.10	0.05	0.00	0.04	0.00	0.05
SCH	0.00	0.01	0.03*	0.01	0.04*	0.01	0.02*	0.01	0.04*	0.01
EMPL	−0.14*	0.06	−0.08	0.05	−0.08	0.07	−0.10*	0.04	−0.03	0.06
AGE	0.08*	0.02	0.02	0.02	0.08*	0.02	0.04*	0.02	0.08*	0.02
NE	−0.12*	0.05	0.09	0.05	−0.03	0.06	0.11*	0.05	0.15*	0.06
MID	0.15*	0.05	−0.04	0.05	−0.20*	0.06	−0.05	0.04	−0.16*	0.06
WEST	−0.43*	0.05	−0.07	0.05	−0.03	0.06	−0.03	0.05	−0.03	0.07
SIZE	−0.00*	0.00								
GOVT	−0.02	0.05								
LOC	−0.05	0.05								
SEMP	0.12	0.07								
FAMINC	0.00	0.00								
ω_{j1}			0.53*	0.02			0.54*	0.02		
ω_{j12}			0.74*	0.03			0.74*	0.02		
ω_{j2}			1.35*	0.04			1.32*	0.04		

covariance between insurance selection and healthcare utilization is not significant in model $\mathcal{M}_2$, as indicated in Table 2.14. Furthermore, the estimated covariance parameters for models $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3$ are similar to each other, as illustrated in Tables 2.13, 2.14, and 2.15.

The estimated treatment effects and marginal likelihood are summarized in Tables 2.16 and 2.17. Model $\mathcal{M}_0$ predicted a positive treatment effect of the FFS plan on the frequency of physician office visits, while the impact on healthcare out-of-pocket expenditure remains unclear. However, the estimated treatment effects using all the other models are not significant. Based on the results using $\mathcal{M}_1$, after restricting that the correlation between insurance selection and healthcare utilization is 0, the average treatment effects estimates are no longer

Table 2.14: MCMC Parameters Estimates Using $\mathcal{M}_2$ (MEPS)

	INS		VIS		EXP	
	Mean	SD	Mean	SD	Mean	SD
CONST	0.00	0.12	0.49*	0.16	4.34*	0.22
EXCH	0.11*	0.04	−0.20*	0.03	−0.09*	0.04
POORH	0.00	0.12	0.33*	0.08	0.40*	0.11
CHRON	0.01	0.02	0.22*	0.01	0.25*	0.02
INJURY	0.01	0.02	0.14*	0.01	0.17*	0.02
BLACK	−0.22*	0.06	−0.17*	0.05	−0.26*	0.07
FEMALE	−0.07	0.04	0.33*	0.03	0.36*	0.03
MARR	0.03*	0.04	−0.03	0.03	−0.05	0.04
SCH	0.00	0.01	0.03*	0.00	0.04*	0.01
EMPL	−0.14*	0.06	−0.10*	0.04	−0.07	0.06
AGE	0.08*	0.02	0.03*	0.01	0.09*	0.02
NE	−0.12*	0.05	0.10*	0.04	0.05	0.05
MID	0.15*	0.05	−0.04	0.04	−0.16*	0.05
WEST	−0.42*	0.05	−0.06	0.06	−0.06	0.08
SIZE	−0.00*	0.00				
GOVT	−0.01	0.05				
LOC	−0.06	0.05				
SEMP	0.11	0.07				
FAMINC	−0.00	0.00				
INSUR			−0.07	0.25	−0.25	0.38
ω_1	0.56*	0.03				
ω_{12}	0.76*	0.03				
ω_2	1.39*	0.06				
ω_{d1}	0.04	0.16				
ω_{d2}	0.18	0.24				

significant. This suggests some evidence of selection bias. However, the difference between the treatment effects estimated by models $\mathcal{M}_2$ and $\mathcal{M}_3$ is not readily apparent. Thus, the presence of selection bias remains uncertain. According to the marginal likelihood results, the constant treatment model, $\mathcal{M}_2$, is recommended.

Table 2.15: MCMC Parameters Estimates Using $\mathcal{M}_3$ (MEPS)

	INS		VIS		EXP	
	Mean	SD	Mean	SD	Mean	SD
CONST	0.00	0.12	0.46*	0.09	4.19*	0.11
EXCH	0.11*	0.04	−0.21*	0.03	−0.10*	0.04
POORH	0.01	0.12	0.33*	0.08	0.41*	0.11
CHRON	0.01	0.02	0.22*	0.01	0.25*	0.02
INJURY	0.01	0.02	0.14*	0.01	0.16*	0.02
BLACK	−0.21*	0.06	−0.16*	0.04	−0.23*	0.06
FEMALE	−0.07	0.04	0.33*	0.03	0.36*	0.03
MARR	0.02	0.04	−0.03	0.03	−0.05	0.04
SCH	0.00	0.01	0.03*	0.00	0.04*	0.01
EMPL	−0.15*	0.06	−0.09*	0.03	−0.05	0.04
AGE	0.08*	0.02	0.03*	0.01	0.09*	0.01
NE	−0.12*	0.05	0.10*	0.03	0.06	0.04
MID	0.15*	0.05	−0.05	0.03	−0.18*	0.04
WEST	−0.43*	0.05	−0.06	0.04	−0.02	0.05
SIZE	−0.00*	0.00				
GOVT	−0.02	0.05				
LOC	−0.05	0.05				
SEMP	0.12	0.07				
FAMINC	−0.00	0.00				
INSUR			−0.00	0.02	0.03	0.03
ω_1	0.54	0.01				
ω_{12}	0.74	0.02				
ω_2	1.34	0.03				

Table 2.16: Impact of FFS on Healthcare Utilization Using $\mathcal{M}_0$ and $\mathcal{M}_1$ (MEPS)

	$\mathcal{M}_0$			$\mathcal{M}_1$		
	Mean	SD	95% CI	Mean	SD	95% CI
ATE_1	0.42	0.07	$(0.28, 0.55)$	−0.01	0.03	$(-0.06, 0.04)$
ATT_1	0.45	0.07	$(0.31, 0.58)$	−0.01	0.03	$(-0.06, 0.05)$
ATE_2	−0.22	0.11	$(-0.44, -0.01)$	0.03	0.03	$(-0.03, 0.10)$
ATT_2	−0.17	0.11	$(-0.39, 0.05)$	0.03	0.03	$(-0.03, 0.10)$
$\ln(m(y))$			−23497.00			−23586.72

Table 2.17: Impact of FFS on Healthcare Utilization Using $\mathcal{M}_2$ and $\mathcal{M}_3$ (MEPS)

	$\mathcal{M}_2$			$\mathcal{M}_3$		
	Mean	SD	95% CI	Mean	SD	95% CI
ATE_1	-0.07	0.25	$(-0.60, 0.42)$	0.00	0.02	$(-0.05, 0.04)$
ATE_2	-0.25	0.38	$(-0.84, 0.63)$	0.03	0.03	$(-0.03, 0.10)$
$\ln(m(y))$			-23493.22			-23497.67

2.5 Conclusion

In conclusion, this chapter has introduced a parametric self-selection model with multiple outcomes within a Bayesian framework to estimate the impact of health insurance on healthcare expenditures. Our model incorporates two outcomes: the number of doctor's office visits and healthcare expenditure. We develop a simulation-based estimation algorithm for parameter estimation. Additionally, we propose an approach to estimate the marginal likelihood for model comparison.

Simulation studies have been conducted to evaluate the performance of the MCMC algorithm and the model comparison techniques. Based on the results, our algorithm can accurately estimate the true parameters, and the model comparison results corroborate the true model.

We have employed this model in two empirical applications and compared the results with those obtained using three other, more parsimonious models. Our findings provide some weak evidence supporting the presence of selection bias in both applications.

Chapter 3

Product Pricing with Consumer Learning

We analyze a price signaling game incorporating consumer learning. Initially, buyers are uninformed about the quality of the seller's product. We introduce a probability factor determining the likelihood of the seller's type being fully disclosed before trading. Optimal strategies for sellers are outlined. We employ the undefeated equilibrium refinement to determine optimal choices. Furthermore, our analysis includes comparative statics, investigating the impact of variations in the probability of type revelation, initial customer review scores, and the quality of both high and low-quality seller products on the expected return for sellers.

3.1 Introduction

Consumers frequently encounter challenges when trying to distinguish between the quality discrepancies among various brands. As a result, they tend to associate higher prices with

superior quality. Nonetheless, the proliferation of the internet and online shopping platforms has empowered consumers to leverage customer reviews as a valuable resource for assessing product quality. In this study, we propose a model that delves into optimal pricing strategies within the context of consumer reviews, wherein customers interpret pricing as a signaling mechanism.

Many researchers found evidence supporting the positive correlation between price and product quality (Mastrobuoni, Peracchi and Tetenov, 2014). Bagwell and Riordan (1991) found that high and declining prices signal high-quality products within a dynamic game framework. Milgrom and Roberts (1986) presents a signaling game using both price and uninformative advertisement as signals. Wolinsky (1983) and Delacroix and Shi (2013) studied the price signaling game with endogenously chosen price and signal pairs. Our model extends this by incorporating the consumer's learning process from previous production outcomes. Additionally, in their model, the cost of production is higher if the product quality is higher. In our model, the cost of production is the same for both the high-quality and low-quality productions to eliminate this factor. In contrast to Bose et al. (2006) exploration of dynamic monopoly pricing where buyers learn from each other's purchase decisions, our model assumes buyers learn from the review scores. Bar-Isaac (2003) examined the influence of reputation on the longevity of sellers. Their model featured a mechanism where consumers determined the price of the product, with the seller having the option to either accept or decline the offer. In contrast, our model employs pricing as a signal of product quality. This model is more relevant to many real-world markets where sellers typically set prices prior to purchase.

In our model, a seller interacts with multiple buyers within a monopolistic setting. The seller determines the price of the product, with the constraint that only a single unit can be sold. Buyers possess limited information regarding the product's quality from this seller and rely on consumer reviews and posted prices to update their beliefs about the seller's quality.

Following a similar framework to the model outlined in Delacroix and Shi (2013), our model incorporates the possibility that, with a certain probability, consumers may acquire complete information about the seller's type during the second stage after the seller posts the price. This scenario mirrors real-world situations where the true quality of a product may become publicly known after its release, perhaps through product quality investigations or disclosures.

For the sake of simplicity and ease of analysis, we adopt a single time period setting within stage two of our model. While we acknowledge that this may not perfectly capture all real-world dynamics, it still offers valuable insights. We anticipate that future extensions of our model could relax this constraint by incorporating a continuous time setting, allowing for a more nuanced exploration of the dynamics over time.

We characterize our model within the context of asymmetric information, wherein the seller possesses perfect information regarding its own type, while buyers lack full information about the seller's type.

We adopt the lexicographically maximum sequential equilibrium (LMSE) or undefeated equilibrium to refine the multiple equilibria (Mailath, Okuno-Fujiwara and Postlewaite, 1993), which is widely used in the signaling games (see Taylor, 1999; Fishman and Hagerty, 2003; Jiang et al., 2016; Bajaj, 2018; Jiang and Yang, 2019; Wu, Zhang and Xie, 2020; Li, Tian and Zheng, 2021). This refinement allows us to select the most profitable equilibrium from the seller's perspective among all the equilibria.

This model finds application in numerous scenarios where buyers cannot observe the product's quality perfectly prior to purchase, and they rely on the price set by the seller as an indicator of quality. For example, with the widespread availability of the Internet, numerous review websites have emerged, offering evaluations and rankings for various products. Consumers can readily access consumer review scores for specific products and infer product

quality by considering both the listed price and the scores they encounter.

The rest of the chapter is organized as follows. In Section 3.2 we introduce our model settings and the equilibria and the refinement. In Section 3.3 we conduct a comparative analysis to examine the characteristics of the equilibrium. Section 3.4 offers concluding remarks.

3.2 Model

Consider a market comprising a single seller and multiple buyers. The seller is assumed to be risk-neutral and can fall into one of two categories: high-quality (denoted as H-type) or low-quality (denoted as L-type). We let S denote the seller. The marginal cost of producing the good is the same for both types of sellers, which is normalized to 0. The H-type seller can make a successful production with probability g and the L-type seller can make a successful production with probability b, where $1 > g > b > c > 0$.

