The Impact of Heteroskedasticity in Observational Studies of Causal Effects

Jieyu Gao* Ivan Jeliazkov
Department of Economics, University of California, Irvine

October 2023

Abstract

There is a large and rapidly growing causal inference literature, yet little is known about the impact of heteroskedascity in popular causal settings. In observational studies where the presence of heteroskedasticity can not be ruled out with certainty, its effects in both treatment assignment and response generation must be studied. Our approach is Bayesian and involves specific modeling whose practical adequacy is then addressed through model comparisons. We build upon and extend existing methods for several well-known settngs such as sharp and fuzzy regression discontinuity designs, the potential outcomes framework, and propensity score matching. Key features of our approach in these settngs include flexible modeling and context-specific computationally efficient estimation algorithms, the ability to recover various functions of the treatment parameters, and an improved efficiency of estimation relative to alternatives that employ only a subset of the data in the analysis. Simulation studies are used to gauge the adequacy of the proposed methods, while their practical applicability is studied in three applications – 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 wellbeing. These applications illustrate the consequences of misspecification and provide strong evidence that the presence of heteroskedascity should not be ignored.

`Keywords:` Bayesian etimation; Markov chain Monte Carlo; Regression discontinuity; Potential outcomes; Propensity score; Academic performance; Healthcare expenditure; Mental health.

1 Introduction

The formulation of an identification framework through which the observed outcomes for the treated and untreated units can be compared plays a central role in observational studies of causal effects because of complications due to non-random treatment assignment, correlation due to unobserved confounders, or the fundamental missingess of counterfactual outcomes at the unit level. A variety of parametric, semi-parametric and nonparametric approaches have been proposed in the literature

*Email addresses: jieyug1@uci.edu and ivan@uci.edu

that have dealt with the effect of continuous, binary, and categorical treatments in non-experimental settings in both classical and Bayesian contexts. Instrumental variable, regression discontinuity, potential outcome, 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. Bayesian methods, which we adopt in this paper, have instead focused on the DGP directly and explicitly, leading to finite-sample inferences; possible misspecification is handled by employing flexible modeling and specification searches through formal model comparisons.

One strand of the Bayesian literature has focused on estimating treatment effects in continuous and discrete (binary, ordinal, and count) instrumental variable models (Koop and Tobias, 2004; Mintz et al., 2013; Li and Tobias, 2014; Vossmeyer, 2014a), settings with sequential outcomes Munkin (2011), as well as models with nonparametric endogeneity (Kline and Tobias, 2008; Chib et al., 2009). Instrumental variable models typically require the exclusion of the instruments from the main regression equation of interest, but this is often the source of much contention in applied research. In this context, Chan and Tobias (2015) propose methods for analyzing models with imperfect instruments which do not necessarily satisfy the exclusion restriction and could appear in both the endogenous variable and main response equation. Extensions to models of endogeneity that also involve sample selection have been presented in Chib et al. (2009), Vossmeyer (2014b), and Vossmeyer (2016) (see also van Hasselt, 2014).

Another branch in the Bayesian treatment effect research has extended the potential outcomes framework for causal analysis (Roy, 1951; Rubin, 1974, 1977, 1978, 2004, 2005) in cross-sectional settings with continuous or discrete outcome variables in the presence of binary or categorical treatments (see, e.g., Munkin and Trivedi, 1999; Chib and Hamilton, 2000; Munkin, 2003; Munkin and Trivedi, 2003; Deb et al., 2006; Li and Tobias, 2008, 2011). Extensions to longitudinal settings have been addressed in (Chib and Hamilton, 2002; Jacobi et al., 2016). Estimation has been approached both by explicitly simulating the counterfactuals from their joint distribution with the observed outcomes (Li et al., 2004) and by only involving the observed outcomes Chib (2007). Heckman et al. (2014) proposed a way to model the joint distribution of potential outcomes by incorporating a latent factor.

Following the seminal work of Rosenbaum and Rubin (1983), much interest has also been devoted to the specification and estimation of propensity score methods (Dehejia and Wahba, 1999; Imai and van Dyk, 2004; Brand and Halaby, 2006; Zhao, 2008; Caliendo and Kopeinig, 2008; An, 2010; Zhao et al., 2020; Chaudhuri and Howley, 2022; Duan et al., 2023). A recent review is offered in Rosenbaum and Rubin (2022). The framework is useful because it is simple and theoretically powerful; yet, in practice results from its application have often been mixed. The potential for misspecification of the propensity score can weaken the performance of matching estimators, which motivates our work in this area.

We also focus on causal analysis done 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 settings where treatment assignments are based on another variable with a known cutoff, and with the assumption that there is a discontinuous treatment assignment rule at the cutoff point. There are many different many applications and extensions in the litereture (see Hahn et al., 2001; Calonico et al., 2014a,b; Cattaneo et al., 2015; Dong, 2015; Dong and Lewbel, 2015; Fletcher and Tokmouline, 2018; Dong, 2019; Wright, 2020; Dong et al., 2023, among others), yet Bayesian analysis has been relatively recent (Chib and Jacobi, 2016; Branson et al., 2019; Geneletti et al., 2019; Chib et al., 2023).

Although there is a large and rapidly growing body of causal methodology, little is known about the impact of heteroskedascity in most popular causal models. One exception is Ferman and Pinto (2019) who show that even in linear specifications such as difference-in-differences, standard methods may perform poorly in the presence of heteroskedasticity, especially in small samples, and derive techniques to correct for it. In non-linear contexts, including those mentioned earlier, however, the adverse effects of heteroskedasticity in both treatment assignment and response generation are expected to be amplified by any non-linearity and therefore require a deeper understanding. We pursue this objective in the Bayesian paradigm through modeling whose practical adequacy is then addressed by model comparisons. The analysis of the treatment effect and all other parameters of interest is then based on the relevant posterior distribution that is obtained as a by-product of MCMC simulation.

The rest of the paper is organized as follows. In Section 2 we build upon and extend existing Bayesian methods for the sharp and fuzzy regression discontinuity designs in which we couple

nonparametric modeling of the running variable with a model for heteroskedasticity. In Section 3 we develop modeling and estimation techniques for the analysis of a heteroskedastic potential outcomes framework. In Section 4 we examine the effect of heteroskedasticity on the performance of propensity score matching. In each of the aforementioned causal contexts, we present MCMC estimation algorithms, study their performance and the impact of heteroskedasticity in simulation studies, and employ them in applications to gauge their practical relevance. We also provide ways for computing marginal likelihoods for formal Bayesian model comparisons. Our applications involve 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 wellbeing.

2 Regression Discontinuity Design

In this section we address both the sharp and fuzzy with several aims in mind. First of all, our model incorporates heteroskedasticity. In line with Chib et al. (2023), our model utilize all the available data rather than concentrating on a limited dataset centered around the cutoff point of the running variable. Our approach also offers flexibility and safeguards aganist potential misspecification through flexible modelling.

2.1 Sharp Regression Discontinuity Design

Consider a sharp RDD with a binary treatment $T \in \{0, 1\}$, and $T = \mathbb{1}\{w \geq w^*\}$, where w is the running variable and w^* is a known cutoff point. We use y_0 and y_1 to denote the potential outcomes when $T = 0$ and $T = 1$ correspondingly. The observed data is thus $y = (1 - T)y_0 + Ty_1$.

We assume that the potential outcome has a function form of

$$y_{ij} = g_j(w_i) + X_i'\beta_j + \epsilon_{ij}, \epsilon_{ij} \sim N(0, \sigma_{ij}^2) \quad \text{for } j \in \{0, 1\} \tag{1}$$

Thus, as specified in Chib et al. (2023), under some continuous assumptions, the RD average treatment effect (RD ATE) is defined as

$$\begin{aligned} \tau_{SRD} &\equiv \lim_{z \downarrow \tau^+} E(Y_1|w, X) - \lim_{z \uparrow \tau^-} E(Y_0|w, X) \\ &= \lim_{w \downarrow w^{*+}} (g_1(w) + x'\beta_1) - \lim_{w \uparrow w^{*-}} (g_1(w) + x'\beta_0) \end{aligned} \tag{2}$$

In a special case when $\beta_1 = \beta_0$, the RD ATE can be identified as

$$\tau_{SRD} = \lim_{w \downarrow w^{*+}} g_1(w) - \lim_{w \uparrow w^{*-}} g_1(w) \tag{3}$$

Suppose we have n observations $\{T_i, y_i, w_i, X_i, Z_i\}_{i=1}^n$ where T_i is the binary treatment varible, w_i is the running variable, y_i is the outcome varible and X_i and Z_i are covariates. We define the following vectors:

$$y_0 \equiv (y_1, ..., y_{n_0})', \quad w_0 \equiv (w_1, ..., w_{n_0})'$$

$$y_1 \equiv (y_{n_0+1}, ..., y_n)', \quad w_1 \equiv (w_{n_0+1}, ..., w_n)'$$

We use ν_j to denote the unique ordered values of the running variable w_j. Thus we have

$$\nu_0 = \begin{pmatrix} w_{0,\min} \\ ... \\ w^* \end{pmatrix} \equiv \begin{pmatrix} v_{01} \\ ... \\ \nu_{0m_0} \end{pmatrix}, \text{ and } \nu_1 = \begin{pmatrix} w^* \\ ... \\ w_{1,\max} \end{pmatrix} \equiv \begin{pmatrix} \nu_{11} \\ ... \\ \nu_{1m_1} \end{pmatrix}$$

where m_j is the number of elements in ν_j. We include w^* in both basis vector to predict the quantity of $g_0(w^*)$ and $g_1(w^*)$ in each MCMC iteration.

Let's use Θ to denote all the parameters, the likelihood function is defined as:

$$f(y|w, X, Z, \Theta) = \prod_{i=1}^{n} f_N(y_i|g_0, \beta_0, \gamma_0, X, Z)^{1-T_i} f_N(y_{n_0+i}|g_1, \beta_1, \gamma_1, X, Z)^{T_i} \tag{4}$$

2.1.1 Model

Following (Chib and Jeliazkov, 2006; Jeliazkov, 2013), we assume that the nonparametric function g has a second-order Markov process prior. Various analogous smoothness priors can also be used here (see e.g. Wahba, 1978; Silverman, 1985; Fan, 1992; Williams, 1998; Gelfand et al., 2005).

For $j \in \{0, 1\}$, $g_j = (g(\nu_{j1}), ..., g(\nu_{jm_j}))' \equiv (g_{j1}, ..., g_{jm_j})'$, and

$$g_{jl} = (1 + \frac{h_{jl}}{h_{jl-1}})g_{jl-1} - \frac{h_{jl}}{h_{jl-1}}g_{jl-2} + \mu_{jl} \tag{5}$$

where $h_{jl} \equiv \nu_{jl} - \nu_{jl-1}$ and $\mu_{jl} \sim N(0, \tau_j^2 h_{jl})$.

We assume that

$$\begin{pmatrix} g_{j1} \\ g_{j2} \end{pmatrix} |\tau_j^2 \sim N\left(\begin{pmatrix} g_{j10} \\ g_{j20} \end{pmatrix}, \tau_j^2 G_{j0} \right) \tag{6}$$

where G_{j0} is a 2×2 symmteric positive definite matrix.

Define

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

We can see that $H_j g_j = \mu_j$, where $\mu_j \sim N(\mu_{j0}, \tau_j^2 \Sigma_j)$, and thus $g_j|\tau_j^2 \sim N(g_{j0}, \tau_j^2 K_j^{-1})$, where $\mu_{j0} = (g_{j10}, g_{j20}, 0, ...0)'$, $K_j = H_j' \Sigma_j^{-1} H_j$ and $g_{j0} = H_j^{-1} \mu_{j0}$.

The model can be rewritten in the following way:

$$y_{ij} = Q_j g_j + X_i \beta_j + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma_{ij}^2) \tag{7}$$

where Q_j is a $n \times m$ incidence matrices with entries $Q_j(i,k) = 1$ if $w_{ji} = \nu_{jk}$, and 0 otherwise.

We specify the other prior distributions as $\tau_j^2 \sim IG(\frac{t_{\nu 0}}{2}, \frac{t_{d0}}{2})$, β_j is specified as $\beta_j \sim N(b_{j0}, B_{j0})$, and $\gamma_j \sim N(\gamma_{j0}, \Gamma_{j0})$.

For observations on either side of the cutoff point w^*, the following model holds:

$$\begin{aligned} & y_j = Q_j g_j + X_j \beta_j + \varepsilon_j, \quad \text{where } \varepsilon_j \sim N(0, \Omega_j), \text{and } \Omega_j = diag(\{\sigma_{ij}^2\}_{i=1}^{n_j}), \\ & g_j|\tau_j^2, \sim N(g_{j0}, \tau_j^2 K_j^{-1}), \quad \beta_j \sim N(b_{j0}, B_{j0}) \\ & \ln(\sigma_{ij}^2) = Z_i \gamma_j, \quad \gamma \sim N(\gamma_0, \Gamma_0), \quad \tau^2 \sim IG(\frac{t_{\nu 0}}{2}, \frac{t_{d0}}{2}) \end{aligned} \tag{8}$$

The joint posterior distribution can be sampled using the MCMC methods. We use computationally efficient precison-based sampling method introduced by Chan and Jeliazkov (2009a). Based on Chan et al. (2006) and Gu et al. (2009), we use the iteratively reweighted least squares algorithm introduced by Gamerman (1997) and Nott and Leonte (2004) to sample the parameter γ. This is a more computationally efficient sampling method compared to the conventional Metropolis-Hasting algorithm, because the optimization process is no longer needed. The algorithm is outlined in Algorithm 1. The homoskedastic model can be estimated using the same algorithm, with the distinction that the matrix Z reduces to a constant vector.

If x is independent of the running variable, the RD average treatment effect (RD ATE) can be estimated as:

$$\hat{\tau}_{SRD} = \frac{1}{M^2} \Big[\sum_{g=1}^{M} \sum_{l=1}^{M} \Big(\lim_{z \downarrow \tau^+} \Big(\frac{1}{n_1} \sum_{i=1}^{n_1} (g_1^{(g)} + x_{1i}' \beta_1^{(g)}) \Big) - \lim_{z \uparrow \tau^-} \Big(\frac{1}{n_0} \sum_{j=1}^{n_0} (g_0^{(l)} + x_{0j}' \beta_0^{(l)}) \Big) \Big) \Big] \tag{9}$$

M is the total number of MCMC iterations, and $\beta_j^{(g)}$ and $g^{(g)}$ denote the sample values of β_j and g_j in g-th iteration of the MCMC sampling process. If X is correlated with the running variable, our estimation in Equation (11) can be calculated solely on the observed X values at the cutoff points.

2.1.2 Binary Outcomes

In this section, we disucss the estimation algorithm for cases in which the outcome variable y takes on binary values. We work under the assumption that there exists a latent variable y^*, and we

Algorithm 1 (Semi-parametric Sharp RDD Model)

---

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

(2) Sample $[\beta_j|y_j, g_j, \gamma_j] \sim N(\hat{\beta}_j, \hat{B}_j)$, where $\hat{B}_j = (B_{j0}^{-1} + X_j'\Omega_j^{-1}X_j)^{-1}$, and $\hat{\beta}_j = \hat{B}_j(B_{j0}^{-1}b_{j0} + X_j'\Omega_j^{-1}(y_j - Q_j g_j))$.

(3) Sample $[\tau_j^2|g_j] \sim IG(\frac{t_{\nu_{j0}}+m_j}{2}, \frac{t_{d_{j0}}+(g_j-g_{j0})'K(g_j-g_{j0})}{2})$.

(4) Sample $[\gamma_j|y_j, g_j, \beta_j]$ using the Metropolis-Hasting algorithm.

(a) Define

$$e_{ij}^c = (y_{ij} - g_j^c(w_i) - X_i'\beta_j^c)^2$$
$$\eta_{ij}^c = Z_i'\gamma_j^c + \frac{e_{ij}^c - \sigma_{ij}^{2c}}{\sigma_{ij}^{2c}}, \quad \eta_j^c = (\eta_{1j}^c, ..., \eta_{n_j}^c)^T.$$

where σ_{ij}^{2c}, g_j^c and β_j^c are the current values of σ_{ij}^2, g_j and β_j.

(b) Using the iteratively reweighted least squares algorithm, we obtain a proposal value γ_j^p from the proposal density, which is $T_4(\gamma_j^p|\hat{\gamma}_j^p, V_j^p)$, where $V_j^p = (\Gamma_{0j}^{-1} + \frac{1}{2}Z_j'Z_j)^{-1}$ and $\hat{\gamma}_j^p = V_j^p(\Gamma_{j0}^{-1}\gamma_{j0} + \frac{1}{2}Z_j'\eta_j^c)$. We obtain $q(\gamma_j^c|\hat{\gamma}_j^c, V_j^c)$ in the reverse direction moving from the proposed value to the current value.

(c) The acceptance rate α is defined as follows:

$$\alpha = \frac{f(y_j|g_j^c, \beta_j^c, \gamma_j^p)\pi(\gamma_j^p|\gamma_{j0}, \Gamma_{j0})}{f(y_j|g_j^c, \beta_j^c, \gamma_j^c)\pi(\gamma_j^c|\gamma_{j0}, \Gamma_{j0})} \times \frac{q(\gamma_j^c|\hat{\gamma}_j^c, V_j^c)}{q(\gamma_j^p|\hat{\gamma}_j^p, V_j^p)}$$

If the proposed value is not accepted, γ_j^c is repeated in the next iteration.

---

define y_i as the indicator function: $y_i = \mathbb{1}\{y^* \geq 0\}$. Moreover, we presume that the definition of y^* aligns with the equation as stated in Equation (1).

In the homosekdastic model, we impose a contraint that fixes the variance of the latent variable at 1 for the sake of identification. For the same reason, we assume that the matrix Z, which plays a role in determining the variance, does not include a constant term.

Let's use Θ to denote all the parameters, the likelihood function is defined as:

$$\begin{aligned}
&\log f(y|w, X, Z, \Theta) \\
&= \sum_{i=1}^{n}(1-T_i)\log\left(\Phi(\frac{g_j(\omega_i)+x_i'\beta_0}{\sigma_{i0}})\mathbb{1}\{y_i^*\geq 0\}+(1-\Phi(\frac{g_0(\omega_i)+x_i'\beta_0}{\sigma_{i0}}))\mathbb{1}\{y_i^*<0\}\right) \\
&+ T_i\log\left(\Phi(\frac{g_j(\omega_i)+x_i'\beta_0}{\sigma_{i1}})\mathbb{1}\{y_i^*\geq 0\}+(1-\Phi(\frac{g_j(\omega_i)+x_i'\beta_0}{\sigma_{i1}}))\mathbb{1}\{y_i^*<0\}\right)
\end{aligned} \tag{10}$$

The estimation algorithm for Sharp RDD with binary outcome is presented in Algorithm 2. The primary distinction between Algorithm 1 and 2 is that in Algorithm 2 we need to sample the latent variable y^* in each iteration.

Similarly to the discussion in Section 2.1, the RD average treatment effect (RD ATE) with binary outcome variable can be estimated as follows:

$$\hat{\tau}_{SRD} = \frac{1}{M^2}\Big[\sum_{g=1}^{M}\sum_{l=1}^{M}\left(\lim_{z\downarrow\tau^+}\left(\frac{1}{n_1}\sum_{i=1}^{n_1}\Phi(\frac{g_1^{(g)}+x_{1i}'\beta_1^{(g)}}{\sqrt{\exp(z_i'\gamma_1^{(g)})}})\right)-\lim_{z\uparrow\tau^-}\left(\frac{1}{n_0}\sum_{j=1}^{n_0}\Phi(\frac{g_0^{(l)}+x_{0i}'\beta_0^{(l)}}{\sqrt{\exp(z_i'\gamma_0^{(l)})}})\right)\right)\Big] \tag{11}$$

2.1.3 Bayesian Model Comparison and Marginal Likelihood Estimation

In this section, we delve into the techniques used to estimate the marginal likelihood, which enables us to compare various models within the sharp RDD framework. Marginal likelihood is a key component in calculating the bayes factor (Kass and Raftery (1995)). Let's use y to denote the observed data. For two models M_1 and M_2, the posterior ratio is defined as:

$$\frac{P(M_1|y)}{P(M_2|y)} = \frac{m(y|M_1)}{m(y|M_2)}\frac{P(M_1)}{P(M_2)} \tag{12}$$

where $P(M_1)$ and $P(M_2)$ are the prior odds. When we assume that M_1 and M_2 have equal probabilties a priori, the posterior ratio is equal to the bayes factor, which is the ratio of the marginal likelihood of the two models.

The marginal likelihood can be written as:

$$m(y|M) = \frac{f(y|M,\theta)\pi(\theta|M)}{\pi(\theta|y,M)} \tag{13}$$

According to Chib (1995), the marginal likelihood can be estimated by MCMC methods at a point θ^* in the parameter space for model M. Hence, the estimation equation of the log marginal likelihood is represented by Equation (14).

$$\log\hat{m}(y|M) = \log f(y|\theta^*, M) + \ln\pi(\theta^*|M) - \log\hat{\pi}(\theta^*|y, M) \tag{14}$$

Algorithm 2 (Semi-parametric Sharp RDD Model with Binary Outcome Variable)

---

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

(2) Sample $[\beta_j|y_j^*, g_j, \gamma_j] \sim N(\hat{\beta}_j, \hat{B}_j)$, where $\hat{B}_j = (B_{j0}^{-1} + X_j'\Omega_j^{-1}X_j)^{-1}$, and $\hat{\beta}_j = \hat{B}_j(B_{j0}^{-1}b_{j0} + X_j'\Omega_j^{-1}(y_j^* - Q_j g_j))$.

(3) Sample $[\tau_j^2|g_j] \sim IG(\frac{t_{\nu_{j0}}+m_j}{2}, \frac{t_{d_{j0}}+(g_j-g_{j0})'K(g_j-g_{j0})}{2})$.

(4) Sample $[\gamma_j|y_j^*, g_j, \beta_j]$ using the Metropolis-Hasting algorithm.

(a) Define

$$e_{ij}^c = (y_{ij}^* - g_j^c(w_i) - X_i^{'}\beta_j^c)^2$$
$$\eta_{ij}^c = Z_i^{'}\gamma_j^c + \frac{e_{ij}^c - \sigma_{ij}^{2c}}{\sigma_{ij}^{2c}}, \quad \eta_j^c = (\eta_{1j}^c, ..., \eta_{n_j}^c)^{'}.$$

where σ_{ij}^{2c}, g_j^c and β_j^c are the current values of σ_{ij}^2, g_j and β_j.

(b) Using the iteratively reweighted least squares algorithm, we obtain a proposal value γ_j^p from the proposal density, which is $T_4(\gamma_j^p|\hat{\gamma}_j^p, V_j^p)$, where $V_j^p = (\Gamma_{0j}^{-1} + \frac{1}{2}Z_j^{'}Z_j)^{-1}$ and $\hat{\gamma}_j^p = V_j^p(\Gamma_{j0}^{-1}\gamma_{j0} + \frac{1}{2}Z_j^{'}\eta_j^c)$. We obtain $q(\gamma_j^c|\hat{\gamma}_j^c, V_j^c)$ in the reverse direction moving from the proposed value to the current value.

(c) The acceptance rate α is defined as follows:

$$\alpha = \frac{f(y_j^*|g_j^c, \beta_j^c, \gamma_j^p)\pi(\gamma_j^p|\gamma_{j0}, \Gamma_{j0})}{f(y_j^*|g_j^c, \beta_j^c, \gamma_j^c)\pi(\gamma_j^c|\gamma_{j0}, \Gamma_{j0})} \times \frac{q(\gamma_j^c|\hat{\gamma}_j^c, V_j^c)}{q(\gamma_j^p|\hat{\gamma}_j^p, V_j^p)}$$

If the proposed value is not accepted, γ_j^c is repeated in the next iteration.

(5) Sample $[y_{ij}^*|y_{ij}, g_j, \beta_j, \gamma_j] \sim TN(g_j(\omega_i) + x_i'\beta_j, \exp(z_i'\gamma_j))$, where $y_{ij}^* \in (-\infty, 0)$ if $y_i = 0$, and $y_{ij}^* \in (0, \infty)$ if $y_i = 1$.

Under the independence assumpiton, we can demontrate that:

$$\hat{m}(y|M) = \hat{m}(y_0|M)\hat{m}(y_1|M) \tag{15}$$

In the following discussion, we omit the model indicator M for the sake of simplifying the notation. Let $\theta_j = (\beta_j, g_j, \tau_j^2)$. We evaluate the posterior ordinate using the poterior mean.

To estimate the marginal likelihood in the context of a sharp RDD model, we define:

$$\hat{m}(y_j) = \frac{f(y_j|\theta_j^*, \gamma_j^*)\pi(\theta_j^*, \gamma_j^*)}{\pi(\theta_j^*|y_j)\pi(\gamma_j^*|y_j, \theta_j^*)} \tag{16}$$

We employ the CRT method as discussed in Jeliazkov and Lee (2010) to estimate $\pi(\theta^*|y_j)$ as shown in Equation (17) for Sharp RDD with continuous outcome variable, and Equation (18) for the model with binary outcome. We utilize the method proposed in Chib and Jeliazkov (2001) to estimate $\pi(\gamma_j^*|y_j, \theta_j^*)$ because the posterior distribution for γ_j is not standard.

$$\hat{\pi}(\theta_j^*|y) = G^{-1}\sum_{g=1}^{G} K(\theta_j^{(g)}, \theta_j^*|y_j) \tag{17}$$

$$\hat{\pi}(\theta_j^*|y) = G^{-1}\sum_{g=1}^{G} K(\theta_j^{(g)}, \theta_j^*|y_j, y_j^{*(g)}) \tag{18}$$

where

$$K(\theta, \theta^*|y) = \prod_{g=1}^{G} \pi(\theta_r^*|y, \{\theta_s^*\}_{(s<r)}, \{\theta_s^{(g)}\}_{(s>r)})$$

$$K(\theta, \theta^*|y, y^*) = \prod_{g=1}^{G} \pi(\theta_r^*|y, \{\theta_s^*\}_{(s<r)}, \{\theta_s^{(g)}\}_{(s>r)}, y_j^{*(g)})$$

where $\{\theta^{(g)}\}_{g=1}^{G}$ is G draws from the posterior distribution using the MCMC sampling.

Following Chib and Jeliazkov (2001), we estimate $\pi(\gamma_j^*|y_j, \theta_j^*)$ for the model with continuous outcome variable as follows:

$$\hat{\pi}(\gamma_j^*|y_j, \theta_j^*) = \frac{S^{-1}\sum_{s=1}^{S} \alpha(\gamma_j^{(s)}, \gamma_j^*|y, \theta_j^*) q(\gamma^{(s)}, \gamma^*|y, \theta_j^*)}{L^{-1}\sum_{l=1}^{L} \alpha(\gamma^*, \gamma^{(l)}|y, \theta_j^*)} \tag{19}$$

The sample $\{\gamma^{(s)}\}_{s=1}^{S}$ is obtained from the reduced MCMC run by setting the other parameters to be θ_j^*. The denominator sample $\{\gamma^{(s)}\}_{l=1}^{L}$ is sampled from the proposal distribution $q(\gamma_j^*, \gamma_j^{(l)}|y, \theta_j^*)$.

$\pi(\gamma_j^*|y_j, \theta_j^*)$ for the model with binary outcome variable can be estimated as follows:

$$\hat{\pi}(\gamma_j^*|y, \theta_j^*) = \frac{S^{-1}\sum_{i=1}^{S} \alpha(\gamma_j^{(s)}, \gamma_j^*|y, y^{*(s)}, \theta_j^*) q(\gamma_j^{(s)}, \gamma_j^*|y, y^{*(s)}, \theta_j^*)}{L^{-1}\sum_{l=1}^{L} \alpha(\gamma_j^*, \gamma_j^{(L)}|y, \theta_j^*, y^*)} \tag{20}$$

where the sample $\{(\gamma_j^{(s)}, y^{*(s)})\}_{s=1}^{S}$ is sampled from the reduced run where we set all the other parameters to be θ_j^*. The sample $\{\gamma_j^l\}_{l=1}^{L}$ is sampled from the distribution $q(\gamma_j^*, \gamma_j|y, y^*, \theta_j^*)\pi(y^*|y, \theta_j^*)$.

2.1.4 Simulation Results

In this section, we performed a simulation study to assess the influence of heteroskedasticity in the context of sharp RDD. We report mean, 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).

We simulate the data from following generating process: $y_j = g_j(w) + x_i'\beta_j + \epsilon_{ij}$ for $j = 0, 1$, where $\epsilon_{ij} \sim N(0, \sigma_{ij}^2)$, and $\ln(\sigma_{ij}^2) = z_i'\gamma_j$.

The function g is generated using the following equations:

$$\begin{aligned} g_0(w) &= 1 - \sin(w+1) + (w+1)^2 \\ g_1(w) &= -1 - \sin(w) + w^2 \\ w &\sim U[-1, 1] \end{aligned} \tag{21}$$

The other parameters are generated in the following way:

$$\gamma_j \sim N(\mathbf{0}, \mathbf{I}); \beta_j \sim N(\mathbf{0}, \mathbf{I}); X \sim N(\mathbf{0}, \mathbf{I}); Z = \begin{pmatrix} \mathbf{1} & w & X1 \end{pmatrix}$$

where X1 is the first column of the generated covariates X. Since X is sampled independenly from w, we can estimate the treatment effect by averaging over all X values in the sample.

The marginal likelihood and the estimated treatment effect is presented in Table 1. The heteroskedastic model is recommended in all scenarios. As the sample size increases, both homoskedastic and heteroskedastic models yield point estimates that approach the true treatment effect. With larger sample sizes, we gain stronger evidence in favor of the heteroskedastic model based on the marginal likelihood estimates. Both Bayesian homoskedastic and heteroskedastic models provide more precise estimates compared to RDRobust. This is primarily because RDRobust relies on data points around the cutoff, while our model utilizes all available data points.

We did a robustness check by employing non-infomrative priors. We modified the prior distribution of β to be $N(0, 1000)$ and γ to be $N(0, 100)$. The results are presented in Appendix A Table 15, and they closely resemble the results in Table 1. Additional simulation results are available in Appendix A.

We then explore a more extreme scenario with fewer data points around the cutoff and a more pronounced presence of heteroskedasticity. We collect a sample of data with a sample size of 5000

Table 1: RD ATE with Continuous Outcome Variable ($\beta_1 \neq \beta_0$)

	Model	True ATE	RD ATE	SD	95% Credible Interval	Marginal Likelihood
n = 500	Homoskedastic	-2.2357	-2.3638	0.2514	(-2.8657, -1.8774)	-765.22
	Heteroskedastic	-2.2357	-2.2616	0.1774	(-2.6127, -1.9173)	-617.77
	RDRobust	-2.2357	-2.6305	0.6780	(-4.1775, -1.0734)	
n = 5000	Homoskedastic	-2.2195	-2.0420	0.1289	(-2.2936, -1.7872)	-8711.93
	Heteroskedastic	-2.2195	-2.0996	0.1062	(-2.3055, -1.8885)	-7929.93
	RDRobust	-2.2195	-2.1892	0.2142	(-2.7119, -1.7242)	
n = 50000	Homoskedastic	-2.2092	-2.1987	0.0409	(-2.2793, -2.1191)	-68715.40
	Heteroskedastic	-2.2092	-2.1993	0.0283	(-2.2550,-2.1439)	-59044.64
	RDRobust	-2.2092	-2.0633	0.0690	(-2.2046, -1.8831)	

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

from the following DGP.