Buyers' valuation of a success is 1 and of a failure is 0. Each buyer can consume at most one unit of the product. Each buyer is capable of consuming at most one unit of the product, and they are considered homogeneous and price-takers. We assume a single consumer in each period, with all buyers being rational. We use B to denote the buyers. The success or failure of the product can be determined after it has been produced and the trade is realized.

In a one-period market, the seller S is born with an initial common belief λ representing the probability that the seller is of the H-type. S has perfect information about their type, and subsequently sets a price p for its product. The seller's type is fully revealed to the public with probability ϕ. The buyer B enters the market and observes the price set by S. Upon observing this price and other available information, B updates their belief regarding the seller's type denoted as μ. Subsequently, B decides to make a purchase if the observed price p is less than or equal to the expected value of the product, calculated as $\mu g + (1 - \mu) b$.

If B chooses to proceed with the purchase, S produces one unit of the product, and the transaction is completed. Following the consumption of the good by B, the outcome of the production is revealed to the public.

We assume that if S sets a price p different from $p^*(\lambda)$—where $p^*(\lambda)$ represents the equilibrium price determined by the H-type seller—the resulting posterior belief μ is set to zero. Conversely, if the posted price matches the equilibrium price, S is then identified as an H-type seller with a non-zero probability, in other words, $\mu > 0$. It's evident that S would select a price $p \in [b, g]$.

The expected payoff function for the L-type seller is given by

$$V^L(p, \mu) = (1 - \phi)\, p \mathbb{1}\{p \leq \mu g + (1 - \mu)\, b\} + \phi b \mathbb{1}\{p = b\},$$

where with probability ϕ, the trade only occurs if the price posted is b. Similarly, the expected payoff function for the H-type seller is expressed as

$$V^H(p, \mu) = (1 - \phi)\, p \mathbb{1}\{p \leq \mu^s g + (1 - \mu)\, b\} + \phi p,$$

where with probability ϕ, the trade occurs as long as the posted price p is less than or equal to g.

We employ backward induction to determine the equilibrium in this model. Initially, we solve for the set of perfect Bayesian equilibria, after which we apply the undefeated equilibrium refinement to further refine the equilibrium set.

3.2.1 Perfect Bayesian Equilibrium

In this section, we use backward induction to solve for the equilibrium for this model. We discuss three potential types of equilibrium, including separating equilibrium, semi-pooling equilibrium, and pooling equilibrium.

3.2.1.1 Separating Equilibrium

To attain the separating equilibrium, the L-type opts to set a price b with $\mu = 0$, while the H-type seller selects a price $p^*(\lambda) \neq b$ with $\mu = 1$. Establishing this equilibrium necessitates making it prohibitively expensive for the L-type seller to imitate the pricing strategy of the H-type seller. In essence, this requires that $b \geq (1-\phi)p^*(\lambda)$. This implies that $p^*(\lambda) \leq \left\{\frac{b}{1-\phi}, g\right\}$. If $\phi > \frac{g-b}{g}$, $p^*(\lambda) \leq g$; otherwise, $p^*(\lambda) \leq \frac{b}{1-\phi}$.

This result intuitively aligns with expectations, as when the probability of type revelation is significantly high, it becomes more costly for the L-type seller to mimic the pricing strategy of the H-type seller. This is due to the increased likelihood that the trade will not occur, making it less profitable for the L-type seller to deviate from their own pricing strategy.

3.2.1.2 Semi-pooling Equilibrium

To achieve the semi-pooling equilibrium, the L-type seller faces a situation of indifference between setting the price b and mimicking the pricing strategy of the H-type seller. We let $d(\lambda)$ denote the probability that the L-type seller mimics the H-type seller. This implies that $b = (1-\phi)\,p^*(\lambda)$, or $p^*(\lambda) = \frac{b}{1-\phi}$. This equilibrium exists only under conditions where $\phi < \frac{g-b}{g}$ and $\mu \geq \frac{\phi b}{(1-\phi)(g-b)}$, where $\mu = \frac{\lambda_2}{\lambda_2+(1-\lambda_2)d(\lambda_2)}$. This condition implies that $d(\lambda) \leq \frac{\lambda g(1-\phi)-\lambda b}{(1-\lambda)b\phi}$. Therefore, in the semi-pooling equilibrium, when $\phi \leq \frac{g-b}{g}$, the H-type seller selects the price $p^*(\lambda) = \frac{b}{1-\phi}$, while the L-type seller posts a price of $\frac{b}{1-\phi}$ with probability

$d(\lambda) \leq \frac{\lambda g(1-\phi)-\lambda b}{(1-\lambda)b\phi}$, and chooses the price b with probability $1 - d(\lambda)$.

3.2.1.3 Pooling Equilibrium

The pooling equilibrium exists when both types of sellers set the same price. In this scenario, the updated belief μ equals λ, as both types of sellers set the same price with certainty, leading to no further revelation of information of the seller's type. Within this equilibrium, the L-type seller benefits by emulating the pricing strategy of the H-type seller. This suggests that $b \leq (1-\phi)\, p^*(\lambda)$, or equivalently, $p^*(\lambda) \geq \frac{b}{1-\phi}$, which also infers that $\phi \leq \frac{g-b}{g}$. This condition occurs because when the probability of revelation is sufficiently low enough, it's more advantageous for the L-type to mimic the H-type seller. Moreover, the buyer's willingness to pay is $\mu g + (1-\mu)\, b$, or equivalently, $\lambda g + (1-\lambda)\, b$. Thus, we derive $p^*(\lambda) \leq \lambda g + (1-\lambda)\, b$. This also implies that $\lambda \geq \frac{b\phi}{(1-\phi)(g-b)}$ In summary, when $\phi \leq \frac{g-b}{g}$ and $\lambda \geq \frac{b\phi}{(1-\phi)(g-b)}$, the pooling equilibrium price $p^*(\lambda) \in \left[\frac{b}{1-\phi}, \lambda g + (1-\lambda)\, b\right]$.

3.2.2 Refinement

In this section, we leverage the undefeated equilibrium (Mailath, Okuno-Fujiwara and Postlewaite, 1993) to refine the perfect Bayesian Equilibria discussed in Section 3.2.1. An equilibrium is undefeated if and only if no alternative equilibrium exists wherein either the H-type seller finds it more beneficial to deviate to a separating equilibrium, or both types of sellers have incentives to deviate to a pooling equilibrium. The undefeated equilibrium serves as a commonly employed refinement tool in signaling games. It enables us to select the equilibrium that maximizes the profit of the H-type seller—an important consideration in many market scenarios.

If $\phi > \frac{g-b}{g}$, the separating equilibrium price for the H-type seller lies in the interval $(b, g]$,

while the L-type seller consistently sets the price at b. Among the set of separating perfect Bayesian equilibria, the optimal choice is made based on maximizing the profit of the H-type seller.

In this scenario, imagine the H-type seller initially setting $p^*(\lambda)$ within the range $(b, g]$, while the L-type seller sets the price to be b. In this situation, the H-type seller will deviate to an equilibrium where $p^*(\lambda) = g$. Consequently, all other separating equilibria, except the one where $p^*(\lambda) = g$, are defeated. Since only the separating equilibria can exist in this scenario, the undefeated equilibrium is that the H-type seller would set a price $p^*(\lambda) = \frac{b}{1-\phi}$, and the L-type seller would set a price b.

If $\phi \leq \frac{g-b}{g}$, all three types of equilibria become feasible. The separating equilibrium price for the H-type seller is within the range of $\left(b, \frac{b}{1-\phi}\right]$. All the separating equilibria where the H-type seller's price $p^*(\lambda) < \frac{b}{1-\phi}$ are defeated by the equilibrium where the H-type seller sets the price $\frac{b}{1-\phi}$ and the L-type seller posts the price b. Thus, the separating equilibrium price for the H-type seller is $\frac{b}{1-\phi}$.

Among the semi-pooling equilibria, we argue that all the other equilibria are defeated by the equilibrium where the H-type seller sets the price $p^*(\lambda) = \frac{b}{1-\phi}$. Suppose that the H-type seller deviates to another semi-pooling equilibrium where $p^*(\lambda) < \frac{b}{1-\phi}$; in such a case, they would be strictly worse off since the optimal profit is achieved by setting the price at $\frac{b}{1-\phi}$.

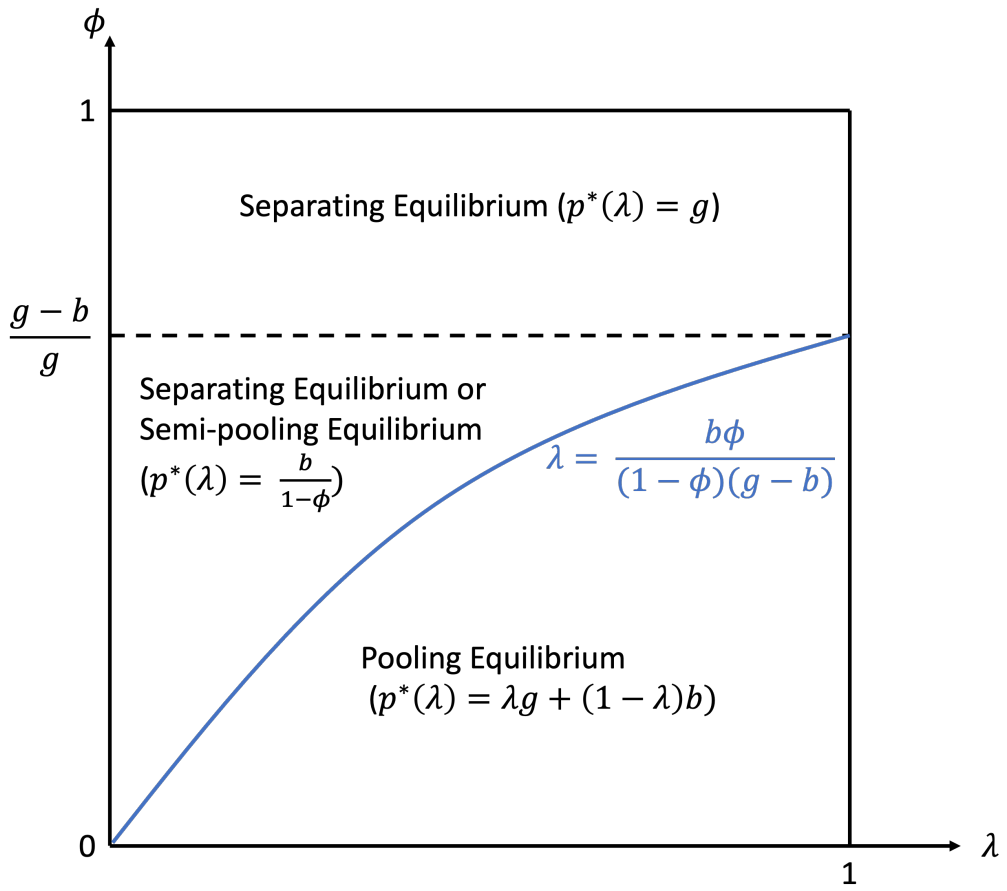
Similarly, all the other pooling equilibrium is defeated by the one where the H-type seller sets the price $\lambda g + (1-\lambda) b$. Suppose the seller sets a price $p^*(\lambda) < \lambda g + (1-\lambda) b$, both types of the seller would benefit from deviating to the pooling equilibrium where they set the price $p^*(\lambda) = \lambda g + (1-\lambda) b$.

When $\lambda < \frac{b\phi}{(1-\phi)(g-b)}$, only the semi-pooling equilibrium and the separating equilibrium are feasible. In this scenario, both the H-type and L-type seller finds themselves indifferent between the two equilibria. When $\lambda \geq \frac{b\phi}{(1-\phi)(g-b)}$, the pooling equilibrium becomes feasible.

In this scenario, the pooling equilibrium defeats both the separating and the semi-pooling equilibria. This is because the price in the pooling equilibrium, $\lambda g + (1-\lambda)\, b$, is weakly greater than the price in the separating and semi-pooling equilibrium, $\frac{b}{1-\phi}$. Consequently, the H-type seller would be better off staying in the pooling equilibrium.