The function g is sampled from the following equations:

$$g_0(w) = \sin(w) + \exp(-20(w + 0.5)^2)$$
$$g_1(w) = 1.2 - \sin(w) - \exp(-20(w - 0.5)^2)$$

The other parameters are generated in the following way:

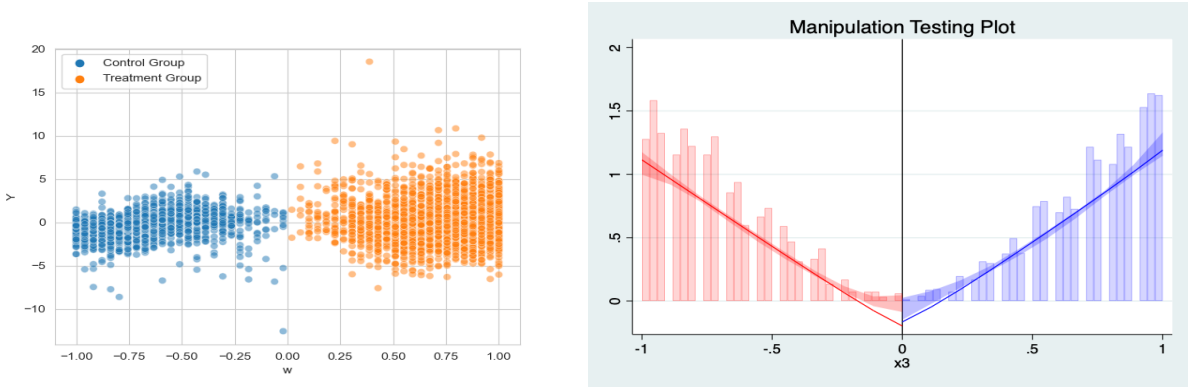
$$\gamma_0 = \begin{pmatrix} -2 & 2 & 1 \end{pmatrix}'; \gamma_1 = \begin{pmatrix} -2 & 2 & -1 \end{pmatrix}'; \beta_j \sim N(\mathbf{0}, \mathbf{I}); X \sim N(\mathbf{0}, \mathbf{I}); Z = \begin{pmatrix} \mathbf{1} & ||w| - 1.5| & X1 \end{pmatrix}$$

Instead of sampling the running variable from a uniform distribution, we sampled the data so that there are fewer points around the cutoff points. The data and density is presened in Figure 1. The generated data passed the frequentist density test (McCrary, 2008) with p-value 0.4655.

The estimated RD ATE is provided in Table 2. The estimated parameter for the nonparamteric function g can be found in Figure 2. Upon examining the figure, it becomes apparent that the homoskedastic model estimates were significantly affected by the outliers near the cutoff points, which ultimately resulted in biased estimates for the RD ATE. The heteroskedastic model, on the other hand, can produce a good estimate for the parameters. This is due to the fact that the data points are weighted differently in the estimation process for the homoskedastic model and heteroskedastic model. Thus, ignoring the heteroskedasticity can be problematic in this case. In this example, RDRobust yielded a notably wide 95% confidence interval, primarily due to the limited number of data points near the cutoff. While the true treatment effect does fall within

the reported confidence interval, the heteroskedastic model offers a more efficient estimation of the treatment effect.

Figure 1: Data and Running Variable Density



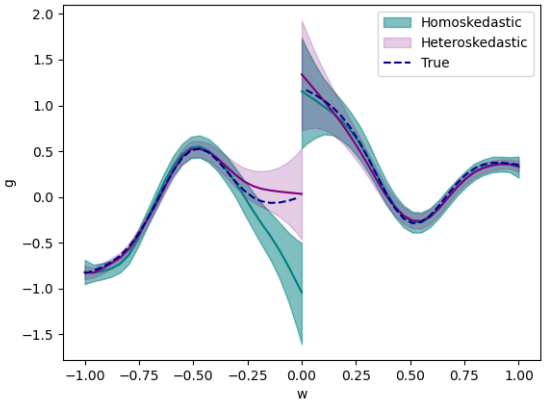
(a) Data

(b) Running Variable Density

Table 2: RD ATE Results

Model	True ATE	RD ATE	SD	95% CI	Marginal Likelihood
Homoskedastic	1.2005	2.2085	0.4174	(1.3831, 3.0230)	-7642.76
Heteroskedastic	1.2005	1.3185	0.3986	(0.5307, 2.0935)	-6066.56
RDRobust	1.2005	2.0958	0.8682	(-0.1510, 4.1174)	

Figure 2: Estimated Parameters $\hat{g}$



We also simulated the data for models with binary outcomes as follows. The function g is

generated using the following equations:

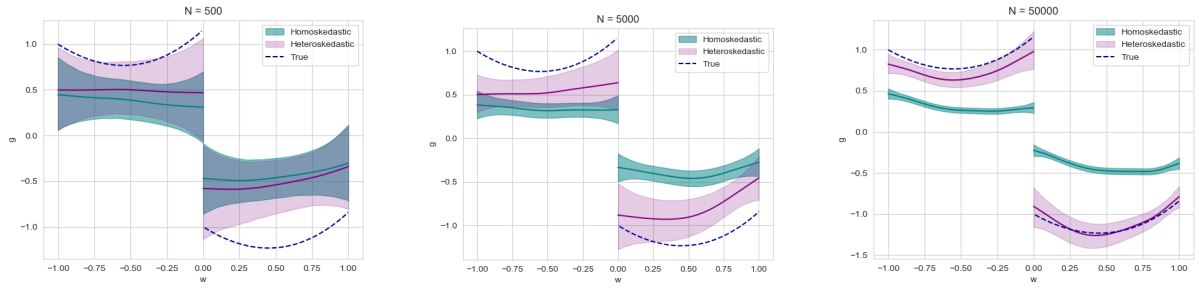
$$g_0(w) = 1 - \sin(w+1) + (w+1)^2$$
$$g_1(w) = -1 - \sin(w) + w^2$$
$$w \sim U[-1, 1]$$

The other parameters are generated in the following way:

$$\gamma_0 = \gamma_1 = 2; \beta_j \sim N(\mathbf{0}, \mathbf{I}); X \sim N(\mathbf{0}, \mathbf{I}); Z = \big(||w| - 1.5|\big)$$

The estimated nonparametric parameter $\hat{g}_0$ and $\hat{g}_1$ are depicted in Figure 3. Table 3 presents the RD ATE and marginal likelihood statistics. Heteroskedastic model provides consistent estimates when the sample size increases, whereas both the homoskedastic model and RDRobust exhibit inconsistencies in this context. With a larger sample size, strong evidence emerges in support of the heteroskedastic model.

Figure 3: Estimated g with Binary Outcome Variable



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

Table 3: RD ATE with Binary Outcome Variable

	Model	True ATE	RD ATE	SD	95% Credible Interval	Marginal Likelihood
n = 500	Homoskedastic	-0.2303	-0.2114	0.0845	(-0.3741,-0.0432)	-308.03
	Heteroskedastic	-0.2303	-0.2284	0.0842	(-0.3906,-0.0607)	-305.56
	RDRobust	-0.2303	-0.2262	0.1318	(-0.5405,0.0801)	
n = 5000	Homoskedastic	-0.2579	-0.2202	0.0364	(-0.2916, -0.1488)	-2812.35
	Heteroskedastic	-0.2579	-0.2700	0.0372	(-0.3422,-0.1963)	-2775.49
	RDRobust	-0.2579	-0.2218	0.0480	(-0.3411,-0.1167)	
n = 50000	Homoskedastic	-0.2566	-0.16885	0.0159	(-0.2002,-0.1377)	-27844.71
	Heteroskedastic	-0.2566	-0.24269	0.0177	(-0.2775,-0.2079)	-27470.32
	RDRobust	-0.2566	-0.1756	0.0174	(-0.2194,-0.1386)	

2.2 Empirical Application

Many universities use academic probation as a tool to motivate students, and help them improve their performance. Researchers commonly employ regression discontinuity design to evaluate the impact of academic probation on academic performance (see e.g. Fletcher and Tokmouline, 2018; Wright, 2020). We use the data from Texas Higher Education Opportunity Project (THEOP) to study the impact of academic probation on students' short-run academic performance.

We consider the treated group as the students who receive academic probation at the end of their first semester. The outcome variables are the students' GPA of two subsequent semesters as well as their graduation rate. We use the longitudinal administrative data from University of Texas, Austin (UT Austin) with students admitted from 1991 through 2000. This data set include a rich set of academic and demographic variables, including students' gender, ethnicity, high school class rank, private highs school indicator, SAT or ACT test scores, and citizenship. At UT Austin, students receive the treatment if their cumulative GPA is under 2.0. The students must raise their GPA above the threshold, otherwise they can be dismissed from the university. This policy may vary across different schools.

One key underlying assumption for the sharp RD design is that the students near the threshold can not manipulate their GPA. Additionally, since the GPA data is rounded to the nearest tenth, it is possible that some students with a GPA of 1.95-1.99, who received the treatment, but are categorized as the control group show a 2.0 GPA. To address this issue, we eliminate the students who have a first semester cumulative GPA of exact 2.0. We only use the covariates X and Z where students who have a first semester GPA of excat 2.0 for treatment effect estimation. Following McCrary (2008), we run a density test of the running variable. The t-test statistic is -1.0027 with a p-value of 0.3160. Thus, there is no evidence showing that student manupulate their GPA to avoid the treatment.

We include the covariates, including the student's gender, citizenship, ethnic, standardized SAT score, high school decile, private high school indicator, and if they have a major in first semester. The summary of statistics for the data are presented in Table 4, 5 and 6.

The estimated function $\hat{g}_0$ and $\hat{g}_1$ are represented in Figure 4, and the RD ATE estimates and the marginal likelihood are provided in Table 7. Notably, the RDRobust model yields larger standard errors compared to the Bayesian models. Academic probation are practically relevent

Table 4: Summary of Statistics: Second Semester GPA

Variable	Control (37980) Mean	Control (37980) Std Dev	Treatment (5789) Mean	Treatment (5789) Std Dev
Female	0.516	0.500	0.386	0.487
Non US Citizen	0.003	0.052	0.002	0.039
Minority	0.032	0.175	0.068	0.252
SAT (Standardized)	0.081	0.997	-0.398	0.911
Second Decile (High School)	0.264	0.441	0.309	0.462
Third Decile (High School)	0.119	0.324	0.209	0.406
Fourth Decile or Below (High School)	0.082	0.275	0.224	0.417
Private High School	0.051	0.220	0.047	0.212
Has Major	0.754	0.431	0.696	0.460
Second semester term GPA	3.003	0.762	2.026	0.892
First semester term GPA	3.205	0.546	1.335	0.495

Table 5: Summary of Statistics: Third Semester GPA

Variable	Control (36738) Mean	Control (36738) Std Dev	Treatment (4366) Mean	Treatment (4366) Std Dev
Female	0.515	0.500	0.390	0.488
Non US Citizen	0.003	0.053	0.002	0.043
Minority	0.032	0.176	0.066	0.248
SAT (Standardized)	0.068	0.997	-0.401	0.917
Second Decile (High School)	0.263	0.440	0.311	0.463
Third Decile (High School)	0.119	0.323	0.207	0.405
Fourth Decile or Below (High School)	0.082	0.274	0.227	0.419
Private High School	0.051	0.221	0.049	0.217
Has Major	0.755	0.430	0.694	0.461
Third semester term GPA	2.949	0.805	2.168	0.892
First semester term GPA	3.213	0.544	1.414	0.445

Table 6: Summary of Statistics: Graduation

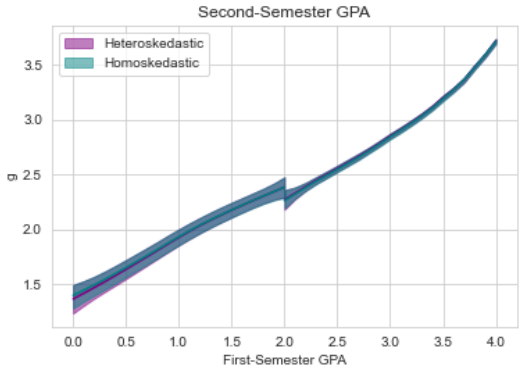
Variable	Control (38525) Mean	Control (38525) Std Dev	Treatment (6494) Mean	Treatment (6494) Std Dev
Female	0.517	0.500	0.387	0.487
Non US Citizen	0.003	0.052	0.001	0.037
Minority	0.032	0.175	0.066	0.248
SAT (Standardized)	0.084	0.997	-0.378	0.917
Second Decile (High School)	0.264	0.441	0.305	0.461
Third Decile (High School)	0.119	0.324	0.213	0.409
Fourth Decile or Below (High School)	0.082	0.275	0.225	0.418
Private High School	0.051	0.220	0.048	0.214
Has Major	0.753	0.431	0.693	0.461
4-Year Graduation	0.527	0.499	0.126	0.332
Graduation	0.802	0.399	0.329	0.470

to the subsequent semester GPAs in all the Bayesian models, while the impact is not statistically significant at 5% level based on the RDRobust result. The impacts of academic probation on graduation rates are not obvious according to the results of all models. Moreover, the marginal likelihood results indicate that the heteroskedastic model is more apporiate than its homoskedastic couterpart in analyzing the dataset.

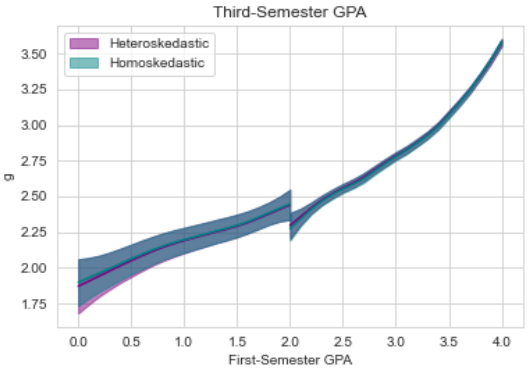
Table 7: RD ATE Results

Outcomes	Models	RD ATE	Std.	95% CI	Marginal Likelihood
Second-Semester GPA	Homoskedastic	0.1709	0.0489	(0.0745, 0.2664)	-42714.0778
	Heteroskedastic	0.1731	0.0556	(0.0641, 0.2825)	-41248.4293
	RDRobust	-0.1434	0.1022	(-0.4538, 0.2644)	
Third-Semester GPA	Homoskedastic	0.1567	0.0570	(0.0461, 0.2700)	-44410.9014
	Heteroskedastic	0.1391	0.0597	(0.0239, 0.2581)	-43617.5593
	RDRobust	-0.1357	0.1070	(-0.4276, 0.3372)	
4-Year Graduation	Homoskedastic	-0.0198	0.0224	(-0.0633, 0.0246)	-27662.01551
	Heteroskedastic	-0.0270	0.0218	(-0.0691, 0.0165)	27650.87316
	RDRobust	-0.0343	0.0535	(-0.3026, 0.0796)	
Graduation	Homoskedastic	0.0213	0.0238	(-0.0254, 0.0679)	-11776.01168
	Heteroskedastic	0.0076	0.0233	(-0.0380, 0.0532)	-21762.14031
	RDRobust	-0.0074	0.0627	(-0.3038, 0.1396)	

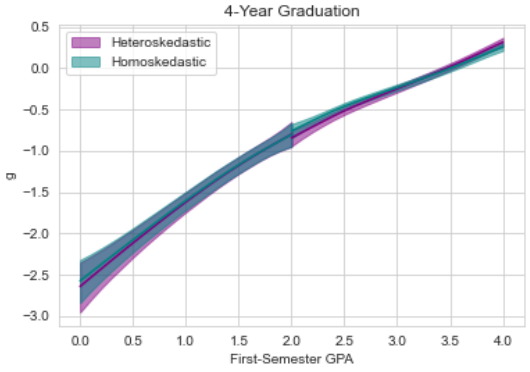
Figure 4: Estimated Nonparmetric Parameters



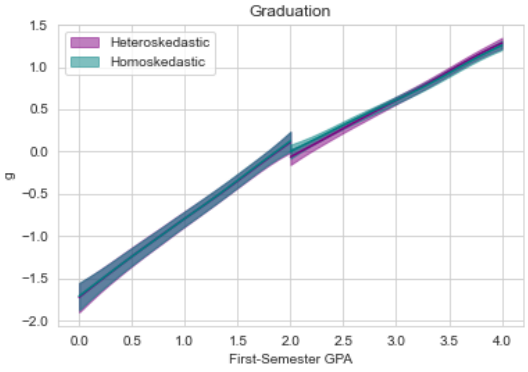
(a) Second-Semester GPA



(b) Third-Semester GPA



(c) 4-Year Graduation Rate



(d) Graduation Rate

2.3 Fuzzy Regression Discontinuity Design

In this section, we present a Bayesian fuzzy RDD model and outline the corresponding estimation algorithm. To help illustrate this model, we employ simulations. In the context of the fuzzy RD design, we work under the assumption that the population consists of three distinct groups: compliers, never-takers, and always-takers. Following Chib et al. (2023), we posit the existence of an unobserved discrete confounding variable s which categorizes individuals into one of these three types: compliers (denoted as c), never-takers (denoted as n), and always-takers (denoted as a).

Consider a fuzzy RD design with a binary treatment $T \in \{0, 1\}$, and the treatment status for compliers is $\mathbb{1}\{w \geq w^* | s = c\}$, where w is the running variable and w^* is a known cutoff point. We use y_0 and y_1 to denote the potential outcomes when $T = 0$ and $T = 1$ correspondingly. The observed data is thus $y = (1 - T)y_0 + Ty_1$. We assume that the potential outcome has a function form of $y_{ij} = g_j(w_i) + X_i\beta_j + \epsilon_{ij}$ for compliers, where g_j are continuous at the cutoff point. The model is specified as follows:

$$
\begin{aligned}
&s = c : T = \mathbb{1}\{w \geq w^*\}, y_{ij} = g_j(w) + X\beta_j + \epsilon_{ij}, j \in \{0, 1\}, \text{ where } \epsilon_{ij} \sim N(0, \sigma_{ij}^2), \text{ and } \ln(\sigma_{ij}^2) = Z'_{ij}\gamma_j \\
&s = a : T = 1, y_i = g_a(w) + X\beta_a + \epsilon_{ia}, \text{ where } \epsilon_{ia} \sim N(0, \sigma_{ia}^2), \text{ and } \ln(\sigma_{ia}^2) = Z'_{ia}\gamma_a \\
&s = n : T = 0, y_i = g_n(w) + X\beta_n + \epsilon_{in}, \text{ where } \epsilon_{in} \sim N(0, \sigma_{in}^2), \text{ and } \ln(\sigma_{in}^2) = Z'_{in}\gamma_n \\
&P(s = k) = q_k > 0, \forall k \in \{c, a, n\} \text{ and } q_c + q_n + q_a = 1
\end{aligned}
\tag{22}
$$

Thus, as specified in Chib et al. (2023), under some smoothness conditions, the RD average treatment effect (ATE) is defined as follows:

$$
\begin{aligned}
\tau_{FRD} &\equiv \lim_{z \downarrow \tau^+} E(Y_1 | w, X, s = c) - \lim_{z \uparrow \tau^-} E(Y_0 | w, X, s = c) \\
&= \lim_{w \downarrow w^{*+}} (g_1(w) + X\beta_{c1}) - \lim_{w \uparrow w^{*-}} (g_0(w) + X\beta_{c0})
\end{aligned}
\tag{23}
$$

The sample data in the fuzzy RD case is summarized by Table 8.

Table 8: Fuzzy RD Data

	$w < w*$	$w \geq w*$
T = 0	c, n	n
T = 1	a	c, a

The likelihood function is defined as:

$$L = \prod_{i \in I_{00}} (q_c \phi(y_i|g_0, \beta_0, \tau_0^2, \gamma_0) + q_n \phi(y_i|g_n, \beta_n, \tau_n^2, \gamma_n)) \prod_{i \in I_{10}} q_n \phi(y_i|g_n, \beta_n, \tau_n^2, \gamma_n)$$
$$\prod_{i \in I_{01}} q_a \phi(y_i|g_a, \beta_a, \tau_a^2, \gamma_a) \prod_{i \in I_{11}} (q_c \phi(y_i|g_1, \beta_1, \tau_1^2, \gamma_1) + q_a \phi(y_i|g_a, \beta_a, \tau_a^2, \gamma_a)) \tag{24}$$

where I_{iT} is the group of observations where $i = \mathbb{1}\{w \geq w^*\}$, T is the treatment variable, and ϕ is the probabilty density function (PDF) of the normal distribution.

We assume that the prior distirbution of $q = (q_a, q_n, q_c)$ follows a Dirichlet distirbution with parameters (n_{a0}, n_{n0}, n_{c0}). All the other parameters follows the same prior distirbution as discussed in section 2.1.1. g_a and g_n are vectors with dimension $m = m_0 + m_1$,where m_0 is the dimension of vector ν_0, and m_1 is the dimension of vector ν_1. ν_0 and ν_1 are the unique ordered values of the running variable w_0 and w_1, where w_0 is the vector of all w where $w < w^*$, and w_1 is the vector of all w where $w \geq w^*$. g_0 and g_1 are vectors with dimension m_0 and m_1 correspondingly. In the estimation process, the nonparatmetric function g in each group are updated using both the prior information and observations that were categorized into this specific group in each iteration.

The model is summarized as follows:

$$\begin{aligned}
&y_j = Q_j g_j + X_j \beta_j + \varepsilon_j, \quad \text{where } \varepsilon_j \sim N(0, \Omega_j), \text{and } \Omega_j = diag(\{\sigma_{ij}^2\}_{i=1}^{n_j}) \\
&y_n = Q_n g_n + X_n \beta_n + \varepsilon_n, \quad \text{where } \varepsilon_n \sim N(0, \Omega_n), \text{and } \Omega_n = diag(\{\sigma_{in}^2\}_{i=1}^{n_n}) \\
&y_a = Q_a g_a + X_a \beta_a + \varepsilon_a, \quad \text{where } \varepsilon_a \sim N(0, \Omega_a), \text{and } \Omega_a = diag(\{\sigma_{ia}^2\}_{i=1}^{n_a}) \\
&g_j|\tau_j^2, \sim N(g_{j0}, \tau_j^2 K_j^{-1}), g_n|\tau_n^2, \sim N(g_{n0}, \tau_n^2 K_n^{-1}), g_a|\tau_a^2, \sim N(g_{a0}, \tau_a^2 K_a^{-1}) \\
&\beta_j \sim N(b_{j0}, B_{j0}), \beta_n \sim N(b_{n0}, B_{n0}), \beta_a \sim N(b_{a0}, B_{a0}) \\
&\ln(\sigma_{ij}^2) = Z_{ij}\gamma_j, \ln(\sigma_{in}^2) = Z_{in}\gamma_n, \ln(\sigma_{ia}^2) = Z_{ia}\gamma_a \\
&\gamma_j \sim N(\gamma_{0j}, \Gamma_{0j}), \gamma_n \sim N(\gamma_{0n}, \Gamma_{0n}), \gamma_a \sim N(\gamma_{0a}, \Gamma_{0a}) \\
&\tau_j^2 \sim IG(\frac{t_{\nu j0}}{2}, \frac{t_{dj0}}{2}), \tau_n^2 \sim IG(\frac{t_{\nu n0}}{2}, \frac{t_{dn0}}{2}), \tau_a^2 \sim IG(\frac{t_{\nu a0}}{2}, \frac{t_{da0}}{2})
\end{aligned} \tag{25}$$

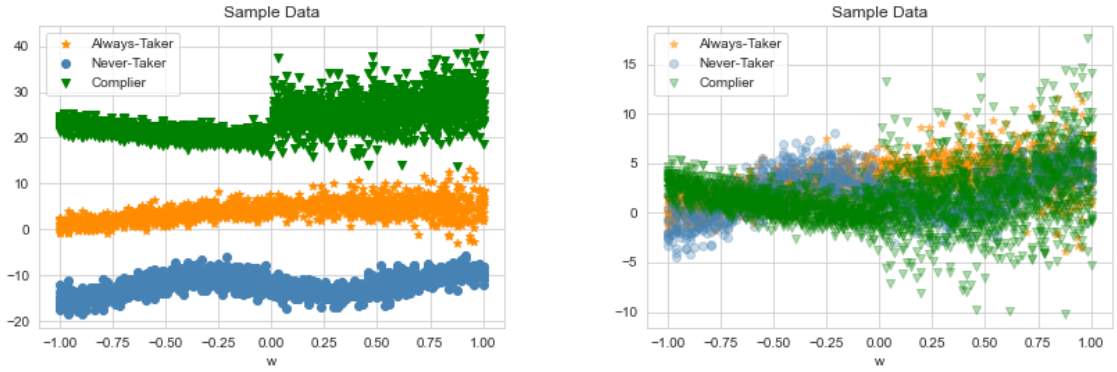
The posterior distribution for the type variable s is as follows:

$$\begin{aligned}
&Pr(s_i = c|y_i, g_j, \beta_j, \tau_j^2, \gamma_j) \propto q_c \phi(y_i|g_i, \beta_j, \tau_j^2, \gamma_j) \\
&Pr(s_i = n|y_i, g_n, \beta_n, \tau_n^2, \gamma_n) \propto q_n \phi(y_i|g_n, \beta_n, \tau_n^2, \gamma_n) \\
&Pr(s_i = a|y_i, g_a, \beta_a, \tau_a^2, \gamma_a) \propto q_a \phi(y_i|g_a, \beta_a, \tau_a^2, \gamma_a)
\end{aligned} \tag{26}$$

The joint posterior distrbution can be sampled using the MCMC methods. In each iteration, every observation can be categorized as complier, always-takers or never-taker. The alogrithm is outlined in Algorithm 3.

To illustrate our model, we consider two data samples. The simulated data are visualized in Figure 5. In the first sample, the data for each group are well-separated, while in the second sample, the data for each group are mixed together. The estimated nonparametric functions are presented in Figure 6 and Figure 7. It's important to highlight that we encountered the label switching problem during the MCMC algorithm. This issue arises when, during the MCMC process, some labels of the mixture components switch, especially in the samples with poorly separated clusters (Celeux, 1998). As of now, there is no consensus solution to address this challenge. Further research is needed to address this issue. It's noteworthy that the heteroskedastic model shows promise in enhancing performance and mitigating label switching concerns.

Figure 5: Simulated Data



(a) Well-separated Sample

(b) Poorly-separated Sample

3 Potential Outcome Framework

In this section, we introduce a potential outcome framework with 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 $T = \mathbb{1}\{T^* \geq 0\}$.

Algorithm 3 (Semi-parametric Fuzzy RDD Model)

---

(1) Sample the type variable s, where the posterior distribution of s is summarized by Equation (26).

(2) Sample $q = (q_a, q_n, q_c) \sim Dir(n_{a0}+n_a, n_{n0}+n_n, n_{c0}+n_c)$, where n_a, n_n and n_c are the sample size of the observations that are categorized into always-takers, never-takers and compliers correspondingly in the previous step.

(3) For all possible values of ν, we update g_j using the samples that were categorized as compliers in the previous step. We sample $[g_j|y_j, \beta_j, \tau_j^2, \gamma_j] \sim N(\hat{g}_j, \hat{G}_j)$, where $\hat{G}_j = (\frac{K_j}{\tau_j^2} + Q_j'\Omega_j^{-1}Q_j)^{-1}$ and $\hat{g}_j = \hat{G}_j(\frac{1}{\tau_j^2}K_j g_{j0} + Q_j'\Omega_j^{-1}(y_j - X_j\beta_j))$. We repeat this step for all the compliers and never-takers.

(4) Sample $[\beta_j|y_j, g_j, \gamma_j] \sim N(\hat{\beta}_j, \hat{B}_j)$, where $\hat{B}_j = (B_{j0}^{-1} + X_j'\Omega_j^{-1}X_j)^{-1}$, and $\hat{\beta}_j = \hat{B}_j(B_{j0}^{-1}b_{j0} + X_j'\Omega_j^{-1}(y_j - Q_j g_j))$. Sample $[\beta_a|y_a, g_a, \gamma_a])$, and $[\beta_n|y_n, g_n, \gamma_n]$ in a similar way.

(5) Sample $[\tau_j^2|g_j] \sim IG(\frac{t_{\nu_{j0}}+m_j}{2}, \frac{t_{d_{j0}}+(g_j-g_{j0})'K_j(g_j-g_{j0})}{2})$. Repeat this step to sample $[\tau_a^2|g_a]$ and $[\tau_n^2|g_n]$.

(6) Sample $[\gamma_j|y_j^*, g_j, \beta_j]$ using the Metropolis-Hasting algorithm.

(a) Define

$$e_{ij}^c = (y_{ij}^* - g_j^c(w_i) - X_i'\beta_j^c)^2$$
$$\eta_{ij}^c = Z_i'\gamma_j^c + \frac{e_{ij}^c - \sigma_{ij}^{2c}}{\sigma_{ij}^{2c}}, \quad \eta_j^c = (\eta_{1j}^c, ..., \eta_{n_j}^c)'.$$

where σ_{ij}^{2c}, g_j^c and β_j^c are the current values of σ_{ij}^2, g_j and β_j.

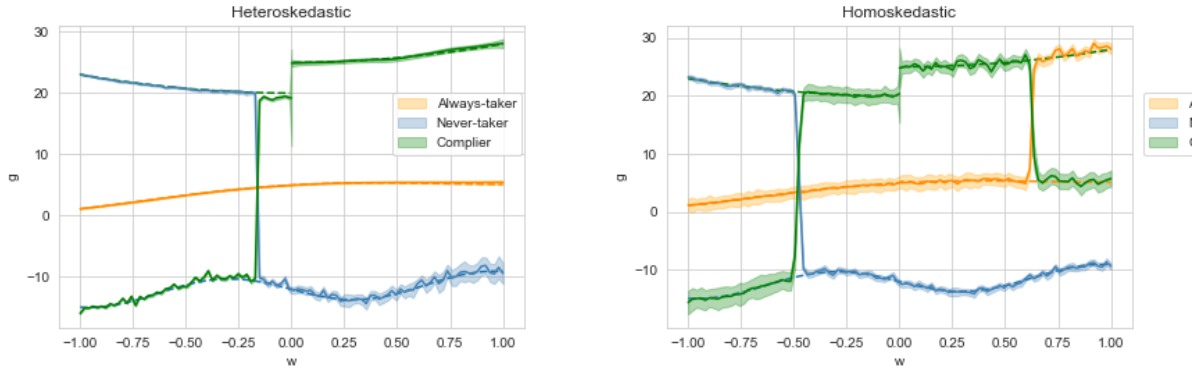
(b) Using the iteratively reweighted least squares algorithm, we obtain a proposal value γ_j^p from the proposal density, which is $T_4(\gamma_j^p|\hat{\gamma}_j^p, V_j^p)$, where $V_j^p = (\Gamma_{0j}^{-1} + \frac{1}{2}Z_j'Z_j)^{-1}$ and $\hat{\gamma}_j^p = V_j^p(\Gamma_{j0}^{-1}\gamma_{j0} + \frac{1}{2}Z_j'\eta_j^c)$. We obtain $q(\gamma_j^c|\hat{\gamma}_j^c, V_j^c)$ in the reverse direction moving from the proposed value to the current value.