To summarize, in scenarios where $\phi > \frac{g-b}{g}$, the separating equilibrium with $p^*(\lambda) = g$ stands as the undefeated equilibrium. In cases where $\phi \leq \frac{g-b}{g}$ and $\lambda < \frac{b\phi}{(1-\phi)(g-b)}$, both the separating and semi-pooling equilibria with $p^*(\lambda) = \frac{b}{1-\phi}$ emerge as undefeated equilibria. Conversely, in scenarios where $\phi \leq \frac{g-b}{g}$ and $\lambda \geq \frac{b\phi}{(1-\phi)(g-b)}$, the pooling equilibrium with $p^*(\lambda) = \lambda g + (1-\lambda)\, b$ prevails as the undefeated equilibrium. These results are summarized in Figure 3.1.

Figure 3.1: Undefeated Equilibrium



3.3 Comparative Statics

In the preceding section, we demonstrated that equilibrium prices are contingent on variables λ, ϕ, g, and b. In this section, we delve into the comparative statics analysis of these equilibria.

Proposition 1

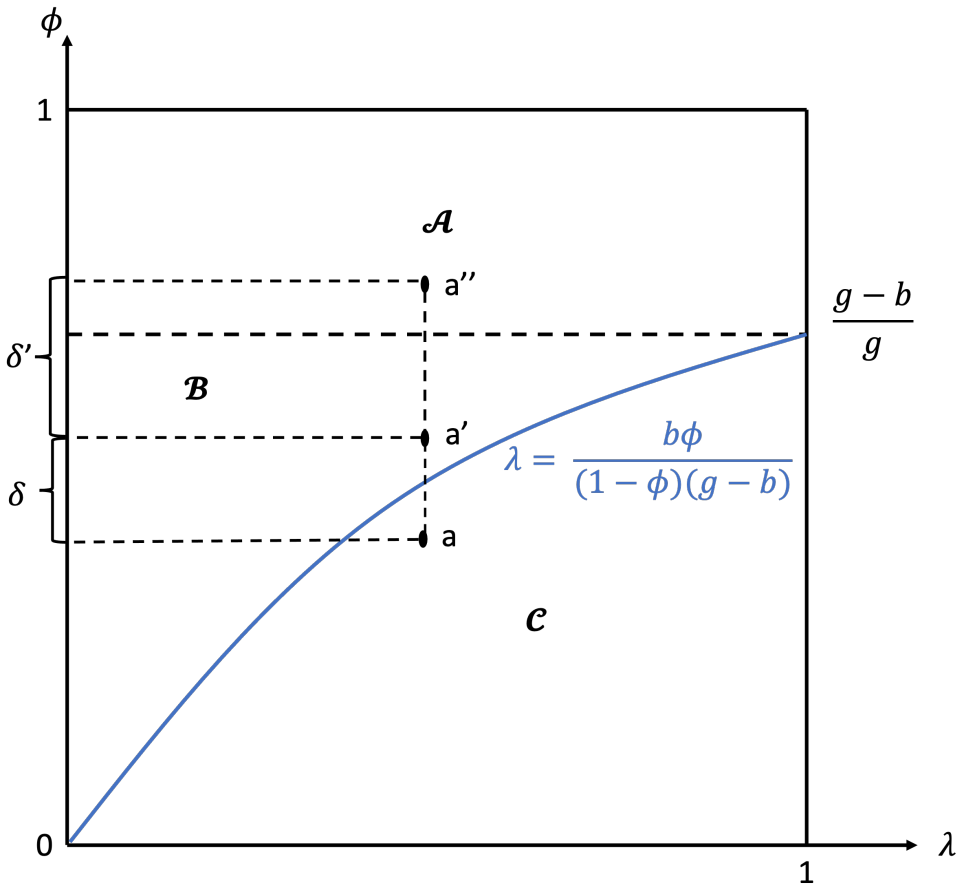
We denote the expected payoffs at equilibrium as $\left(\pi^H, \pi^L\right)$, where π^H represents the equilibrium expected return for the H-type seller, and π^L signifies the equilibrium expected return for the L-type seller. We have

1. $\frac{\partial \pi^H}{\partial \phi} \geq 0$, and $\frac{\partial \pi^L}{\partial \phi} \leq 0$;
2. $\frac{\partial \pi^H}{\partial \lambda} \geq 0$, and $\frac{\partial \pi^L}{\partial \lambda} \geq 0$;
3. $\frac{\partial \pi^H}{\partial g} \geq 0$, and $\frac{\partial \pi^L}{\partial g} \geq 0$;
4. $\frac{\partial \pi^H}{\partial b} \geq 0$, and $\frac{\partial \pi^L}{\partial b} \geq 0$.

The first part of the condition indicates that the equilibrium expected return for both seller types exhibits a weak increase as the probability of type disclosure rises, while the equilibrium expected return for the L-type seller decreases in response to the probability of disclosure. Since we normalized the cost of production to be 0, the expected return for the H-type seller is just the price they set. We illustrate the changes using Figure 3.2.

Based on the figure, consider a seller, denoted as a, is born in the region $\mathcal{C}$ where the pooling equilibrium prevails as the undefeated equilibrium. The equilibrium expected return for the H-type seller is $\lambda g + (1-\lambda)\, b$, and for the L-type seller is $\pi^L = (1-\phi)\left(\lambda g + (1-\lambda)\, b\right)) > b$. If we increase the value of ϕ while the seller remains in region $\mathcal{C}$, we find $\frac{\partial \pi^H}{\partial \phi} = 0$ and $\frac{\partial \pi^L}{\partial \phi} = -\left(\lambda g + (1-\lambda)\, b\right) < 0$.

Figure 3.2: ϕ Increases



Suppose we increase ϕ by δ, leading the seller to transition to point a' in region $\mathcal{B}$. In this region, either the semi-pooling or the separating equilibrium prevails as the undefeated equilibrium, with the expected return being b for the L-type seller and $\frac{b}{1-\phi}$ for the H-type seller. Additionally, in region $\mathcal{B}$, we observe that $\lambda < \frac{b\phi}{(1-\phi)(g-b)}$ and $\phi < \frac{g-b}{g}$, implying $g > \frac{b}{1-\phi} > \lambda g + (1-\lambda)\, b$. Hence, transitioning from region $\mathcal{C}$ to region $\mathcal{B}$ leads to an increase in the expected return of the H-type seller and a decrease in that of the L-type seller.

Furthermore, consider the scenario where we increase the value of ϕ while the seller remains in region $\mathcal{B}$. In this case, the expected return for the H-type seller would increase because $\frac{\partial \frac{b}{1-\phi}}{\partial \phi} > 0$, while the expected return for the L-type seller would remain unchanged.

Considering the further increase of ϕ by δ', this transition moves the seller to point a'' in

region $\mathcal{A}$. In this region, the separating equilibrium is the undefeated equilibrium, characterized by the H-type seller setting the price at g and the L-type seller setting it at b. In region $\mathcal{A}$, the expected return for the H-type seller is g, representing the highest possible expected return for this type of seller, while the L-type seller's expected return remains at b. Consequently, transitioning from region $\mathcal{B}$ to $\mathcal{A}$ results in an increase in the expected return for the H-type seller, with no change in the return for the L-type seller.

In summary, the expected return for the H-type seller experiences a weak increase as ϕ rises, whereas the expected return for the L-type seller decreases with an increase in ϕ. This intuition aligns with real-life scenarios, where a higher likelihood of type disclosure before a purchase benefits the H-type seller due to reduced signaling efforts, while simultaneously disadvantaging the L-type seller.

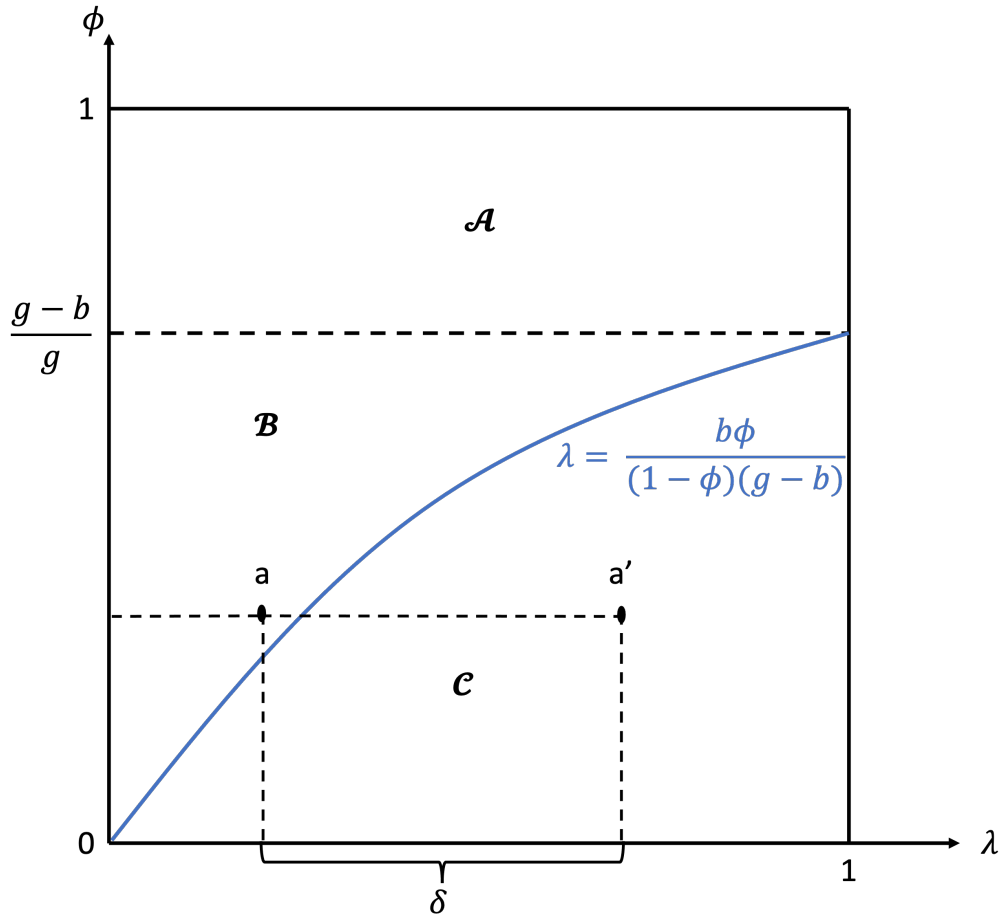
The second part of Proposition 1 indicates that both types of sellers would experience a weak benefit from an increase in the initial reputation parameter λ. We illustrate the changes in Figure 3.3.

Referring to the figure, We consider a seller, labeled as a, born in region $\mathcal{B}$. If we increase the value of λ while seller a remains in the same region, the expected return for both the H-type seller and the L-type seller would remain unchanged, at $\frac{b}{1-\phi}$ and b respectively.

If we increase the value of λ by δ, the seller a moves to position a' within region $\mathcal{C}$. In this scenario, the expected return for both the H-type seller and the L-type seller increases to $\lambda g + (1-\lambda)\, b$ and $(1-\phi)\,(\lambda g + (1-\lambda)\, b)$ respectively. As discussed in Section 3.2.1, we know that in this region, $\lambda g + (1-\lambda)\, b > \frac{b}{1-\phi}$, and $(1-\phi)\,(\lambda g + (1-\lambda)\, b) > b$. Therefore, the expected return for both types of sellers increases.

If we increase ϕ while the seller a remains in the region $\mathcal{C}$, the expected return for both types of sellers increases. This is because $\frac{\partial \lambda g + (1-\lambda) b}{\partial \lambda} = g - b > 0$, and $\frac{\partial (1-\phi)(\lambda g + (1-\lambda) b)}{\partial \lambda} = (1-\phi)\,(g-b) > 0$.