(c) The acceptance rate α is defined as follows:

$$\alpha = \frac{f(y_j^*|g_j^c, \beta_j^c, \gamma_j^p)\pi(\gamma_j^p|\gamma_{j0}, \Gamma_{j0})}{f(y_j^*|g_j^c, \beta_j^c, \gamma_j^c)\pi(\gamma_j^c|\gamma_{j0}, \Gamma_{j0})} \times \frac{q(\gamma_j^c|\hat{\gamma}_j^c, V_j^c)}{q(\gamma_j^p|\hat{\gamma}_j^p, V_j^p)}$$

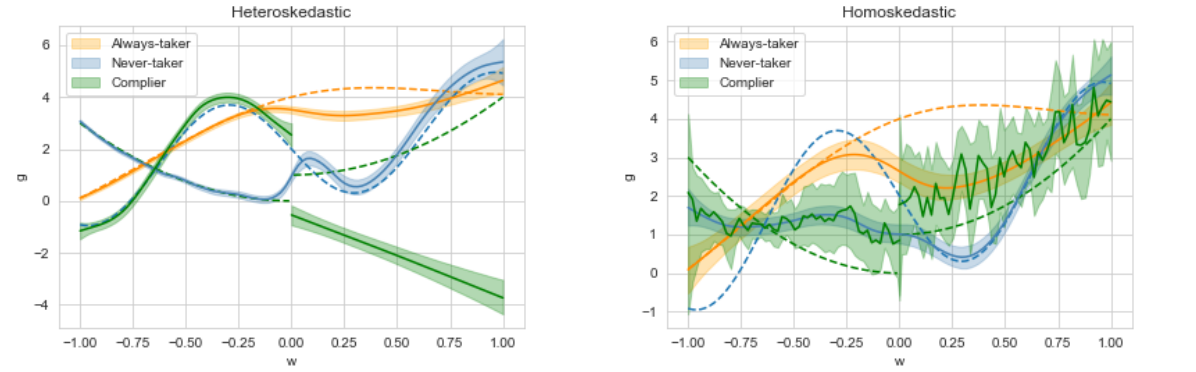
If the proposed value is not accepted, γ_j^c is repeated in the next iteration. We repeat this step for always-takers and never-takers to sample $[\gamma_a|y_a^*, g_j, \beta_a]$ and $[\gamma_n|y_n^*, g_n, \beta_n]$.

Figure 6: Well-separated Data



(a) Heteroskedastic (b) Homoskedastic

Figure 7: Poorly-separated Data



(a) Heteroskedastic (b) Homoskedastic

The model can be represented as outlined below:

$$Y_i = X_i\beta + \epsilon_i, \epsilon_i \sim N(0, \Omega_i) \tag{27}$$

where

$$Y_i = \begin{pmatrix} T_i^* \\ Y_{0i} \\ Y_{1i} \end{pmatrix}, X_i = \begin{pmatrix} x'_{Ti} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & x'_{0i} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & x'_{1i} \end{pmatrix}, \beta = \begin{pmatrix} \beta'_T \\ \beta'_0 \\ \beta'_1 \end{pmatrix}, \text{ and } \epsilon = \begin{pmatrix} \epsilon_{Ti} \\ \epsilon_{0i} \\ \epsilon_{1i} \end{pmatrix}$$

We use N_j to denote the sample $\{i : T_i = j\}$ and n_j to denote the cardinality of N_j, where $j \in \{0, 1\}$. The covariance matrix Ω_i is defined as:

$$\Omega_i = \begin{pmatrix} \omega_{TTi} & \omega_{T0i} & \omega_{T1i} \\ \omega_{0di} & \omega_{00i} & \omega_{01i} \\ \omega_{1di} & \omega_{10i} & \omega_{11i} \end{pmatrix} \tag{28}$$

Due to the missing data, we can't identify ω_{01i}.

To simplify the notation in the subsequent discussion, we introduce the following matrices:

$$\Omega_{0i} = \begin{pmatrix} \omega_{TTi} & \omega_{T0i} \\ \omega_{0Ti} & \omega_{00i} \end{pmatrix}, \Omega_1 = \begin{pmatrix} \omega_{TTi} & \omega_{T1i} \\ \omega_{1Ti} & \omega_{11i} \end{pmatrix} \tag{29}$$

$$J_0 = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix}, \text{ and } J_1 = \begin{pmatrix} \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix}$$

$$\tilde{X}_{i0} = \begin{pmatrix} x'_{Ti} & \mathbf{0} \\ \mathbf{0} & x'_{0i} \end{pmatrix}, \tilde{X}_{i1} = \begin{pmatrix} x'_{Ti} & \mathbf{0} \\ \mathbf{0} & x'_{1i} \end{pmatrix}, \tilde{y}_{i0} = \begin{pmatrix} T_i^* \\ y_{0i} \end{pmatrix}, \text{ and } \tilde{y}_{i1} = \begin{pmatrix} T_i^* \\ y_{1i} \end{pmatrix}$$

Thus we have $J_0\beta = (\beta'_T, \beta'_0)$ and $J_1\beta = (\beta'_T, \beta'_1)$.

The complete data density function is specified as:

$$f(Y_0, Y_1, T* | \beta_0, \beta_1, \Omega_0, \Omega_1) = \left[\prod_{i \in N_0} f(\tilde{y}_{i0} | \beta_0, \Omega_{i0}) \mathbb{1}\{T_i^* < 0\} \right] \left[\prod_{i \in N_1} f(\tilde{y}_{i1} | \beta_1, \Omega_{i1}) \mathbb{1}\{T_i^* \geq 0\} \right] \tag{30}$$

3.1 Homoskedastic Model

In the homoskedastic model, we impose a restrcition where $\omega_{TTi} = 1$ for the purpose of identification. We define the covariance matrices and several variables in the following manner:

$$\Omega_0 = \begin{pmatrix} 1 & \omega_{T0} \\ \omega_{0T} & \omega_{00} \end{pmatrix}, \Omega_1 = \begin{pmatrix} 1 & \omega_{T1} \\ \omega_{1T} & \omega_{11} \end{pmatrix} \tag{31}$$

$$\Omega_{22 \cdot 1} = \omega_{11} - \omega_{1T}\omega_{T1}$$

$$\Omega_{22 \cdot 0} = \omega_{00} - \omega_{0T}\omega_{T0}$$

If $\Omega_{22 \cdot j}$ is a matrix, we assume that the prior distribution for $\Omega_{22 \cdot j}$ is $\Omega_{22 \cdot j} \sim IW(\nu_j, O_j)$ for $j \in \{0, 1\}$. Otherwise, we assume that $\Omega_{22 \cdot j} \sim IG(\frac{\nu_j}{2}, \frac{O_j}{2})$. Under this assumption, the conditional distribution of $\omega_{jT} | \Omega_{jj \cdot 2} \sim N(q_j, \Omega_{22 \cdot j})$. Furthermore, we specify the prior distribution of β as β is $\beta \sim N(b_0, B_0)$.

The estimation algorithm is summarized in Alorightm 4.

3.2 Heteroskedastic Model

In this section, we expand the model to accommodate heteroskedasticity. Building upon the insights presented in Chan and Jeliazkov (2009b), we discussed the MCMC sampler for the covariance matrix as follows:

$$\Omega_{0i} = L_0 G_{0i} L'_0, \quad \Omega_{1i}^{-1} = L_1 G_{1i} L'_1,$$

Algorithm 4 (Bayesian Potential Outcome Framework Model with Homoskedasticity)

---

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

(2) Sample $T_i^* \sim TN(\mu_{Tij}, \hat{\omega}_{TT})$, where $T_i^* \in (-\infty, 0)$ if $i \in N_0$, $T_i^* \in [0, \infty)$ if $i \in N_1$, $\mu_{2ij} = x_{Ti}' \beta_T + \omega_{jT} \omega_{jj}^{-1} (y_{ji} - x_{ji}' \beta_j)$, and $\hat{\omega}_{22} = 1 - \omega_{jT} \omega_{jj}^{-1} \omega_{jT}$.

(3) For $i \in N_j$:

$$\pi(\omega_{j2} | \Omega_{22 \cdot j}, \beta, y_i, z_i) = \pi(\omega_{jT} | \Omega_{22 \cdot j}) \pi(y_j | z) = f_N(\omega_{jT} | q_t, \Omega_{22 \cdot j}) \prod_{i \in N_j} f_N(y_{ji} | \mu_{ji|2}, \Omega_{22 \cdot j})$$

where $\mu_{ji|2} = x_{ji}' \beta_j + \omega_{jT}(T_i^* - x_{iT}' \beta_T)$. Thus the posterior distribution for ω_{jT} is $\omega_{jT} \sim N(\hat{q}_j, \hat{\omega}_{22 \cdot j})$, where $\hat{\omega}_{22 \cdot j} = [\Omega_{22 \cdot j}^{-1} + (\sum_{i=1}^{n_j} (\epsilon_{Ti}^2 \Omega_{22 \cdot j}^{-1}))^{-1}]^{-1}$, and $\hat{q}_j = \hat{\omega}_{22 \cdot j} [\Omega_{22 \cdot j}^{-1} q_t + \Omega_{22 \cdot j}^{-1} \sum_{i=1}^{n_j} \epsilon_{Ti} \epsilon_{ji}]$, where $\epsilon_{ji} \equiv y_{ji} - x_{ji}' \beta_j$ and $\epsilon_{Ti} \equiv T_i^* - x_{Ti}' \beta_T$.

(4) For $i \in N_j$:

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

The posterior distribution is as follows: $\Omega_{22 \cdot j} \sim IG(\frac{\hat{\nu}}{2}, \frac{\hat{O}_j}{2})$, where $\hat{\nu} = \nu + 1 + n_j$, and $\hat{O}_j = O_j + (\omega_{jT} - q_t)^2 + \sum_{i \in N_j} (\epsilon_{ji} - \omega_{jT} \epsilon_{Ti})^2$.

---

where for $j \in \{0, 1\}$

$$L_j \equiv \begin{pmatrix} 1 & 0 \\ a_{jT} & 1 \end{pmatrix}, G_{ji} \equiv \begin{pmatrix} \lambda_{Ti} & 0 \\ 0 & \lambda_{ji} \end{pmatrix}$$

We can rewrite our model as follows:

$$\begin{pmatrix} T_i^* \\ y_{ji} \end{pmatrix} = \begin{pmatrix} x_{Ti}' & \mathbf{0} \\ \mathbf{0} & x_{ji}' \end{pmatrix} \begin{pmatrix} \beta_T \\ \beta_j \end{pmatrix} + L_j \begin{pmatrix} \psi_{Ti} \\ \psi_{ji} \end{pmatrix}, \text{where } \begin{pmatrix} \psi_{Ti} \\ \psi_{ji} \end{pmatrix} \sim N(\mathbf{0}, G_{ji}) \tag{32}$$

This model can be alternatively represented as:

$$\begin{aligned} T_i^* &= x_{Ti}' \beta_T + \psi_{Ti} \\ y_{ji} &= x_{ji}' \beta_j + a_{jT} \psi_{Ti} + \psi_{ji} \\ \psi_{ji} &\sim N(0, \lambda_{ji}), \psi_{Ti} \sim N(0, \lambda_{Ti}) \\ \lambda_{ji} &= \exp(Z_{ji}' \gamma_j), \text{and } \lambda_{Tj} = \exp(Z_{Ti}' \gamma_T) \end{aligned} \tag{33}$$

where Z_{ji} and Z_{zi} are covariate matrices that determine the variance.

The prior distributions are specified as follows:

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

The estimation algorithm is detailed in Algorithm 5.

3.3 Treatment Effects Estimation

In our model:

$$ATE = E(Y_1 - Y_0) = E(x_1'\beta_1 - x_0'\beta_0)$$
$$ATT = E(Y_1 - Y_0|D = 1) = E(x_1'\beta_1 - x_1'\beta_0)$$

ATE and ATT can be estimated by

$$\hat{ATE} = \frac{1}{M^2}\Big(\sum_{g=1}^{M}\sum_{l=1}^{M}\frac{1}{n}(\sum_{i=1}^{n} x_i'\beta_1^{(g)}) - \frac{1}{n}(\sum_{i=1}^{n} x_i'\beta_0^{(l)})\Big) \tag{34}$$

$$\hat{ATT} = \frac{1}{M^2}\Big(\sum_{g=1}^{M}\sum_{l=1}^{M}\frac{1}{n_1}(\sum_{i=1}^{n_1} x_{1i}'\beta_1^{(g)}) - \frac{1}{n_1}(\sum_{i=1}^{n_1} x_{1i}'\beta_0^{(l)})\Big) \tag{35}$$

where M is the total number of MCMC iterations, and $\beta_j^{(g)}$ and $\beta_j^{(l)}$ denote the sample draws of β_j in the g-th and l-th iterations of the MCMC sampling, respectively.

3.4 Model Comparison: Marginal Likelihood

In the homoskedastic case, the posterior ordinate can be estimated in a similar way as presented in Section 2.1.3 Equation (18).

To estimate the marginal likelihood in the heteroskedastic model, let's use θ to denote the parameters (β, a_{1T}, a_{0T}). The posterior ordinate of $\hat{\pi}(\theta^*|y, T^*)$ can also be estimated using Equation (18). The posterior orderdate of $[\gamma_T|y, \theta^*, T^*]$, $[\gamma_1|y, \theta^*, T^*]$, and $[\gamma_0|y, \theta^*, T^*]$ can be sampled in a similar way as outlined in Section 2.1.3 Equation (20).

3.5 Simulation

In this section, we performed a simulation study to assess the influence of heteroskedasticity in the context of potential outcome framework. We report mean, standard deviations, and 95% credible intervals of the posterior distribution for the treatment effects in each model.

Algorithm 5 (Bayesian Potential Outcome Framework Model with Heteroskedasticity)

---

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

(2) Sample $T_i^* \sim TN(\mu_{2ij}, \hat{\omega}_{22i})$, where $T_i^* \in (-\infty, 0)$ if $i \in N_0$, $T_i^* \in [0, \infty)$ if $i \in N_1$, $\mu_{2ij} = x_{Ti}'\beta_T + \omega_{jTi}\omega_{jji}^{-1}(y_{ji} - x_{ji}'\beta_j)$, and $\hat{\omega}_{TTi} = \omega_{TTi} - \omega_{jTi}\omega_{jji}^{-1}\omega_{jTi}$.

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

(4) Sample $[\gamma_T | a_{0T}, a_{1T}, \beta_T, T^*]$ using the Metropolis-Hasting algorithm.

(a) Define

$$e_{Ti}^c = (T_i^* - X_{Ti}'\beta_T^c)^2$$
$$\eta_{Ti}^c = Z_{Ti}'\gamma_T^c + \frac{e_{Ti}^c - \omega_{TTi}^c}{\omega_{TTi}^c}, \quad \eta_T^c = (\eta_{1j}^c, ..., \eta_n^c)'.$$

where ω_{TTi}^c and β_T^c are the current values of ω_{TTi} and β_T.

(b) Using the iteratively reweighted least squares algorithm, we obtain a proposal value γ_T^p from the proposal density, which is $T_\nu(\gamma_T^p | \hat{\gamma}_T^p, V_T^p)$, where $V_T^p = (\Gamma_{0T}^{-1} + \frac{1}{2}Z_T'Z_T)^{-1}$ and $\hat{\gamma}_T^p = V_T^p(\Gamma_{T0}^{-1}\gamma_{T0} + \frac{1}{2}Z_T'\eta_T^c)$. We obtain $q(\gamma_T^c | \hat{\gamma}_T^c, V_T^c)$ in the reverse direction moving from the proposed value to the current value.

(c) The acceptance rate α is defined as follows:

$$\alpha = \frac{f(T_i^* | a_{0T}, a_{1T}, \beta_T^c, \gamma_T^p)\pi(\gamma_T^p | \gamma_{T0}, \Gamma_{T0})}{f(T_i^* | a_{0T}, a_{1T}, \beta_T^c, \gamma_T^c)\pi(\gamma_T^c | \gamma_{T0}, \Gamma_{T0})} \times \frac{q(\gamma_T^c | \hat{\gamma}_T^c, V_T^c)}{q(\gamma_T^p | \hat{\gamma}_T^p, V_T^p)}$$

If the proposed value is not accepted, γ_T^c is repeated in the next iteration.

(5) We define e_{ji}^c and η_{ji}^c as follows:

$$e_{ji}^c = (y_{ji} - X_{ji}'\beta_j^c - a_{jT}^c\psi_{Ti})^2$$
$$\eta_{ji}^c = Z_{ji}'\gamma_j^c + \frac{e_{ji}^c - \omega_{jji}^c}{\omega_{jji}^c}, \quad \eta_j^c = (\eta_{j1}^c, ..., \eta_{jn_j}^c)'.$$

$[\gamma_j | a_{jT}, \beta_j]$ can be sampled in a similar manner as described in Step (4).

We run 3 simulation sets with sample size n = $\{500, 5000, 50000\}$. The data is generated using

the following DGP:

$$X_0 \sim (1, N(0,1)), X_1 = X_0, X_d = (X_0, N(0,1)); \beta \sim N(\mathbf{0}_{7\times1}, \mathbf{I}),$$

$$\gamma_0 \sim N(\mathbf{0}_{2\times1}, \mathbf{I}), \gamma_1 \sim N(\mathbf{0}_{2\times1}, \mathbf{I}), \gamma_d \sim N(\mathbf{0}_{3\times1}, \mathbf{I}), a_{0T} = -0.2 \text{ and } a_{1T} = 0.2$$

The estimated ATE and ATT are summarized by Table 9. In all cases, the heteroskedastic model is recommended based on the marginal likelihood results. These results highlight that when heteroskedasticiy exists, disregarding it leads to inconsistent and biased estimates of ATE and ATT.

Table 9: Treatment Effects Estimation (Simulation)

	Model		TRUE	Estimated	Std.Dev	95% CI	Marginal Likelihood
n = 500	Heteroskedastic	**ATE**	-2.3573	-2.3030	0.1910	(-2.6775, -1.9268)	-798.15
		ATT	-1.4987	-1.4693	0.2512	(-1.9608, -0.9737)	
	Homoskedastic	**ATE**	-2.3573	-1.5441	0.2865	(-2.1134, -0.9846)	-853.74
		ATT	-1.4987	-0.8222	-0.3577	(-1.5287, -0.1122)	
n = 5000	Heteroskedastic	**ATE**	-0.4869	-0.4704	0.1476	(-0.7504, -0.1610)	-7073.77
		ATT	-0.5669	-0.5367	0.1546	(-0.8296, -0.2127)	
	Homoskedastic	**ATE**	-0.4869	-1.4841	0.3421	(-2.1683, -0.8130)	-8108.51
		ATT	-0.5669	-1.5631	0.3610	(-2.2862, -0.8552)	
n = 50000	Heteroskedastic	**ATE**	-1.7010	-1.6709	0.0328	(-1.7359, -1.6069)	-77644.71
		ATT	-2.6833	-2.6595	0.0330	(-2.7244, -2.5950)	
	Homoskedastic	**ATE**	-1.7010	-3.1775	0.0403	(-3.2569, -3.0985)	-90039.30
		ATT	-2.6833	-3.3340	0.0308	(-3.3937, -3.2732)	

3.6 Empirical Application

We employ our model to examine the influence of private health insurance on healthcare expenditures, focusing on the elderly population. For individuals aged 65 and above, Medicare provides coverage, but some seniors opt to purchase private insurance, also known as Medigap policies, to supplement their Medicare benefits. Medigap policies typically offer enhanced coverage compared to the basic Medicare policy, and individuals often choose them in the belief that they can reduce out-of-pocket healthcare costs.

Given the onset of the COVID-19 pandemic towards the end of 2019, we partitioned the data into two distinct sets. One sample comprises survey data from 2020, while the other spans the years 2018 to 2019. 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 affect the purchase of the private insurance, and it would not affect thehealth care utilization directly. The variable definition summary of statistics are summarized by Table 10.

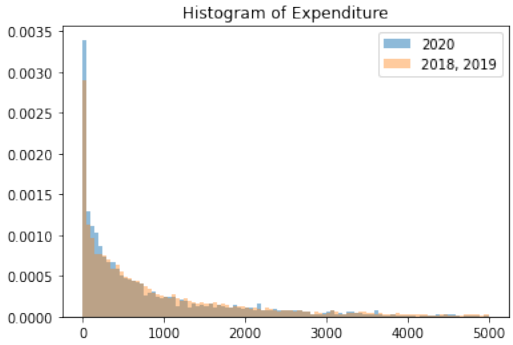
We standardized the variable age. We also transformed the expenditure using function that $Y_{new} = \frac{\sqrt{Y_i}}{1000}$ because, as indicated by Figure 8, the expenditure distributions are right-skewed in both samples. In addition, we assume that the variance of the selection only varies based on family income, and the variance of the expenditure depends on age and the number of chronic condition.

Table 10: Variable Definition and Summary Statistics

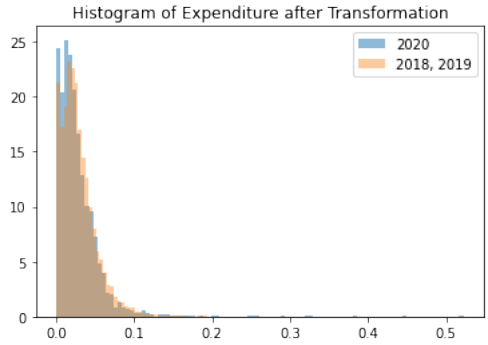
Variables	Description	2020 5019		2018, 2019 10226	
		Mean	Std.Dev.	Mean	Std.Dev.
AGE	AGE	73.52	6.24	73.65	6.42
FAMINC	Family income (as percentage of poverty line)	4.11	3.88	4.19	3.90
NUM_VISIT	# Office-based provider visits	10.11	14.53	12.19	16.63
NUM_CHRON	# Chronic conditions	3.93	2.26	3.82	2.25
EXCHLTH	=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 of self-perceived mental health is excellent	0.28	0.45	0.30	0.46
POORMHLTH	=1 of self-perceived mental health is poor	0.02	0.14	0.02	0.15
EMPLOYEED	=1 if the person is employed	0.20	0.40	0.19	0.39
PRIVATE	=1 if the person 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 the person is African American	0.12	0.33	0.13	0.33
MARRIED	=1 if the person is married	0.50	0.50	0.53	0.50
COLLEGE	=1 if the person has a college degree	0.33	0.47	0.30	0.46
MEDICAID	=1 if the person is covered by medicaid	0.14	0.35	0.14	0.34
ANYLIM	=1 if the person has a condition which limits activities of daily living	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 11. The heteroskedastic model suggests a negative impact of Medigap on health care expenditure, while the homoskedastic model indicates a positive impact. The marginal likelihood results recommend the heteroskedastic model for both samples. Based on the results, after we takes into consideration of the heteroskedasticity and self-selection, Medigap is expected to reduce the health care out-of-pocket expenditure. The detailed coefficients are reported in Appendix B.

Figure 8: Expenditure Distribution



(a) Expenditure Before Transformation



(b) Expenditure After Transformation

Table 11: Treatment Effects Estimation

Year	Model		Estimate	Std.	95% CI	Marginal Likelihood
2018, 2019	Homoskedastic	ATE	0.0162	0.0014	(0.0133, 0.0188)	22243.61
		ATT	0.0170	0.0013	(0.0144, 0.0195)	
	Heteroskedastic	ATE	-0.0058	0.0009	(-0.0076, -0.0040)	25365.03
		ATT	-0.0016	0.0009	(-0.0034, 0.0001)	
2020	Homoskedastic	ATE	0.0075	0.0029	(0.0023, 0.0139)	10602.64
		ATT	0.0066	0.0031	(0.0015, 0.0136)	
	Heteroskedastic	ATE	-0.0057	0.0011	(-0.0080, -0.0035)	12181.05
		ATT	-0.0014	0.0011	(-0.0036, 0.0008)	

4 Propensity Score Matching

Propensity score matching, as outlined by Rosenbaum and Rubin (1983), stands as a widely favored method for estimating causal treatment effects. This approach is instrumental in mitigating selection bias, effectively enabling the creation of comparable control and treatment groups by leveraging the propensity score as a balance score. We illustrate the relationships in Figure 9. Figure 10 shows how to estimate the treatment effects. Y_1 and Y_0 denote the observed treated and untreated outcome variables, and $\hat{Y}_1$ and $\hat{Y}_0$ denote the matched samples.

The average treatment effect can be calculated as:

$$\begin{aligned} ATE &= \frac{1}{N_{1,matched} + N_{0,matched}}[N_{1,matched}ATT + N_{0,matched}ATU] \\ &= \frac{1}{N_{1,mached} + N_{0,matched}}[\sum_{i \in N_1^*}(Y_{1i} - \hat{Y}_{0i}) + \sum_{i \ in N_0^*}(\hat{Y}_{1i} - Y_{0i})] \end{aligned} \tag{36}$$

where N_0^* represents the matched sample for each untreated units, and N_1^* denotes the matched

samples for each treated unit. In other words, for $i \in N_0^*$, we have a mathced pair $(\hat{Y}_{1i}, Y_{0i})$, and for $i \in N_1^*$, we have a mathed pair $(Y_{1i}, \hat{Y}_{0i})$. $N_{1,matched}$ and $N_{0,matched}$ are the cardinality of the sets N_1^* and N_0^* correspondingly.

Figure 9: Data Illustration

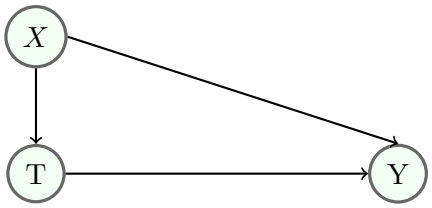
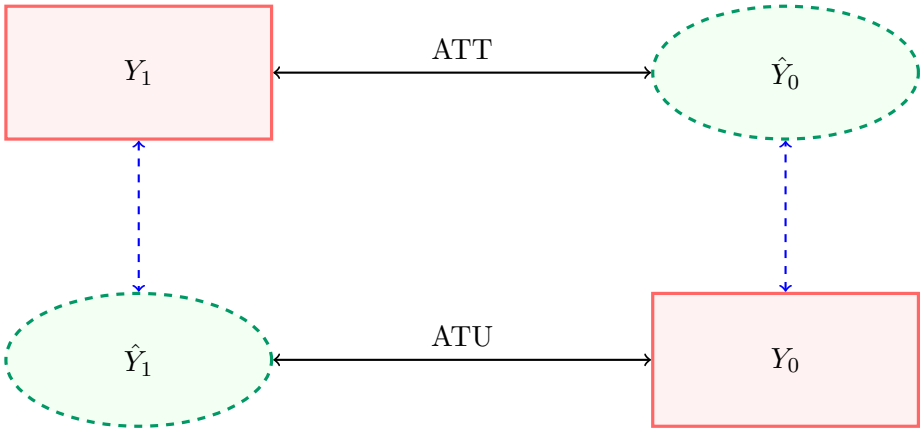


Figure 10: Treatment Effects Estimation



4.1 Estimation

In this section, we introduce a Beyesian probit model that accounts for heteroskedasticity in estimating the propensity score. We utilize simulation results to assess the consequences of disregarding heteroskedasticity when it is present in the treatment selection process within the context of propensity score matching.

The model is specified as follows:

$$T_i = \mathbb{1}\{T_i^* \geq 0\} = \mathbb{1}\{x_i'\beta + \nu_i \geq 0\}$$

In the homoskedastic model, we impose a constraint that variance of ν_i is 1 for the purpose of identification. In the heterosekdastic model, we assume that $var(\nu_i) = \exp(z_i'\gamma)$. We specify the prior distribution for β as $\beta \sim N(b_0, B_0)$, and for γ as $\gamma \sim N(\gamma_0, \Gamma_0)$.

The propensity score estimation algorithms are outlined in Algorithm 6 and 5.

Algorithm 6 (Bayesian Propensity Score Estimation with Homoskedasticity)

---

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

(2) Sample $T_i^* \sim TN(\mu_i, 1)$, where $T_i^* \in (-\infty, 0)$ if $T_i = 0$, and $T_i^* \in [0, \infty)$ if $T_i = 1$, where $\mu_i = x_i'\beta$, and the variance is 1.

Algorithm 7 (Bayesian Propensity Score Estimation with Heteroskedasticity)

---

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

(2) Sample $T_i^* \sim TN(\mu_i, \exp(z_i'\gamma))$, where $T_i^* \in (-\infty, 0)$ if $T_i = 0$, $T_i^* \in [0, \infty)$ if $T_i = 1$, $\mu_i = x_i'\beta$.

(3) Sample $[\gamma | \beta, T^*]$ using the Metropolis-Hasting algorithm.

(a) Define

$$e_i^c = (T_i^* - X_i'\beta^c)^2$$
$$\eta_i^c = Z_i'\gamma^c + \frac{e_{Ti}^c - exp(z_i'\gamma^c)}{\exp(z_i'\gamma^c)}, \quad \eta^c = (\eta_i^c, ..., \eta_n^c)'.$$

where γ^c and β^c are the current values of γ and β.

(b) Using the iteratively reweighted least squares algorithm, we obtain a proposal value γ^p from the proposal density, which is $T_\nu(\gamma^p | \hat{\gamma}^p, V^p)$, where $V^p = (\Gamma_0^{-1} + \frac{1}{2}Z'Z)^{-1}$ and $\hat{\gamma}^p = V^p(\Gamma_0^{-1}\gamma_0 + \frac{1}{2}Z'\eta^c)$. We obtain $q(\gamma^c | \hat{\gamma}^c, V^c)$ in the reverse direction moving from the proposed value to the current value. Z is defined as $Z \equiv (z_0', z_1, ..., z_n)'$, which contains covariates that determines the variance.

(c) The acceptance rate α is defined as follows:

$$\alpha = \frac{f(T_i^* | \beta^c, \gamma^p)\pi(\gamma^p | \gamma_0, \Gamma_0)}{f(T_i^* | \beta^c, \gamma^c)\pi(\gamma^c | \gamma_0, \Gamma_0)} \times \frac{q(\gamma^c | \hat{\gamma}^c, V^c)}{q(\gamma^p | \hat{\gamma}^p, V^p)}$$

If the proposed value is not accepted, γ^c is repeated in the next iteration.

4.2 Model Comparison: Marginal Likelihood

The posterior ordinate for the homosekdastic model can be calculated in the same way as presented in Section 2.1.3 Equation (18).

In the heteroskedastic case, the posterior ordinate of $\hat{\pi}(\beta^*|T, T^*)$ can also be estimated using Equation (18). The posterior orderdate of $[\gamma|T, \beta^*, T^*]$ can be sampled in the same way as outlined in Section 2.1.3 Equation (20).

4.3 Simulation

Emprical researchers normally incorporate the higher-order and interaction terms to improve the balance of the matched samples if it failed in the begining.(see e.g. Dehejia and Wahba, 1999; Caliendo and Kopeinig, 2008). In this section, we illustrate that neglecting heteroskedasticity when it exists can lead to the emergence of imbalanced samples. We also state the hypothesis that incorporating higher-order or interaction terms to enhance balance might be attributed to the presence of heteroskedasticity.

We utilize the sample data used in Dehejia and Wahba (1999) for the simulation study. This data set is a combination of data from 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, education (educ), if the subject is Black or Hispanic, if the subject is married (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 observatios in the control group.

We assume that DGP for the treatment assignment is:

$$\begin{aligned} T_i = \{ & -2 - 0.17age_i - 0.001educ_i + 0.3744nodegree_i - 0.9630married_i + \\ & 1.2285black_i + 1.219hispanic_i - 0.000005RE74_i - 0.0001RE74_i + \nu_i \geq 0\} \end{aligned} \tag{37}$$

where $\nu_i \sim N(0, age_i)$.

We use the standardized mean difference (SMD) as a balance measure. (Rosenbaum and Rubin (1985) and Thoemmes (2012)). A SMD excedding 0.1 can be considered as a sign of imbalance. (Zhang et al. (2019)). SMD is calculated as follows:

$$SMD = \frac{\bar{X}_T - \bar{X}_C}{\sqrt{\frac{S_T^2 + S_C^2}{2}}} \tag{38}$$

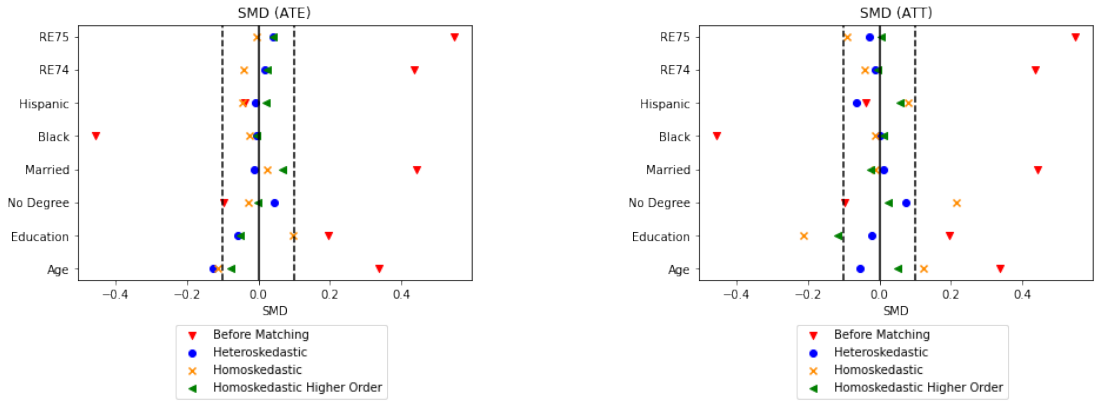
where $\bar{X}_T$ and $\bar{X}_C$ are the sample averages, and S_T^2 and S_C^2 are the standard deviations for the treatment and control groups correspondingly.