Figure 3.3: λ Increases



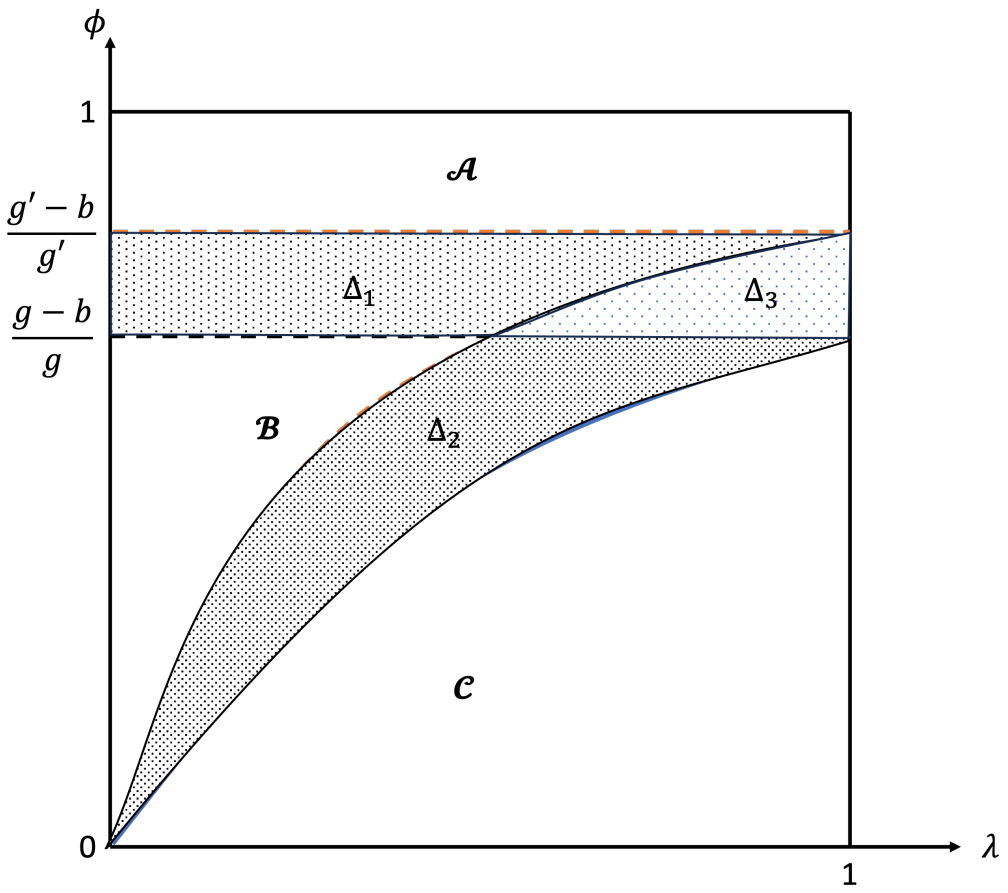
Similarly, if a seller is situated in the region $\mathcal{A}$, the expected return for the H-type seller remains at g and for the L-type seller stays at b as λ increases.

Therefore, the expected return for both types of sellers experiences a weakly increase as λ increases. This aligns with the intuition that both types of sellers would benefit from higher review scores.

The third part of Proposition 1 suggests that both types of sellers benefit from quality improvements in the H-type seller's product. We illustrate these changes using Figure 3.4. Suppose we increase the value of g to g', where $g' > g$. As a result, the region Δ_1, previously within region $\mathcal{A}$, now belongs to part of region $\mathcal{B}$. Simultaneously, the region Δ_2 and Δ_3, previously within region $\mathcal{B}$ and $\mathcal{A}$ respectively, become part of region $\mathcal{C}$. We define the

regions as $\mathcal{A}' = \mathcal{A} \setminus (\Delta_1 \cup \Delta_3)$, $\mathcal{B}' = (\mathcal{B} \setminus \Delta_2) \cup \Delta_1$, and $\mathcal{C}' = \mathcal{C} \cup \Delta_2 \cup \Delta_3$.

Figure 3.4: g Increases



Suppose a seller a is born in region Δ_1 with the expected return for the H-type seller being g, and for the L-type seller being b. When we increase g to g', the expected return for the H-type seller becomes $\frac{b}{1-\phi}$, where $g < \frac{b}{1-\phi} < g'$. Meanwhile, the expected return for the L-type seller remains at b. Thus, during this transition, as g increases, the expected return for the H-type seller increases while that for the L-type seller remains the same.

We then consider a seller a born in region Δ_2, with the expected return for the H-type seller being $\frac{b}{1-\phi}$ and for the L-type seller being b. If we increase g to g', the updated expected return at the undefeated pooling equilibrium for the H-type seller becomes $\lambda g' + (1-\lambda)\, b > \frac{b}{1-\phi}$, and that of the L-type seller becomes $(1-\phi)\,(\lambda g' + (1-\lambda)\, b) > b$. Thus, as g increases,

both types of sellers in region Δ_2 would benefit.

Similarly, consider a seller a born in region Δ_3, with the expected return for the H-type seller being g and for the L-type seller being b. If we increase the value of g to g', the updated expected return at the undefeated pooling equilibrium becomes $\lambda g' + (1-\lambda)\, b > \frac{b}{1-\phi} > g$, while that of the L-type seller becomes $(1-\phi)\,(\lambda g' + (1-\lambda)\, b) > b$. Thus, as g increases, both types of sellers in region Δ_3 would be better off.

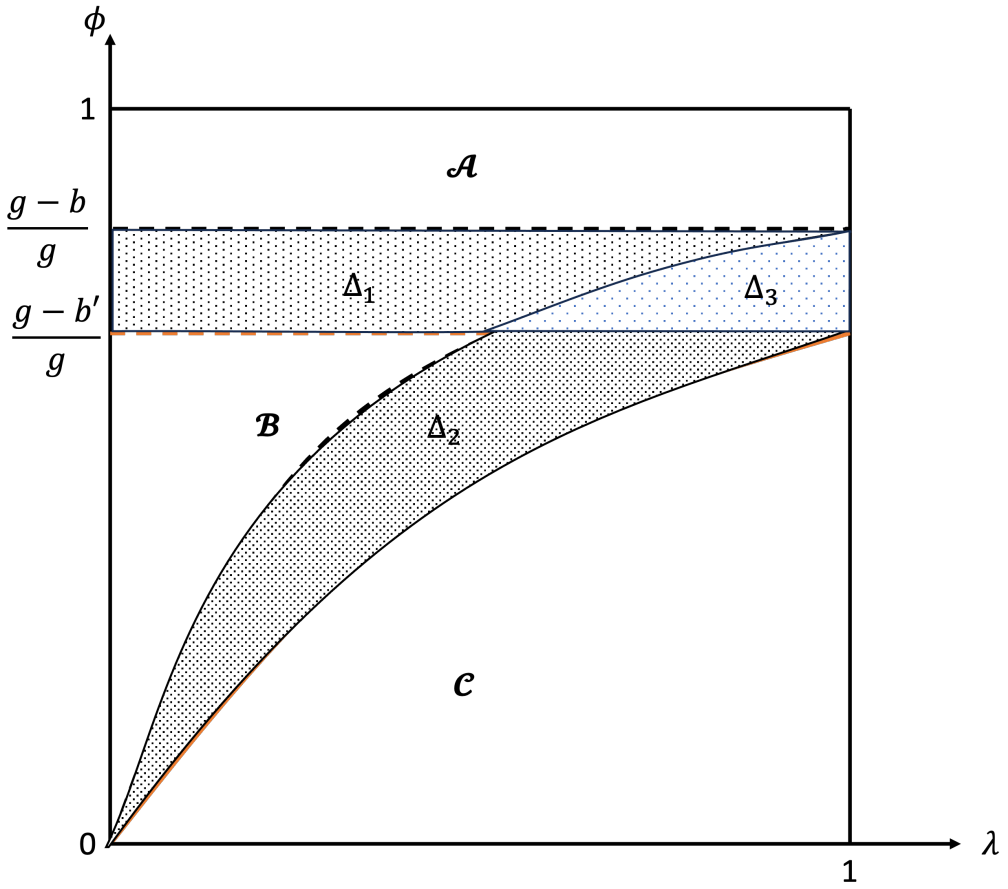
Suppose we increase g to g'. For a seller a remaining in region $\mathcal{A}'$, the expected return for the H-type seller becomes $g' > g$, while for the L-type seller, it remains b. For a seller a staying in region $\mathcal{C}'$, the expected return for both types of sellers increases. This is because $\frac{\partial \lambda g + (1-\lambda) b}{\partial g} = \lambda > 0$, and $\frac{\partial (1-\phi)(\lambda g + (1-\lambda) b)}{\partial g} = (1-\phi)\lambda > 0$. For a seller a remaining in region $\mathcal{B}'$, the expected return for both types of sellers remains unchanged.

In conclusion, as g increases, the expected return for both types of sellers experiences a weak increase.

The final part of Proposition 1 indicates that both types of sellers would benefit from an increase in the product quality of the L-type seller. We illustrate these changes using Figure 3.5. Suppose we increase the value of b to b', where $b' > b$. Consequently, the regions Δ_1 and Δ_3, previously within regions $\mathcal{B}$ and $\mathcal{C}$ respectively, become part of region $\mathcal{A}$. Simultaneously, the region Δ_2, previously within region $\mathcal{C}$, becomes part of region $\mathcal{B}$. We define the updated regions as $\mathcal{A}' = \mathcal{A} \cup \Delta_1 \cup \Delta_3$, $\mathcal{B}' = (\mathcal{B} \setminus \Delta_1) \cup \Delta_2$, and $\mathcal{C}' = \mathcal{C} \setminus (\Delta_2 \cup \Delta_3)$.

Suppose a seller a is born in region Δ_1. If the seller is of the H-type, their expected return at the undefeated equilibrium is $\frac{b}{1-\phi}$, while for the L-type seller, it remains b. If we increase the value of b to b', the updated expected return at the undefeated equilibrium for the H-type seller becomes $g > \frac{b}{1-\phi}$, and for the L-type seller, it becomes $b' > b$. Thus, both types of sellers experience a benefit from the increase in b.

Figure 3.5: b Increases



We consider a seller a born in region Δ_3, with the expected return for the H-type seller being $\lambda g + (1-\lambda)\, b$ and for the L-type seller being $(1-\phi)\,(\lambda g + (1-\lambda)\, b)$. If we increase b to b', the expected return for the H-type seller becomes $g \geq \lambda g + (1-\lambda)\, b$, and for the L-type seller it becomes b'. Based on the discussion in Section 3.2.1, one can deduce that in this region $\frac{b'}{1-\phi} > \lambda g + (1-\lambda)\, b' > \lambda g + (1-\lambda)\, b$, which also implies $b' > (1-\phi)\,(\lambda g + (1-\lambda)\, b)$. Thus, both types of sellers benefit from the increase in b.

We now consider a seller a born in region Δ_2, where the expected return for the H-type is $\lambda g + (1-\lambda)\, b$, and for the L-type seller, it's $(1-\phi)\,(\lambda g + (1-\lambda)\, b)$. Suppose we increase the value of b to b'. The new equilibrium expected return for the H-type seller becomes $\frac{b'}{1-\phi}$, and for the L-type seller, it's b'. As discussed previously, we have $b' > (1-\phi)\,(\lambda g + (1-\lambda)\, b)$,

indicating the L-type seller benefits from the increase. Also, based on the previous discussion in Section 3.2.1, we know that in this region, $\frac{b'}{1-\phi} > \lambda g + (1-\lambda)\, b'$, implying $\frac{b'}{1-\phi} > \lambda g + (1-\lambda)\, b$. Thus, both types of sellers benefit from the increase.

Suppose we increase b to b'. For a seller a in region $\mathcal{A}'$, the expected return for the H-type seller remains g, while for the L-type seller it increases to $b' > b$. In region $\mathcal{C}'$, the expected return for both types of sellers increases. This is because $\frac{\partial \lambda g + (1-\lambda) b}{\partial b} = 1 - \lambda \geq 0$, and $\frac{\partial (1-\phi)(\lambda g + (1-\lambda) b)}{\partial b} = (1-\phi)\,(1-\lambda) \geq 0$. Similarly, in region $\mathcal{B}'$, the expected return for both types of sellers increases since $\frac{\partial \frac{b}{1-\phi}}{\partial b} = \frac{1}{1-\phi} \geq 0$, and $\frac{\partial b}{\partial b} = 1 > 0$.