We employ nearest nighbour matching with replacment with a radius of 0.2 times the standard deviation of the estimated propensity score. (Austin (2011) and Chaudhuri and Howley (2022))

Three models are estimated: the heteroskedastic model with correct specification, the homosekdastic model with all the covariates, and the homoskedastic model with all the covariates, age^2 and all the interaction terms between age and other covariates.

Figure 11 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 exxample, falls short of achieving balance in the ATT estimation sample. However, by including higher order and interaction terms, we can improve the balance.

Figure 11: SMD



(a) SMD (ATE)

(b) SMD (ATT)

4.4 Empirical Application

We use the sample data used in Chaudhuri and Howley (2022) to evaluate the impact of COVID-19 vaccination on mental health. The treatment variable is if the subject received any dose of COVID-19 vaccine. This is a sample of wave 7 and 8 of COVID-19 surveys by the UK Household Longitudinal Study (University of Essex, Institute for Social and Economic Research, 2021). This survey recorded the vaccination, demographic and mental health information.

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, it is important to note that the GHQ scores have been reversed, such that a score of 36 indicates the highest level of mental health, while a score of 0 signifies a markedly diminished level of mental well-being in the evaluated individuals. The summary of statistics of the key variables are summarized in Table 12.

Table 12: Summary of Statistics

	Control Group (12423)		**Treatment Group (9562)**	
Variables	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
GHQ-12	23.1388	6.1476	24.0396	5.6400
AGE	49.1924	15.6416	61.6679	13.8353
Burn in UK	0.8653	0.3414	0.9012	0.2984
Clinical Vulnerable	0.3417	0.4743	0.5646	0.4958
Male	0.4204	0.4936	0.4147	0.4927
Key Worker	0.2491	0.4325	0.2509	0.4335
Couple	0.6871	0.4637	0.7287	0.4446
Willingness to take vaccine	0.9059	0.2920	0.9507	0.2164

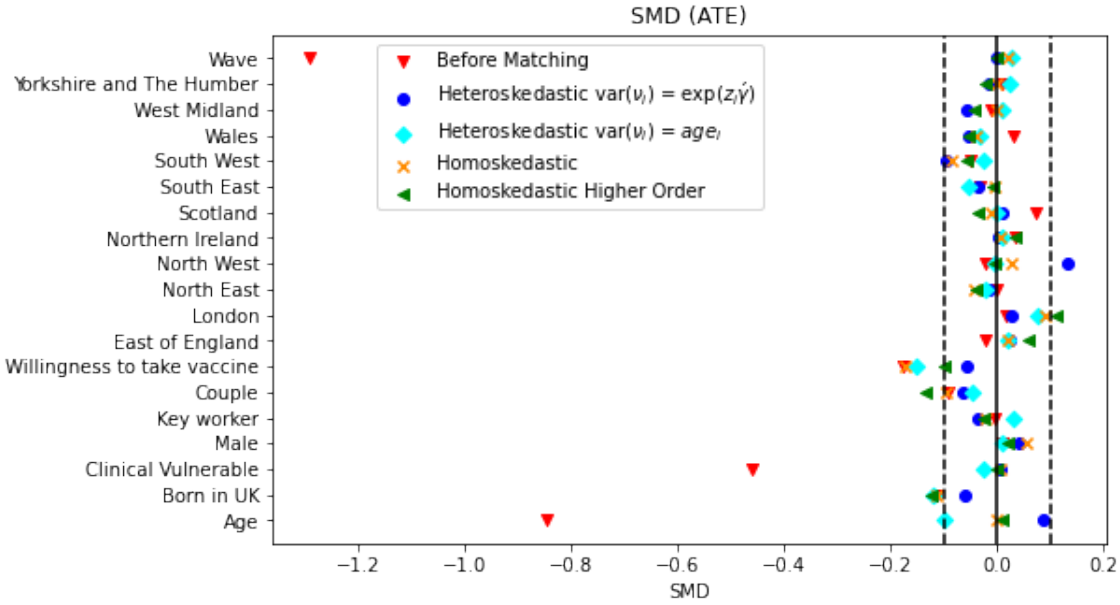
We estimated the treament effects using four models. The first one is the heteroskedastic model with variance $var(\nu_i) = \exp(z_i'\gamma)$, while the second model is more parsimonous with variance $var(\nu_i) = age_i$. The third model is a homoskedastic model with all the covariates, and the fourth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

The estimated impact of COVID-19 vaccination on mental well-being is presented in Table 13. The marginal likelihood results suggest that the heteroskedastic model with $var(\nu_i) = \exp(z_i'\gamma)$ fits the data better. All the results suggest that COVID-19 Vaccination is expected to improve the mental health. Figure 12 shows the SMD before and after matching. The figure shows that for the ATE samples, all the models can improve the balance. For the ATT samples, the heteroskedastic model with $var(\nu_i) = \exp(z_i'\gamma)$ performs the best.

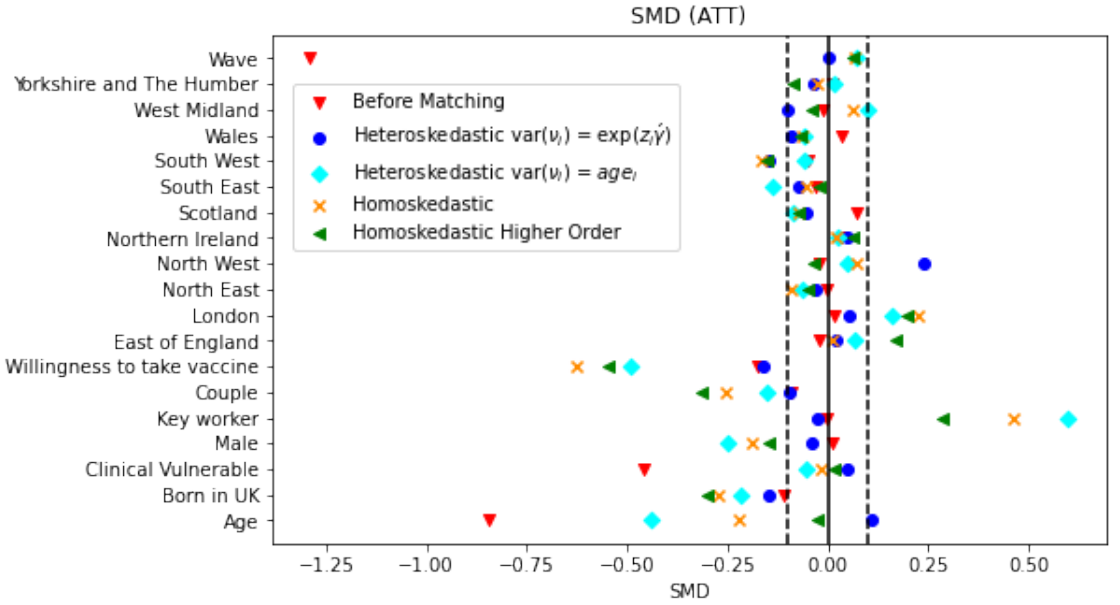
Table 13: Impact of COIVD-19 Vaccination on Mental Health

	ATE			**ATT**			
Model	**Mean**	**Std.dev**	**95% CI**	**Mean**	**Std.dev**	**95% CI**	**Marginal Likelihood**
Heteroskedastic ($var(\nu_i) = \exp(z_i'\gamma)$)	1.2018	0.2206	(0.7585, 1.6391)	2.5916	0.4681	(1.6778, 3.5173)	-7687.86
Heteroskedastic ($var(\nu_i) = age_i^2$)	0.5227	0.1232	(0.2993, 0.7653)	1.3593	0.2237	(0.9183, 1.7994)	-9337.68
Homoskedastic	0.3851	0.1179	(0.1525, 0.6109)	1.1477	0.2229	(0.7168, 1.6037)	-8753.30
Homoskedastic (Higher-order)	0.3628	0.1251	(0.1306, 0.6353)	0.9407	0.2316	(0.5139, 1.4600)	-8030.50

Figure 12: SMD (COVID-19)



(a) SMD (ATE Sample)



(b) SMD (ATT Sample)

5 Conclusion

In this paper, we have developed Bayesian treatment models and their corresponding estimation algorihtms. Utilizing simulation results, we have assessed the impact of heteroskedasticity across various causal models, and found that the presence of heteroskedasticity should not be ignored. Additionally, we have presented methods for estimating the marginal likelihood to facilitate formal model comparisons. We have utilized those models in applications and found evidence that supports the presence of heteroskedasticity.

In the sharp RDD setting, our investigation has revealed that disregarding heteroskedasticity within the dataset when it presents can lead to a biased estimation of treatment effects. We has employed the model to evaluate the impact of academic probation on subsequent semester GPAs and graduation rates. We found that the academic probation is expected to improve the students subsequent term GPAs, while the impact on graduation rate is not obvious.

Our findings have emphasized that neglecting heteroskedasticity within the dataset may lead to a biased estimation of the treatment effect within the potential outcome framework. We assessed the impact of Medigap on out-of-pocket healthcare expenditures using the model. Our model, accounting for both self-selection and heteroskedasticity, demonstrates that, on average, Medigap is expected to lower the out-of-pocket healthcare expenditures.

Furthermore, our study has revealed that overlooking heteroskedasticity has the potential to generate an imbalanced matched samples in propensity score matching. Nonetheless, this issue can be effectively mitigated by the inclusion of higher-order and interaction terms. We used our model to evaluate the impact of COVID-19 vaccination on mental health conditions, and our analysis indicated that on average, COVID-19 vaccination improves mental health.

A Sharp Regression Discontinutiy Design: Other Simulation Results

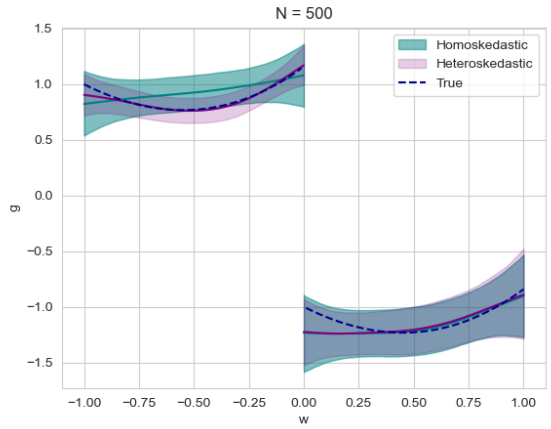
Table 14: ATE with Continuous Outcome Variable ($\beta_1 = \beta_0$)

	Model	True ATE	RD ATE	SD	95% CI	Marginal Likelihood
n = 500	Homoskedastic	-2.1585	-2.3144	0.2276	(-2.7632, -1.8710)	-784.62
	Heteroskedastic	-2.1585	-2.3964	0.1764	(-2.7449, -2.0534)	-680.02
	RDRobust	-2.1585	-2.3896	0.2250	(-2.8610, -1.8196)	
n = 5000	Homoskedastic	-2.1585	-2.0998	0.1131	(-2.3201, -1.8770)	-7811.57
	Heteroskedastic	-2.1585	-2.1142	0.1097	(-2.3281, -1.8978)	-7621.34
	RDRobust	-2.1585	-1.9219	0.1603	(-2.2203, -1.5017)	
n = 50000	Homoskedastic	-2.1585	-2.148	0.0393	(-2.2242, -2.0698)	-54592.15
	Heteroskedastic	-2.1585	-2.1579	0.0292	(-2.2149, -2.1002)	-48880.91
	RDRobust	-2.1585	-2.1474	0.0323	(-2.2132, -2.0644)	

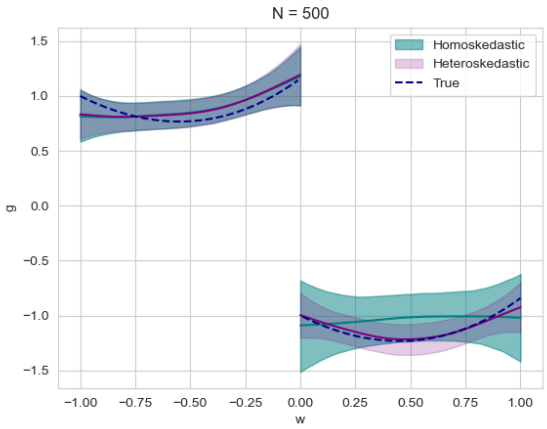
Table 15: ATE with Continuous Outcome Variable (Non-informative Priors)

	Model	True ATE	RD ATE	SD	95% CI	Marginal Likelihood
n = 500	Homoskedastic	-2.2357	-2.3684	0.2494	(-2.8652, -1.8849)	-806.77
	Heteroskedastic	-2.2357	-2.4144	0.2309	(-2.8728, 1.9686)	-1035.63
	RDRobust	-2.2357	-2.6305	0.6780	(-4.1775, -1.0734)	
n = 5000	Homoskedastic	-2.2195	-2.0407	0.1271	(-2.2888, -1.7895)	-8739.52
	Heteroskedastic	-2.2195	-2.0941	0.1061	(-2.2996, -1.8834)	-7965.18
	RDRobust	-2.2195	-2.1892	0.2142	(-2.7119, -1.7242)	
n = 50000	Homoskedastic	-2.2092	-2.1986	0.0409	(-2.2793, -2.1191)	-68740.12
	Heteroskedastic	-2.2092	-2.1991	0.0288	(-2.2557, -2.1427)	-59082.58
	RDRobust	-2.2092	-2.0633	0.0690	(-2.2046, -1.8831)	

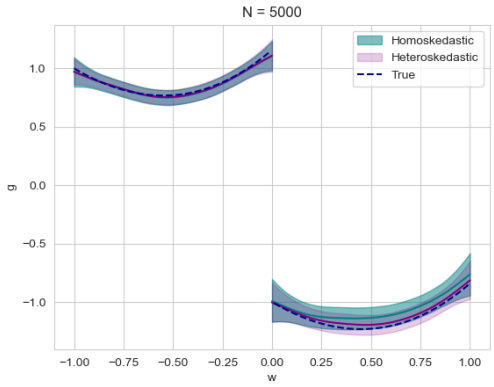
Figure 13: $\hat{g}$ with Continuous Outcome Variable