Hence, as b increases, both the H-type and L-type sellers are better off.

3.4 Conclusion

Our model extends the signal game by incorporating the review score, a significant aspect in online markets where customers rely on both price and customer scores to gauge seller quality. Additionally, we introduce a probability factor representing the likelihood of the seller's quality being fully disclosed to the public. This aspect becomes particularly pertinent in scenarios where government interventions may occur to investigate and reveal product quality information.

We have discovered that when the probability of quality revelation is low, the pooling equilibrium emerges as the undefeated equilibrium under conditions of sufficiently high customer reviews. Conversely, when customer reviews are not favorable, the separating or semi-pooling equilibrium prevails. However, in instances where the probability of type revelation is high, the separating equilibrium becomes the undefeated equilibrium. This is due to the significant cost incurred by low-quality sellers in attempting to mimic high-quality sellers.

We have showed the comparative statics of a seller's expected payoff concerning the probability of type revelation, the initial customer review rate, the quality of the high-quality seller's product, and the quality of the low-quality seller's product. We observe that the H-type seller benefits from an increase in the probability of type revelation, whereas the L-type seller experiences a loss. We also demonstrate that both types of sellers benefit from higher customer reviews, as well as from an increase in the quality of both high-quality and low-quality seller's products.

Bibliography

Ahamada, Ibrahim, and Emmanuel Flachaire. 2010. *Non-Parametric Econometrics.* Oxford: Oxford University Press.

Albert, J., and S. Chib. 1993. "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association*, 88: 669–679.

Amaratunga, Dhammika, and Javier Cabrera. 2001. "Outlier Resistance, Standardization, and Modeling Issues for DNA Microarray Data." *Statistics in Genetics and in the Environmental Sciences*, 17–26.

Andrews, D. F., and C. L. Mallows. 1974. "Scale Mixtures of Normal Distributions." *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(1): 99–102.

Antoniak, Charles E. 1974. "Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems." *The Annals of Statistics*, 1152–1174.

An, Weihua. 2010. "Bayesian Propensity Score Estimators: Incorporating Uncertainties in Propensity Scores into Causal Inference." *Sociological Methodology*, 40(1): 151–189.

Austin, Peter C. 2011. "Optimal Caliper Widths for Propensity-Score Matching When Estimating Differences in Means and Differences in Proportions in Observational Studies." *Pharmaceutical Statistics*, 10(2): 150–161.

Bagwell, Kyle, and Michael H Riordan. 1991. "High and Declining Prices Signal Product Quality." *The American Economic Review*, 224–239.

Bajaj, Ayushi. 2018. "Undefeated Equilibria of the Shi–Trejos–Wright Model under Adverse Selection." *Journal of Economic Theory*, 176: 957–986.

Bar-Isaac, Heski. 2003. "Reputation and Survival: Learning in a Dynamic Signalling Model." *The Review of Economic Studies*, 70(2): 231–251.

Bose, Subir, Gerhard Orosel, Marco Ottaviani, and Lise Vesterlund. 2006. "Dynamic Monopoly Pricing and Herding." *The RAND Journal of Economics*, 37(4): 910–928.

Brand, Jennie E, and Charles N Halaby. 2006. "Regression and Matching Estimates of the Effects of Elite College Attendance on Educational and Career Achievement." *Social Science Research*, 35(3): 749–770.

Branson, Zach, Maxime Rischard, Luke Bornn, and Luke W Miratrix. 2019. "A Nonparametric Bayesian Methodology for Regression Discontinuity Designs." *Journal of Statistical Planning and Inference*, 202: 14–30.

Caliendo, Marco, and Sabine Kopeinig. 2008. "Some Practical Guidance for the Implementation of Propensity Score Matching." *Journal of Economic Surveys*, 22(1): 31–72.

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik. 2014*a*. "Robust Data-Driven Inference in the Regression-Discontinuity Design." *The Stata Journal*, 14(4): 909–946.

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik. 2014*b*. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica*, 82(6): 2295–2326.

Calonico, Sebastian, Matias D Cattaneo, Max H Farrell, and Rocio Titiunik. 2017. "Rdrobust: Software for Regression-Discontinuity Designs." *The Stata Journal*, 17(2): 372–404.

Cardon, James H, and Igal Hendel. 2001. "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey." *RAND Journal of Economics*, 408–427.

Cattaneo, Matias D., Brigham R. Frandsen, and Rocío Titiunik. 2015. "Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the U.S. Senate." *Journal of Causal Inference*, 3(1): 1–24.

Celeux, Gilles. 1998. "Bayesian Inference for Mixture: The Label Switching Problem." 227–232. Heidelberg:Physica-Verlag HD.

Chan, David, Robert Kohn, David Nott, and Chris Kirby. 2006. "Locally Adaptive Semiparametric Estimation of the Mean and Variance Functions in Regression Models." *Journal of Computational and Graphical Statistics*, 15(4): 915–936.

Chan, Joshua CC, and Ivan Jeliazkov. 2009*a*. "Efficient Simulation and Integrated Likelihood Estimation in State Space Models." *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1-2): 101–120.

Chan, Joshua C. C., and Justin L. Tobias. 2015. "Priors and Posterior Computation in Linear Endogenous Variable Models with Imperfect Instruments." *Journal of Applied Econometrics*, 30(4): 650–674.

Chan, Joshua Chi-Chun, and Ivan Jeliazkov. 2009*b*. "MCMC Estimation of Restricted Covariance Matrices." *Journal of Computational and Graphical Statistics*, 18(2): 457–480.

Chaudhuri, Kausik, and Peter Howley. 2022. "The Impact of COVID-19 Vaccination for Mental Well-Being." *European Economic Review*, 150: 104293.

Chesnaye, Nicholas C, Vianda S Stel, Giovanni Tripepi, Friedo W Dekker, Edouard L Fu, Carmine Zoccali, and Kitty J Jager. 2022. "An introduction to Inverse Probability of Treatment Weighting in Observational Research." *Clinical Kidney Journal*, 15(1): 14–20.

Chib, Siddhartha. 1995. "Marginal Likelihood from the Gibbs Output." *Journal of the American Statistical Association*, 90(432): 1313–1321.

Chib, Siddhartha. 2007. "Analysis of Treatment Response Data without the Joint Distribution of Potential Outcomes." *Journal of Econometrics*, 140(2): 401–412.

Chib, Siddhartha, and Barton H. Hamilton. 2000. "Bayesian Analysis of Cross-Section and Clustered Data Treatment Models." *Journal of Econometrics*, 97(1): 25–50.

Chib, Siddhartha, and Barton H. Hamilton. 2002. "Semiparametric Bayes Analysis of Longitudinal Data Treatment Models." *Journal of Econometrics*, 110(1): 67–89.

Chib, Siddhartha, and Edward Greenberg. 2010. "Additive Cubic Spline Regression with Dirichlet Process Mixture Errors." *Journal of Econometrics*, 156(2): 322–336.

Chib, Siddhartha, and Ivan Jeliazkov. 2001. "Marginal Likelihood from the Metropolis–Hastings Output." *Journal of the American Statistical Association*, 96(453): 270–281.

Chib, Siddhartha, and Ivan Jeliazkov. 2006. "Inference in Semiparametric Dynamic Models for Binary Longitudinal Data." *Journal of the American Statistical Association*, 101(474): 685–700.

Chib, Siddhartha, and Liana Jacobi. 2016. "Bayesian Fuzzy Regression Discontinuity Analysis and Returns to Compulsory Schooling." *Journal of Applied Econometrics*, 31(6): 1026–1047.

Chib, Siddhartha, Edward Greenberg, and Anna Simoni. 2023. "Nonparametric Bayes Analysis of the Sharp and Fuzzy Regression Discontinuity Designs." *Econometric Theory*, 39(3): 481–533.

Chib, Siddhartha, Edward Greenberg, and Ivan Jeliazkov. 2009. "Estimation of Semiparametric Models in the Presence of Endogeneity and Sample Selection." *Journal of Computational and Graphical Statistics*, 18(2): 321–348.

Chib, Siddhartha, Edward Greenberg, and Rainer Winkelmann. 1998. "Posterior Simulation and Bayes Factors in Panel Count Data Models." *Journal of Econometrics*, 86(1): 33–54.

Cutler, David M., and Richard J. Zeckhauser. 1998. "Adverse Selection in Health Insurance." *Forum for Health Economics & Policy*, 1(1).

Deb, Partha, Murat K. Munkin, and Pravin K. Trivedi. 2006. "Bayesian Analysis of the Two-Part Model with Endogeneity: Application to Health Care Expenditure." *Journal of Applied Econometrics*, 21(7): 1081–1099.

Dehejia, Rajeev H, and Sadek Wahba. 1999. "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs." *Journal of the American Statistical Association*, 94(448): 1053–1062.

Delacroix, Alain, and Shouyong Shi. 2013. "Pricing and Signaling with Frictions." *Journal of Economic Theory*, 148(4): 1301–1332.

Dong, Yingying. 2015. "Regression Discontinuity Applications with Rounding Errors in the Running Variable." *Journal of Applied Econometrics*, 30(3): 422–446.

Dong, Yingying. 2019. "Regression Discontinuity Designs with Sample Selection." *Journal of Business & Economic Statistics*, 37(1): 171–186.

Dong, Yingying, and Arthur Lewbel. 2015. "Identifying the Effect of Changing the Policy Threshold in Regression Discontinuity Models." *The Review of Economics and Statistics*, 97(5): 1081–1092.

Dong, Yingying, Ying-Ying Lee, and Michael Gou. 2023. "Regression Discontinuity Designs with a Continuous Treatment." *Journal of the American Statistical Association*, 118(541): 208–221.

Dreze, J. H., and J.-F. Richard. 1983. "Bayesian Analysis of Simultaneous Equation Systems." *Handbook of Econometrics*, 1: 517–598.

Duan, Zhizhou, Yi Feng, Shicun Xu, Desheng Gao, Yan Ji, Xi Sun, Runsen Chen, and Yuanyuan Wang. 2023. "The Role of Childhood Left-behind Experience on Childhood Trauma Exposure and Mental Health Outcomes: A Propensity Score Matching (PSM) Analysis." *Journal of Public Health*, 45(4): 785–793.

Fahrmeir, L., and S. Lang. 2001. "Bayesian Inference for Generalized Additive Mixed Models Based on Markov Random Field Priors." *Journal of the Royal Statistical Society – Series C*, 50: 201–220.

Ferguson, Thomas S. 1973. "A Bayesian Analysis of Some Nonparametric Problems." *The Annals of Statistics*, 209–230.

Ferman, Bruno, and Cristine Pinto. 2019. "Inference in Differences-in-Differences with Few Treated Groups and Heteroskedasticity." *The Review of Economics and Statistics*, 101(3): 452–467.

Fishman, Michael J, and Kathleen M Hagerty. 2003. "Mandatory versus Voluntary Disclosure in Markets with Informed and Uninformed Customers." *Journal of Law, Economics, and Organization*, 19(1): 45–63.

Fletcher, Jason M, and Mansur Tokmouline. 2018. "The Effects of Academic Probation on College Success: Regression Discontinuity Evidence from Four Texas Universities." *Available at SSRN 3097330*.

Frühwirth-Schnatter, Sylvia. 2001. "Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models." *Journal of the American Statistical Association*, 96(453): 194–209.

Frühwirth-Schnatter, Sylvia. 2006. *Finite Mixture and Markov Switching Models.* Springer.

Frühwirth-Schnatter, Sylvia. 2011. "Dealing with Label Switching under Model Uncertainty." *Mixtures: Estimation and Applications*, 213–239.

Frühwirth-Schnatter, Sylvia, and Sylvia Kaufmann. 2008. "Model-Based Clustering of Multiple Time Series." *Journal of Business & Economic Statistics*, 26(1): 78–89.

Gamerman, Dani. 1997. "Sampling from the Posterior Distribution in Generalized Linear Mixed Models." *Statistics and Computing*, 7(1): 57–68.

Gelman, Andrew, and Guido Imbens. 2019. "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs." *Journal of Business & Economic Statistics*, 37(3): 447–456.

Geneletti, Sara, Federico Ricciardi, Aidan G O'Keeffe, and Gianluca Baio. 2019. "Bayesian Modelling for Binary Outcomes in the Regression Discontinuity Design." *Journal of the Royal Statistical Society Series A: Statistics in Society*, 182(3): 983–1002.

Gu, Yuanyuan, Denzil G Fiebig, Edward Cripps, and Robert Kohn. 2009. "Bayesian Estimation of a Random Effects Heteroscedastic Probit Model." *The Econometrics Journal*, 12(2): 324–339.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica*, 69(1): 201–209.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica*, 47(1): 153–161.

Heckman, James J, Hedibert F Lopes, and Rémi Piatek. 2014. "Treatment Effects: A Bayesian Perspective." *Econometric Reviews*, 33(1-4): 36–67.

Imai, Kosuke, and David A van Dyk. 2004. "Causal Inference with General Treatment Regimes." *Journal of the American Statistical Association*, 99(467): 854–866.

Jacobi, Liana, Helga Wagner, and Sylvia Frühwirth-Schnatter. 2016. "Bayesian Treatment Effects Models with Variable Selection for Panel Outcomes with an Application to Earnings Effects of Maternity Leave." *Journal of Econometrics*, 193(1): 234–250.

Jeliazkov, Ivan. 2013. "Nonparametric Vector Autoregressions: Specification, Estimation, and Inference." In *VAR Models in Macroeconomics–New Developments and Applications: Essays in Honor of Christopher A. Sims.* Emerald Group Publishing Limited.

Jeliazkov, Ivan, and Angela Vossmeyer. 2018. "The Impact of Estimation Uncertainty on Covariate Effects in Nonlinear Models." *Statistical Papers*, 59.

Jeliazkov, Ivan, and Esther Hee Lee. 2010. "MCMC Perspectives on Simulated Likelihood Estimation." In *Maximum Simulated Likelihood Methods and Applications*. Emerald Group Publishing Limited.

Jiang, Baojun, and Bicheng Yang. 2019. "Quality and Pricing Decisions in a Market with Consumer Information Sharing." *Management Science*, 65(1): 272–285.

Jiang, Baojun, Lin Tian, Yifan Xu, and Fuqiang Zhang. 2016. "To Share or Not to Share: Demand Forecast Sharing in a Distribution Channel." *Marketing Science*, 35(5): 800–809.

Kass, Robert E, and Adrian E Raftery. 1995. "Bayes Factors." *Journal of the American Statistical Association*, 90(430): 773–795.

Keane, Michael, and Olena Stavrunova. 2016. "Adverse Selection, Moral Hazard and the Demand for Medigap Insurance." *Journal of Econometrics*, 190(1): 62–78.

Kline, Brendan, and Justin L. Tobias. 2008. "The Wages of BMI: Bayesian Analysis of a Skewed Treatment–Response Model with Nonparametric Endogeneity." *Journal of Applied Econometrics*, 23(6): 767–793.

Koop, G., and D. J. Poirier. 2004. "Bayesian Variants of Some Classical Semiparametric Regression Techniques." *Journal of Econometrics*, 123: 259–282.

Koop, Gary, and Justin L. Tobias. 2004. "Learning about Heterogeneity in Returns to Schooling." *Journal of Applied Econometrics*, 19(7): 827–849.

Koop, Gary, Dale J. Poirier, and Justin Tobias. 2005. "Semiparametric Bayesian Inference in Multiple Equation Models." *Journal of Applied Econometrics*, 20(6): 723–747.

Li, Guo, Lin Tian, and Hong Zheng. 2021. "Information Sharing in an Online Marketplace with Co-opetitive Sellers." *Production and Operations Management*, 30(10): 3713–3734.

Li, Mingliang, and Justin L Tobias. 2008. "Bayesian Analysis of Treatment Effects in an Ordered Potential Outcomes Model." In *Modelling and Evaluating Treatment Effects in Econometrics*. Vol. 21, 57–91. Emerald Group Publishing Limited.

Li, Mingliang, and Justin L. Tobias. 2011. "Bayesian Inference in a Correlated Random Coefficients Model: Modeling Causal Effect Heterogeneity with an Application to Heterogeneous Returns to Schooling." *Journal of Econometrics*, 162(2): 345–361.

Li, Mingliang, and Justin L Tobias. 2014. "Bayesian Analysis of Treatment Effect Models." *Bayesian Inference in the Social Sciences*, 63–90.

Li, Mingliang, Dale J Poirier, and Justin L Tobias. 2004. "Do Dropouts Suffer from Dropping out? Estimation and Prediction of Outcome Gains in Generalized Selection Models." *Journal of Applied Econometrics*, 19(2): 203–225.

Mailath, George J, Masahiro Okuno-Fujiwara, and Andrew Postlewaite. 1993. "Belief-Based Refinements in Signalling Games." *Journal of Economic Theory*, 60(2): 241–276.

Mastrobuoni, Giovanni, Franco Peracchi, and Aleksey Tetenov. 2014. "Price as a Signal of Product Quality: Some Experimental Evidence." *Journal of Wine Economics*, 9(2): 135–152.

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics*, 142(2): 698–714.

Milgrom, Paul, and John Roberts. 1986. "Price and Advertising Signals of Product Quality." *Journal of Political Economy*, 94(4): 796–821.

Mintz, Ofer, Imran S. Currim, and Ivan Jeliazkov. 2013. "Information Processing Pattern and Propensity to Buy: An Investigation of Online Point-of-Purchase Behavior." *Marketing Science*, 32(5): 716–732.

Munkin, Murat K. 2003. "The MCMC and SML Estimation of a Self-Selection Model with Two Outcomes." *Computational Statistics and Data Analysis*, 42(3): 403–424. Computational Ecomometrics.

Munkin, Murat K. 2011. "The Endogenous Sequential Probit Model: An Application to the Demand for Hospital Utilization." *Economics Letters*, 112(2): 182–185.

Munkin, Murat K., and Pravin K. Trivedi. 1999. "Simulated Maximum Likelihood Estimation of Multivariate Mixed-Poisson Regression Models, with Application." *The Econometrics Journal*, 2(1): 29–48.

Munkin, Murat K., and Pravin K. Trivedi. 2003. "Bayesian Analysis of a Self-Selection Model with Multiple Outcomes Using Simulation-Based Estimation: An Application to the Demand for Healthcare." *Journal of Econometrics*, 114(2): 197–220.

Nghiem, Son, and Nicholas Graves. 2019. "Selection Bias and Moral Hazard in the Australian Private Health Insurance Market: Evidence from the Queensland Skin Cancer Database." *Economic Analysis and Policy*, 64: 259–265.

Nott, David J, and Daniela Leonte. 2004. "Sampling Schemes for Bayesian Variable Selection in Generalized Linear Models." *Journal of Computational and Graphical Statistics*, 13(2): 362–382.

Panagiotelis, Anastasios, and Michael Smith. 2008. "Bayesian Identification, Selection and Estimation of Semiparametric Functions in High-Dimensional Additive Models." *Journal of Econometrics*, 143(2): 291–316.

Poirier, Dale J. 1973. "Piecewise Regression Using Cubic Spline." *Journal of the American Statistical Association*, 68(343): 515–524.

Reschovsky, James D, Peter Kemper, and Ha Tu. 2000. "Does Type of Health Insurance Affect Health Care Use and Assessments of Care among the Privately Insured?" *Health Services Research*, 35(1 Pt 2): 219.

Ritter, Christian, and Martin A Tanner. 1992. "Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler." *Journal of the American Statistical Association*, 87(419): 861–868.

Rosenbaum, Paul R. 1987. "Model-Based Direct Adjustment." *Journal of the American Statistical Association*, 82(398): 387–394.

Rosenbaum, Paul R, and Donald B Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika*, 70(1): 41–55.

Rosenbaum, Paul R, and Donald B Rubin. 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score." *The American Statistician*, 39(1): 33–38.

Rosenbaum, P R, and D B Rubin. 2022. "Propensity Scores in the Design of Observational Studies for Causal Effects." *Biometrika*, 110(1): 1–13.

Roy, Andrew Donald. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers*, 3(2): 135–146.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology*, 66(5): 688.

Rubin, Donald B. 1977. "Assignment to Treatment Group on the Basis of a Covariate." *Journal of Educational Statistics*, 2(1): 1–26.

Rubin, Donald B. 1978. "Bayesian Inference for Causal Effects: The Role of Randomization." *The Annals of Statistics*, 34–58.

Rubin, Donald B. 2004. "Direct and Indirect Causal Effects via Potential Outcomes." *Scandinavian Journal of Statistics*, 31(2): 161–170.

Rubin, Donald B. 2005. "Causal Inference Using Potential Outcomes." *Journal of the American Statistical Association*, 100(469): 322–331.

Rue, Havard, and Leonhard Held. 2005. *Gaussian Markov Random Fields: Theory and Applications.* CRC press.

Ruppert, David, M. P. Wand, and R. J. Carroll. 2003. *Semiparametric Regression. Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press.

Sapelli, Claudio, and Bernardita Vial. 2003. "Self-Selection and Moral Hazard in Chilean Health Insurance." *Journal of Health Economics*, 22(3): 459–476.

Schwarz, Gideon E. 1978. "Estimating the Dimension of a Model." *Annals of Statistics*, 6(2): 461–464.

Shiller, R. 1984. "Smoothness Priors and Nonlinear Regression." *Journal of the American Statistical Association*, 79: 609–615.

Taylor, Curtis R. 1999. "Time-on-the-Market as a Sign of Quality." *The Review of Economic Studies*, 66(3): 555–578.

Thistlethwaite, Donald L, and Donald T Campbell. 1960. "Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment." *Journal of Educational Psychology*, 51(6): 309.

Thoemmes, Felix. 2012. "Propensity Score Matching in SPSS."

University of Essex, Institute for Social and Economic Research. 2021. "Understanding Society: COVID-19 Study, 2020-2021." *dataset*, Accessed on 31 October 2023.

van Hasselt, Martjn. 2014. "Bayesian Analysis of Sample Selection Models." *Bayesian Inference in the Social Sciences*, 91–113.

Verlinda, J. A. 2006. "A Comparison of Two Common Approaches for Estimating Marginal Effects in Binary Choice Models." *Applied Economics Letters*, 13: 77–80.

Vossmeyer, Angela. 2014*a*. "Determining the Proper Specification for Endogeneous Covariates in Discrete Data Settings." *Advances in Econometrics*, 34: 223–247.

Vossmeyer, Angela. 2014*b*. "Treatment Effects and Informative Missingness with an Application to Bank Recapitalization Programs." *American Economic Review*, 104(5): 212–17.

Vossmeyer, Angela. 2016. "Sample Selection and Treatment Effect Estimation of Lender of Last Resort Policies." *Journal of Business & Economic Statistics*, 34(2): 197–212.

Williams, C. K. I. 1998. "Prediction with Gaussian Processes: From Linear Regression to Linear Prediction and Beyond." In *Learning in Graphical Models.* , ed. Michael I. Jordan, 599–621. Springer Netherlands.

Wolinsky, Asher. 1983. "Prices as Signals of Product Quality." *Review of Economic Studies*, 50(4): 647–658.

Wright, Nicholas A. 2020. "Perform Better, or Else: Academic Probation, Public praise, and Students Decision-Making." *Labour Economics*, 62: 101773.

Wu, Yue, Kaifu Zhang, and Jinhong Xie. 2020. "Bad Greenwashing, Good Greenwashing: Corporate Social Responsibility and Information Transparency." *Management Science*, 66(7): 3095–3112.

Zhang, Zhongheng, Hwa Jung Kim, Guillaume Lonjon, and Yibing Zhu. 2019. "Balance Diagnostics after Propensity Score Matching." *Annals of Translational Medicine*, 7(1).

Zhao, Shandong, David A van Dyk, and Kosuke Imai. 2020. "Propensity Score-Based Methods for Causal Inference in Observational Studies with Non-binary Treatments." *Statistical Methods in Medical Research*, 29(3): 709–727.

Zhao, Zhong. 2008. "Sensitivity of Propensity Score Methods to the Specifications." *Economics Letters*, 98(3): 309–319.

Appendix A

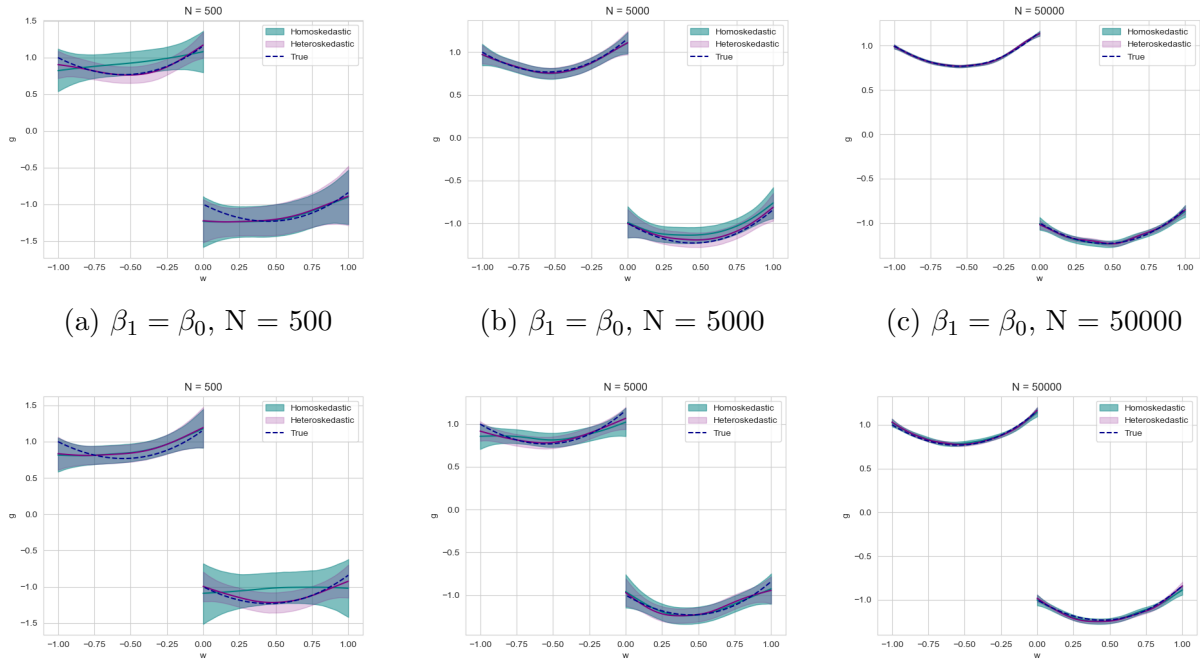
Supplementary Material for Chapter 1

A.1 Additional Simulation Results for Sharp RDD

Table A.1: Prior Sensitivity Analysis

n	**Model**	Δ_{SRD}	$\hat{\Delta}_{SRD}$	**SD**	**95% CI**
500	Homosk.	-2.24	-2.37	0.25	$(-2.87, -1.88)$
	Heterosk.	-2.24	-2.41	0.23	$(-2.87, -1.97)$
	RDRobust	-2.24	-2.63	0.68	$(-4.18, -1.07)$
5000	Homosk.	-2.22	-2.04	0.13	$(-2.29, -1.79)$
	Heterosk.	-2.22	-2.09	0.11	$(-2.30, -1.88)$
	RDRobust	-2.22	-2.19	0.21	$(-2.71, -1.72)$
50000	Homosk.	-2.21	-2.20	0.04	$(-2.28, -2.12)$
	Heterosk.	-2.21	-2.20	0.03	$(-2.26, -2.14)$
	RDRobust	-2.21	-2.06	0.07	$(-2.20, -1.88)$

Figure A.1: Estimated Functions $\hat{g}$ with Continuous Outcome Variable



(a) $\beta_1 = \beta_0$, N = 500 (b) $\beta_1 = \beta_0$, N = 5000 (c) $\beta_1 = \beta_0$, N = 50000

(d) $\beta_1 \neq \beta_0$, N = 500 (e) $\beta_1 \neq \beta_0$, N = 5000 (f) $\beta_1 \neq \beta_0$, N = 50000

A.2 Fuzzy RDD

Here we provide details behind the specification of a heteroskedastic fuzzy RD model, outline the corresponding estimation algorithm, and offer a simulation study of its performance. In the fuzzy RD setting, the combinations of treatments and unit types in the fuzzy RD case are summarized in Table A.2. A general model that allows for category-specific functions,

Table A.2: Fuzzy RD Data

	$w < w*$	$w \geq w*$
$T = 0$	$\mathcal{C}, \mathcal{N}$	$\mathcal{N}$
$T = 1$	$\mathcal{A}$	$\mathcal{C}, \mathcal{A}$

parameters, and heteroskedasticity relationships is specified as

$$
\begin{aligned}
&s=\mathcal{C}: T=\mathbb{1}\{w \geq w^*\}, \quad y_{ij}=g_j\left(w_i\right)+x_i'\beta_j+\varepsilon_{ij}, \quad \varepsilon_{ij} \sim N\left(0, \sigma_{ij}^2\right), \quad \ln\left(\sigma_{ij}^2\right)=Z_{ij}'\gamma_j, \\
&s=\mathcal{A}: T=1, \quad y_{i\mathcal{A}}=g_{\mathcal{A}}\left(w_i\right)+x_i'\beta_{\mathcal{A}}+\varepsilon_{i\mathcal{A}}, \quad \varepsilon_{i\mathcal{A}} \sim N\left(0, \sigma_{i\mathcal{A}}^2\right), \quad \ln\left(\sigma_{i\mathcal{A}}^2\right)=Z_{i\mathcal{A}}'\gamma_{\mathcal{A}}, \\
&s=\mathcal{N}: T=0, \quad y_{i\mathcal{N}}=g_{\mathcal{N}}\left(w_i\right)+x_i'\beta_{\mathcal{N}}+\varepsilon_{i\mathcal{N}}, \quad \varepsilon_{i\mathcal{N}} \sim N\left(0, \sigma_{i\mathcal{N}}^2\right), \quad \ln\left(\sigma_{i\mathcal{N}}^2\right)=Z_{i\mathcal{N}}'\gamma_{\mathcal{N}}, \\
&\Pr(s)=q_s>0 \text{ for } s \in \{\mathcal{C}, \mathcal{A}, \mathcal{N}\} \text{ and } q_{\mathcal{C}}+q_{\mathcal{A}}+q_{\mathcal{N}}=1, \quad j=0,1.
\end{aligned}
\tag{A.1}
$$

The RD ATE is defined as

$$
\begin{aligned}
\Delta_{FRD} &\equiv \lim_{z \downarrow \tau^+} E\left(Y_1 | w, x_i, s=\mathcal{C}\right)-\lim_{z \uparrow \tau^-} E\left(Y_0 | w, x_i, s=\mathcal{C}\right) \\
&=\lim_{w \downarrow w^{*+}} E\left(g_1(w)+x_i'\beta_1\right)-\lim_{w \uparrow w^{*-}} E\left(g_0(w)+x_i'\beta_0\right).
\end{aligned}
$$

The likelihood function is expressed as

$$
\begin{aligned}
L=&\prod_{i:T_i=0, w_i<w^*}\left(q_{\mathcal{C}} f_N\left(y_i | g_0(w_i)+x_i'\beta_0, \sigma_{i0}^2\right)+q_{\mathcal{N}} f_N\left(y_i | g_{\mathcal{N}}(w_i)+x_i'\beta_{\mathcal{N}}, \sigma_{i\mathcal{N}}^2\right)\right) \\
&\prod_{i:T_i=0, w_i \geq w^*} q_{\mathcal{N}} f_N\left(y_i | g_{\mathcal{N}}(w_i)+x_i'\beta_{\mathcal{N}}, \sigma_{i\mathcal{N}}^2\right) \prod_{i:T_i=1, w_i<w^*} q_{\mathcal{A}} f_N\left(y_i | g_{\mathcal{A}}(w_i)+x_i'\beta_{\mathcal{A}}, \sigma_{i\mathcal{A}}^2\right) \\
&\prod_{i:T_i=1, w_i \geq w^*}\left(q_{\mathcal{C}} f_N\left(y_i | g_1(w_i)+x_i'\beta_1, \sigma_{i1}^2\right)+q_{\mathcal{A}} f_N\left(y_i | g_{\mathcal{A}}(w_i)+x_i'\beta_{\mathcal{A}}, \sigma_{i\mathcal{A}}^2\right)\right).
\end{aligned}
$$

The prior distribution of $q=(q_{\mathcal{A}}, q_{\mathcal{N}}, q_{\mathcal{C}})$ is given by $q \sim Dir\left(n_{\mathcal{A}0}, n_{\mathcal{N}0}, n_{\mathcal{C}0}\right)$. All other parameters follow the priors discussed previously. The unknown functions g_s are modeled analogously to Section 1.2 for the different subsets of data. The posterior distribution for the type variables s_i, $i=1, \ldots, n$, is specified as

$$
\begin{aligned}
&\Pr\left(s_i=\mathcal{C} | y_i, g_j, \beta_j, \tau_j^2, \gamma_j\right) \propto q_c f_N\left(y_i | g_i(w_i)+x_i'\beta_j, \sigma_{ij}^2\right), \quad j=0,1, \\
&\Pr\left(s_i=\mathcal{N} | y_i, g_{\mathcal{N}}, \beta_{\mathcal{N}}, \tau_{\mathcal{N}}^2, \gamma_{\mathcal{N}}\right) \propto q_{\mathcal{N}} f_N\left(y_i | g_{\mathcal{N}}(w_i)+x_i'\beta_{\mathcal{N}}, \sigma_{i\mathcal{N}}^2\right), \\
&\Pr\left(s_i=\mathcal{A} | y_i, g_{\mathcal{A}}, \beta_{\mathcal{A}}, \tau_{\mathcal{A}}^2, \gamma_{\mathcal{A}}\right) \propto q_{\mathcal{A}} f_N\left(y_i | g_{\mathcal{A}}(w_i)+x_i'\beta_{\mathcal{A}}, \sigma_{i\mathcal{A}}^2\right).
\end{aligned}
\tag{A.2}
$$

The joint posterior distribution can be sampled as in Algorithm 7.

Algorithm 7 (Semi-parametric Fuzzy RDD)

---

(1) Sample the type variables $\{s_i\}$ from the posterior distribution in Equation (A.2).

(2) Sample $q = (q_{\mathcal{A}}, q_{\mathcal{N}}, q_{\mathcal{C}}) \sim Dir\left(n_{\mathcal{A}0} + n_{\mathcal{A}}, n_{\mathcal{N}0} + n_{\mathcal{N}}, n_{\mathcal{C}0} + n_{\mathcal{C}}\right)$, where $n_{\mathcal{A}}, n_{\mathcal{N}}$ and $n_{\mathcal{C}}$ are the sample sizes of the observations that are categorized into always-takers, never-takers and compliers correspondingly in the previous step.

(3) We update g_j, $j = 0, 1$, using the samples that were categorized as compliers in the previous step. We sample $\left[g_j | y_j, \beta_j, \tau_j^2, \gamma_j\right] \sim N\left(\hat{g}_j, \hat{G}_j\right)$, where $\hat{G}_j = \left(\frac{K_j}{\tau_j^2} + Q_j' \Omega_j^{-1} Q_j\right)^{-1}$ and $\hat{g}_j = \hat{G}_j \left(\frac{1}{\tau_j^2} K_j g_{j0} + Q_j' \Omega_j^{-1} \left(y_j - X_j \beta_j\right)\right)$. We repeat this step for all the compliers and never-takers to sample $[g_{\mathcal{A}} | y_{\mathcal{A}}, \beta_{\mathcal{A}}, \tau_{\mathcal{A}}^2, \gamma_{\mathcal{A}}]$ and $[g_{\mathcal{N}} | y_{\mathcal{N}}, \beta_{\mathcal{N}}, \tau_{\mathcal{N}}^2, \gamma_{\mathcal{N}}]$.

(4) Sample $[\beta_j | y_j, g_j, \gamma_j] \sim N\left(\hat{\beta}_j, \hat{B}_j\right)$, $j = 0, 1$, where $\hat{B}_j = \left(B_{j0}^{-1} + X_j' \Omega_j^{-1} X_j\right)^{-1}$, and $\hat{\beta}_j = \hat{B}_j \left(B_{j0}^{-1} b_{j0} + X_j' \Omega_j^{-1} (y_j - Q_j g_j\right)$. Sample $[\beta_{\mathcal{A}} | y_{\mathcal{A}}, g_{\mathcal{A}}, \gamma_{\mathcal{A}}]$, and $[\beta_{\mathcal{N}} | y_{\mathcal{N}}, g_{\mathcal{N}}, \gamma_{\mathcal{N}}]$ in a similar way.

(5) Sample $\left[\tau_j^2 | g_j\right] \sim IG\left(\frac{\kappa_{j0} + m_j}{2}, \frac{d_{j0} + (g_j - g_{j0})' K_j (g_j - g_{j0})}{2}\right)$, $j = 0, 1$. Repeat this step to sample $[\tau_{\mathcal{A}}^2 | g_{\mathcal{A}}]$ and $[\tau_{\mathcal{N}}^2 | g_{\mathcal{N}}]$.

(6) Sample $[\gamma_j | y_j, g_j, \beta_j]$, $j = 0, 1$, using an MH step by drawing a proposal value $\gamma_j^\dagger \sim q\left(\gamma_j | \hat{\gamma}_j, V_j\right)$, where $\hat{\gamma}_j$ and V_j are computed as in (1.5) using the current value of γ_j. Also use $\gamma_j^\dagger$ in equation (1.5) to produce $\hat{\gamma}_j^\dagger$ and accept the proposed $\gamma_j^\dagger$ with probability

$$\alpha(\gamma_j, \gamma_j^\dagger | y_j, g_j, \beta_j) = \min\left\{1, \frac{f(y_j | g_j, \beta_j, \gamma_j^\dagger) \pi(\gamma_j^\dagger | \gamma_{j0}, \Gamma_{j0})}{f(y_j | g_j, \beta_j, \gamma_j) \pi(\gamma_j | \gamma_{j0}, \Gamma_{j0})} \frac{q(\gamma_j | \hat{\gamma}_j^\dagger, V_j)}{q(\gamma_j^\dagger | \hat{\gamma}_j, V_j)}\right\},$$

otherwise the current value γ_j is repeated in the next MCMC iteration. We repeat this step for always-takers and never-takers to sample $[\gamma_{\mathcal{A}} | y_{\mathcal{A}}, g_{\mathcal{A}}, \beta_{\mathcal{A}}]$ and $[\gamma_{\mathcal{N}} | y_{\mathcal{N}}, g_{\mathcal{N}}, \beta_{\mathcal{N}}]$.

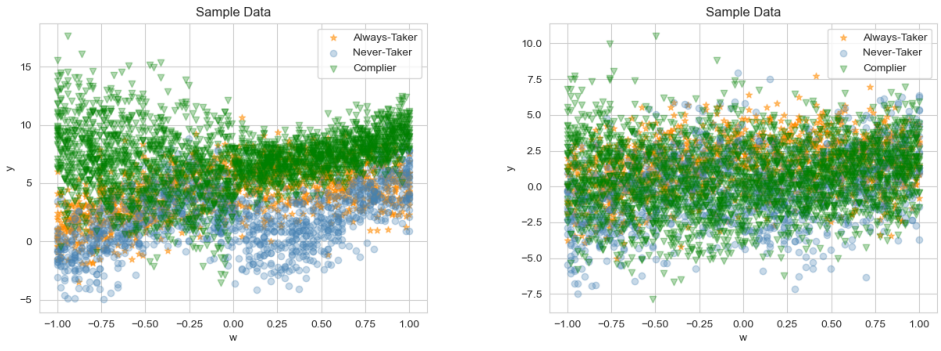
We conducted a simulation study to assess the algorithm's performance and evaluate the influence of heteroskedasticity within the framework of the fuzzy RD model. We sample the data from the model specified in Equation (A.1). The simulated data are visualized in Figure A.2.

In the first sample, the data points for each group exhibit clear distinctions, with parameter values set as follows: $n = 5000$, $q_c = \frac{1}{2}$, $q_a = q_n = \frac{1}{4}$, $\gamma_0 = \gamma_1 = \gamma_a = \gamma_n = (-2, 1)'$, $g_0(w) = 5.5+3w^2$, $g_1(w) = 6.5+3w^2$, $g_{\mathcal{A}}(w) = 2+2w+3\exp(-w^2)$, and $g_{\mathcal{N}}(w) = 2+w+2\sin(-5w)$, along with $\beta_0 = (-1.51, -0.38, -2.34, 0.19, -1.15)'$, $\beta_1 = (0.42, -0.03, -0.24, 0.83, 0.43)'$, $\beta_{\mathcal{A}} = (0.40, -1.17, -0.35, 0.02, -0.97)'$, and $\beta_{\mathcal{N}} = (0.85, 1.24, 0.80, -0.07, -0.57)'$. Covariates x_i are drawn from independent standard normal distributions with a dimension of 5, while covariate z_i comprises a constant term and $||w_i| - 1.5|$.

In the second sample, the data points for each group are intermingled. The parameter values are $\beta_0 = (0.88, 2.10, -0.70, -0.00, -1.00)'$, $\beta_1 = (-0.08, 0.29, 0.42, -1.80, 0.19)'$, $\beta_{\mathcal{A}} = (-0.55, 0.60, 0.11, -1.18, 0.86)'$, $\beta_{\mathcal{N}} = (-0.29, -0.01, 1.68, 0.67, -1.04)'$, $g_0(w) = w^2$, $g_1(w) = 0.5 + w^2$, $g_{\mathcal{A}}(w) = w + 2\exp(-w^2)$, and $g_{\mathcal{N}}(w) = w + \sin(-5w)$. All remaining parameters are kept consistent with those specified for the first sample.

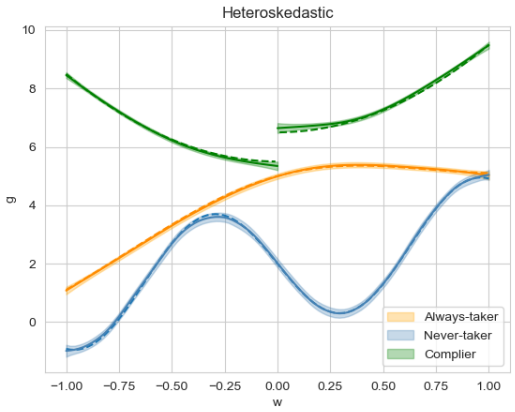
The nonparametric functions estimated for each group are presented in Figures A.3 and A.4. In this context, it is important to highlight the possibility of misclassification and label switching, which can occur when the clusters are not well-separated (Celeux, 1998). Hence, we advise practitioners to carefully examine their results prior to drawing definitive conclusions in applications where clusters are not well-separated, as there is no current consensus solution to the problems present in this context.

Figure A.2: Fuzzy RDD Simulated Data

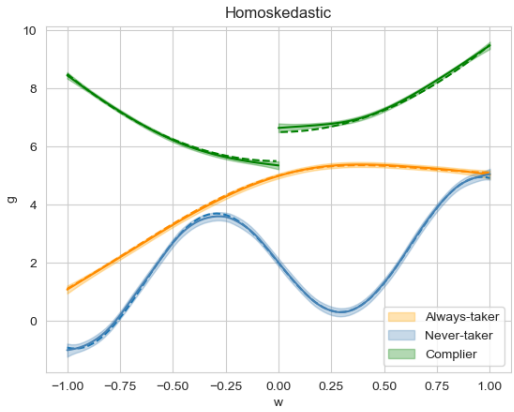


(a) Well-separated Sample (b) Poorly-separated Sample

Figure A.3: Well-Separated Data

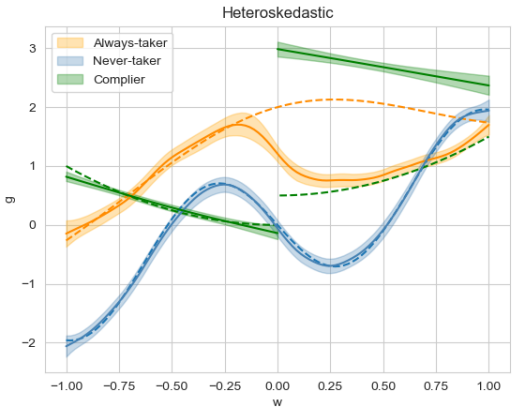


(a) Heteroskedastic

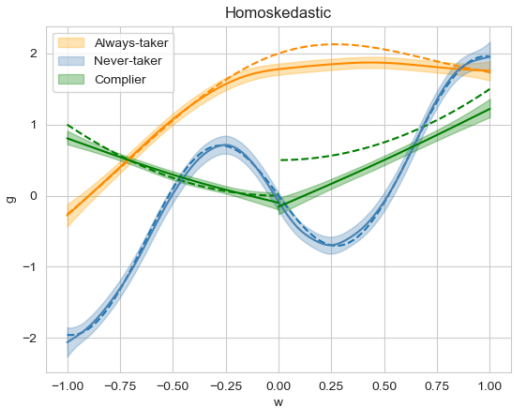


(b) Homoskedastic

Figure A.4: Poorly-Separated Data



(a) Heteroskedastic



(b) Homoskedastic

Appendix B

Supplementary Material for Chapter 2

B.1 $\mathcal{M}_2$ Model Specification

In this section, we outline the model for $\mathcal{M}2$. Assume we have a sample comprising of n independent observations, where y_{1i} and y_{2i} represent the outcome variables, where y_{1i} denotes the discrete potential outcome variable, while y_{2i} represents the potential continuous outcome variable. Let D_i denote the treatment status for individual i, with $D_i = \mathbb{1}\{d_i^* \geq 0\}$, where d_i^* represents a latent variable. Furthermore, let x_{1i} denote the covariates for y_{1i}, x_{2i} denote the covariates for y_{2i}, and x_{di} represent the covariates that determines d_i^* for individual i. We assume that

$$y_{1i} \sim \text{Poisson}\left(\mu_i\right).$$

The model can be represented as

$$g_i = X_i\beta + \varepsilon_i,$$

where

$$g_i = \left(d_i^*, \ln\left(\mu_i\right), y_{2i}\right)', \quad X_i = \begin{pmatrix} x_{di}' & 0 & 0 \\ 0 & x_{1i}' & 0 \\ 0 & 0 & x_{2i}' \end{pmatrix},$$

$\beta = (\beta_d', \beta_1', \beta_2')'$ and $\varepsilon_i = (\varepsilon_{di}, \varepsilon_{1i}, \varepsilon_{2i})'$. Assume that $\varepsilon_i \sim N(0, \Omega)$, where

$$\Omega = \begin{pmatrix} 1 & \omega_{d1} & \omega_{d2} \\ \omega_{1d} & \omega_1 & \omega_{12} \\ \omega_{2d} & \omega_{21} & \omega_2 \end{pmatrix} = \begin{pmatrix} 1 & \Omega_{12} \\ \Omega_{21} & \underset{(2\times 2)}{\Omega_{11}} \end{pmatrix}.$$

The complete data density function is expressed as

$$f\left(y_1, y_2, D, \ln(\mu) | \beta, \Omega\right) = \prod_{i=1}^{n} f\left(g_i | \beta, \Omega\right) f\left(y_{1i} | \ln\left(\mu_{1i}\right)\right) \mathbb{1}\{d_i^* \in \mathcal{B}_i\},$$

where $\mathcal{B}_i = (-\infty, 0)$ if $T_i = 0$ and $\mathcal{B}_i = [0, +\infty)$ if $T_i = 1$.

Define

$$\Omega_{22} = \Omega_{11} - \Omega_{21}\Omega_{12}.$$

Because of the unit restriction, direct sampling of Ω isn't feasible. As detailed in Section 2.2, we instead sample Ω_{22} and Ω_{12} and subsequently reconstruct the covariance matrix Ω accordingly. The estimation algorithm for this model closely resembles the one presented in Section 2.2.1 and the model proposed by Munkin and Trivedi (2003), hence its omission in this section.