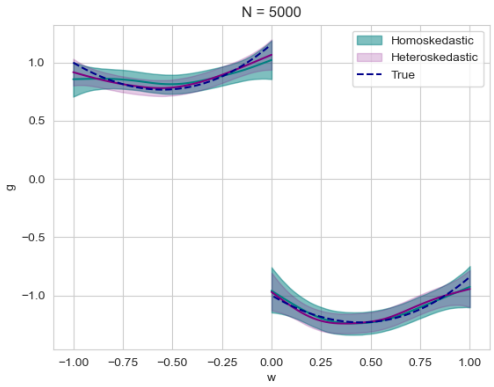
(a) $\beta_1 = \beta_0$, N = 500



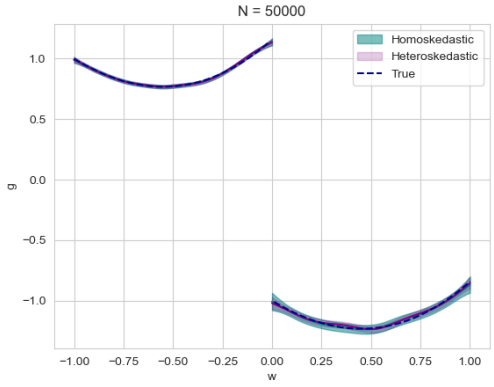
(b) $\beta_1 \neq \beta_0$, N = 500



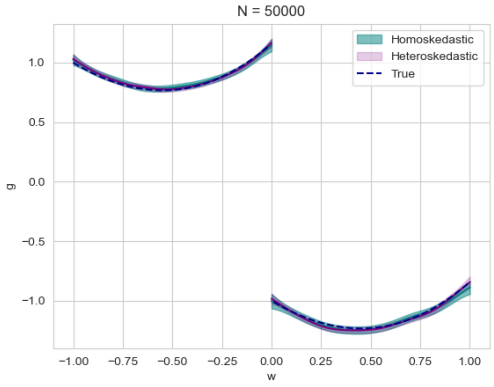
(c) $\beta_1 = \beta_0$, N = 5000



(d) $\beta_1 \neq \beta_0$, N = 5000



(e) $\beta_1 = \beta_0$, N = 50000



(f) $\beta_1 \neq \beta_0$, N = 50000

B MEPS: Coefficients

Table 16: Heteroskedastic Model (2020)

	Parameter	Mean	Std	95% CI			Parameter	Mean	Std	95% CI	
	const	-0.7584	0.0986	-0.9532	-0.5719	γ_T	FAMINC	0.1743	0.0076	0.1594	0.1893
	AGE	-0.0270	0.0321	-0.0893	0.0370		const	-8.6130	0.1125	-8.8395	-8.3974
	FAMINC	0.1734	0.0150	0.1444	0.2032	γ_0	AGE	0.4661	0.0513	0.3662	0.5676
	NUM_VISIT	0.0016	0.0019	-0.0022	0.0054		NUM_CHRON	0.0512	0.0263	-0.0052	0.1001
	NUM_CHRON	0.0123	0.0144	-0.0157	0.0406		const	-9.1443	0.1706	-9.4832	-8.8163
	EXCHLTH	0.0111	0.0991	-0.1841	0.2049	γ_1	AGE	0.4260	0.0743	0.2788	0.5715
	POORHLTH	0.0431	0.1605	-0.2715	0.3550		NUM_CHRON	0.0673	0.0327	0.0019	0.1296
	EXCMHLTH	-0.0145	0.0780	-0.1669	0.1387						
β_T	POORMHLTH	-0.4186	0.2427	-0.9074	0.0434						
	EMPLOYEED	0.1753	0.0899	0.0007	0.3535						
	NORTHEAST	-0.0914	0.0894	-0.2693	0.0821						
	MIDWEST	0.1390	0.0818	-0.0176	0.3036						
	WEST	-0.1954	0.0853	-0.3637	-0.0300						
	MALE	-0.0746	0.0627	-0.1975	0.0494						
	BLACK	-0.0576	0.0942	-0.2406	0.1301						
	MARRIED	0.0515	0.0659	-0.0781	0.1825						
	COLLEGE	0.0568	0.0711	-0.0841	0.1933						
	MEDICAID	-0.9967	0.1169	-1.2317	-0.7726						
	ANYLIM	0.0425	0.0659	-0.0875	0.1697						
	const	0.0036	0.0013	0.0011	0.0062						
	AGE	0.0009	0.0005	0.0000	0.0018						
	NUM_VISIT	0.0004	0.0000	0.0004	0.0005						
	NUM_CHRON	0.0012	0.0002	0.0008	0.0016						
	EXCHLTH	-0.0025	0.0013	-0.0050	0.0000						
	POORHLTH	0.0002	0.0020	-0.0037	0.0041						
	EXCMHLTH	-0.0002	0.0010	-0.0022	0.0019						
	POORMHLTH	0.0009	0.0027	-0.0044	0.0061						
b_0	EMPLOYEED	-0.0048	0.0012	-0.0071	-0.0025						
	NORTHEAST	0.0025	0.0012	0.0002	0.0047						
	MIDWEST	0.0000	0.0011	-0.0022	0.0022						
	WEST	0.0017	0.0011	-0.0005	0.0038						
	MALE	-0.0003	0.0008	-0.0019	0.0013						
	BLACK	-0.0021	0.0012	-0.0044	0.0002						
	MARRIED	-0.0034	0.0009	-0.0050	-0.0017						
	COLLEGE	-0.0013	0.0010	-0.0032	0.0007						
	MEDICAID	-0.0038	0.0011	-0.0059	-0.0017						
	ANYLIM	0.0047	0.0009	0.0030	0.0065						
	const	-0.0042	0.0017	-0.0075	-0.0008						
	AGE	0.0008	0.0005	-0.0002	0.0019						
	NUM_VISIT	0.0004	0.0000	0.0003	0.0004						
	NUM_CHRON	0.0014	0.0002	0.0009	0.0018						
	EXCHLTH	-0.0002	0.0014	-0.0029	0.0024						
	POORHLTH	0.0002	0.0029	-0.0056	0.0059						
	EXCMHLTH	0.0004	0.0011	-0.0018	0.0025						
	POORMHLTH	-0.0054	0.0048	-0.0149	0.0037						
b_1	EMPLOYEED	0.0046	0.0011	0.0025	0.0067						
	NORTHEAST	0.0019	0.0014	-0.0008	0.0046						
	MIDWEST	0.0046	0.0012	0.0022	0.0069						
	WEST	-0.0012	0.0012	-0.0036	0.0012						
	MALE	-0.0038	0.0009	-0.0056	-0.0020						
	BLACK	-0.0043	0.0015	-0.0073	-0.0013						
	MARRIED	0.0014	0.0010	-0.0005	0.0033						
	COLLEGE	0.0043	0.0010	0.0024	0.0061						
	MEDICAID	-0.0250	0.0026	-0.0302	-0.0201						
	ANYLIM	0.0051	0.0010	0.0031	0.0070						
	a_{0T}	-0.0147	0.0007	-0.0160	-0.0134						
	a_{1T}	0.0175	0.0007	0.0163	0.0188						

Table 17: Heteroskedastic Model (2018, 2019)

	Parameter	Mean	Std	95% CI	
β_T	const	-0.7571	0.0636	-0.8834	-0.6312
	AGE	-0.0052	0.0207	-0.0456	0.0349
	FAMINC	0.1512	0.0091	0.1334	0.1691
	NUM_VISIT	0.0020	0.0011	-0.0001	0.0041
	NUM_CHRON	0.0191	0.0093	0.0012	0.0376
	EXCHLTH	-0.0787	0.0610	-0.1980	0.0405
	POORHLTH	-0.0513	0.0915	-0.2332	0.1261
	EXCMHLTH	0.0584	0.0493	-0.0380	0.1543
	POORMHLTH	0.0874	0.1273	-0.1648	0.3352
	EMPLOYEED	0.1563	0.0596	0.0391	0.2740
	NORTHEAST	-0.0607	0.0584	-0.1747	0.0540
	MIDWEST	0.1376	0.0528	0.0324	0.2404
	WEST	-0.0801	0.0531	-0.1852	0.0232
	MALE	-0.0064	0.0407	-0.0860	0.0725
	BLACK	-0.0328	0.0602	-0.1503	0.0863
	MARRIED	0.1239	0.0417	0.0406	0.2057
	COLLEGE	0.1140	0.0491	0.0176	0.2073
	MEDICAID	-1.0296	0.0750	-1.1787	-0.8822
	ANYLIM	-0.0015	0.0435	-0.0879	0.0846
β_0	const	0.0103	0.0010	0.0084	0.0122
	AGE	0.0016	0.0003	0.0010	0.0022
	NUM_VISIT	0.0003	0.0000	0.0003	0.0003
	NUM_CHRON	0.0009	0.0001	0.0006	0.0011
	EXCHLTH	-0.0022	0.0009	-0.0040	-0.0005
	POORHLTH	0.0014	0.0013	-0.0011	0.0039
	EXCMHLTH	0.0001	0.0007	-0.0014	0.0015
	POORMHLTH	-0.0007	0.0019	-0.0044	0.0029
	EMPLOYEED	-0.0055	0.0009	-0.0072	-0.0038
	NORTHEAST	0.0023	0.0008	0.0007	0.0040
	MIDWEST	-0.0003	0.0008	-0.0018	0.0013
	WEST	0.0011	0.0008	-0.0005	0.0026
	MALE	-0.0034	0.0006	-0.0045	-0.0022
	BLACK	-0.0022	0.0008	-0.0038	-0.0005
	MARRIED	-0.0032	0.0006	-0.0044	-0.0020
	COLLEGE	0.0008	0.0008	-0.0007	0.0022
	MEDICAID	-0.0078	0.0008	-0.0094	-0.0061
	ANYLIM	0.0046	0.0006	0.0033	0.0058
β_1	const	-0.0041	0.0012	-0.0064	-0.0018
	AGE	0.0012	0.0004	0.0005	0.0020
	NUM_VISIT	0.0004	0.0000	0.0003	0.0004
	NUM_CHRON	0.0017	0.0002	0.0014	0.0020
	EXCHLTH	-0.0013	0.0009	-0.0031	0.0004
	POORHLTH	0.0025	0.0017	-0.0009	0.0059
	EXCMHLTH	0.0015	0.0007	0.0001	0.0030
	POORMHLTH	0.0030	0.0025	-0.0020	0.0079
	EMPLOYEED	0.0054	0.0008	0.0039	0.0068
	NORTHEAST	0.0031	0.0009	0.0013	0.0048
	MIDWEST	0.0044	0.0008	0.0029	0.0060
	WEST	0.0000	0.0008	-0.0016	0.0017
	MALE	-0.0034	0.0006	-0.0046	-0.0022
	BLACK	-0.0060	0.0010	-0.0079	-0.0040
	MARRIED	0.0032	0.0007	0.0019	0.0046
	COLLEGE	0.0065	0.0007	0.0052	0.0079
	MEDICAID	-0.0264	0.0018	-0.0298	-0.0229
	ANYLIM	0.0022	0.0007	0.0008	0.0035
	a_{0T}	-0.0140	0.0006	-0.0151	-0.0128
	a_{1T}	0.0190	0.0006	0.0179	0.0201

	Parameter	Mean	Std	95% CI	
γ_T	FAMINC	0.1347	0.0055	0.1240	0.1457
γ_0	const	-8.4420	0.0786	-8.5970	-8.2916
	AGE	0.3319	0.0311	0.2727	0.3930
	NUM_CHRON	0.0452	0.0138	0.0181	0.0719
γ_1	const	-8.9472	0.1254	-9.2025	-8.7126
	AGE	0.5289	0.0501	0.4290	0.6243
	NUM_CHRON	0.0300	0.0226	-0.0136	0.0744

Table 18: Homoskedastic Model

		2020				2018, 2019			
	Parameter	Mean	Std	95% CI		Mean	Std	95% CI	
β_T	const	-0.5358	0.0632	-0.6607	-0.4121	-0.4230	0.0431	-0.5069	-0.3388
	AGE	-0.0383	0.0202	-0.0773	0.0020	-0.0285	0.0141	-0.0557	-0.0004
	FAMINC	0.0652	0.0061	0.0534	0.0771	0.0483	0.0036	0.0412	0.0554
	NUM_VISIT	0.0023	0.0014	-0.0004	0.0050	0.0012	0.0008	-0.0003	0.0028
	NUM_CHRON	0.0149	0.0093	-0.0032	0.0328	0.0143	0.0063	0.0020	0.0267
	EXCHLTH	0.0139	0.0578	-0.0991	0.1266	-0.0540	0.0386	-0.1310	0.0208
	POORHLTH	-0.0268	0.1084	-0.2380	0.1874	-0.0843	0.0655	-0.2127	0.0430
	EXCMHLTH	0.0189	0.0461	-0.0702	0.1084	0.0484	0.0312	-0.0112	0.1096
	POORMHLTH	-0.3962	0.1640	-0.7170	-0.0755	0.0069	0.0919	-0.1762	0.1891
	EMPLOYEED	0.2225	0.0504	0.1244	0.3218	0.2120	0.0358	0.1424	0.2816
	NORTHEAST	0.0283	0.0555	-0.0801	0.1371	-0.0053	0.0381	-0.0785	0.0703
	MIDWEST	0.1533	0.0508	0.0526	0.2521	0.1160	0.0357	0.0468	0.1862
	WEST	-0.1175	0.0511	-0.2160	-0.0182	-0.0710	0.0348	-0.1399	-0.0032
	MALE	-0.0560	0.0389	-0.1324	0.0199	-0.0066	0.0264	-0.0582	0.0449
	BLACK	-0.0461	0.0625	-0.1686	0.0775	-0.0333	0.0421	-0.1147	0.0497
	MARRIED	0.1168	0.0409	0.0356	0.1948	0.1613	0.0279	0.1081	0.2159
	COLLEGE	0.1204	0.0435	0.0366	0.2053	0.1667	0.0301	0.1080	0.2256
	MEDICAID	-1.0880	0.0747	-1.2330	-0.9413	-1.0081	0.0518	-1.1115	-0.9081
	ANYLIM	0.0093	0.0424	-0.0742	0.0931	-0.0292	0.0289	-0.0854	0.0277
β_0	const	0.0113	0.0024	0.0059	0.0156	0.0091	0.0012	0.0067	0.0114
	AGE	0.0007	0.0005	-0.0004	0.0017	0.0016	0.0004	0.0009	0.0023
	NUM_VISIT	0.0005	0.0000	0.0005	0.0006	0.0004	0.0000	0.0003	0.0004
	NUM_CHRON	0.0010	0.0002	0.0005	0.0014	0.0007	0.0002	0.0004	0.0011
	EXCHLTH	-0.0024	0.0016	-0.0055	0.0008	-0.0014	0.0011	-0.0035	0.0006
	POORHLTH	-0.0023	0.0024	-0.0071	0.0024	0.0014	0.0015	-0.0017	0.0044
	EXCMHLTH	0.0003	0.0013	-0.0021	0.0028	0.0006	0.0009	-0.0011	0.0022
	POORMHLTH	0.0001	0.0032	-0.0060	0.0064	-0.0005	0.0022	-0.0048	0.0037
	EMPLOYEED	-0.0004	0.0016	-0.0035	0.0027	-0.0039	0.0010	-0.0059	-0.0019
	NORTHEAST	0.0032	0.0014	0.0004	0.0060	0.0030	0.0010	0.0010	0.0049
	MIDWEST	0.0019	0.0015	-0.0010	0.0048	0.0006	0.0010	-0.0013	0.0025
	WEST	0.0022	0.0013	-0.0004	0.0048	0.0025	0.0009	0.0007	0.0043
	MALE	-0.0004	0.0010	-0.0024	0.0016	-0.0033	0.0007	-0.0047	-0.0019
	BLACK	-0.0037	0.0015	-0.0066	-0.0008	-0.0030	0.0010	-0.0050	-0.0010
	MARRIED	0.0001	0.0011	-0.0022	0.0022	-0.0021	0.0007	-0.0036	-0.0006
	COLLEGE	0.0050	0.0013	0.0024	0.0075	0.0042	0.0009	0.0025	0.0059
	MEDICAID	-0.0119	0.0020	-0.0154	-0.0075	-0.0087	0.0010	-0.0107	-0.0066
	ANYLIM	0.0058	0.0011	0.0036	0.0079	0.0051	0.0008	0.0036	0.0066
β_1	const	0.0213	0.0028	0.0157	0.0266	0.0204	0.0017	0.0170	0.0238
	AGE	0.0015	0.0007	0.0001	0.0028	0.0011	0.0004	0.0002	0.0019
	NUM_VISIT	0.0004	0.0000	0.0003	0.0005	0.0004	0.0000	0.0003	0.0004
	NUM_CHRON	0.0010	0.0003	0.0003	0.0016	0.0012	0.0002	0.0008	0.0016
	EXCHLTH	-0.0008	0.0018	-0.0044	0.0028	0.0001	0.0011	-0.0020	0.0022
	POORHLTH	-0.0001	0.0041	-0.0080	0.0080	0.0043	0.0021	0.0001	0.0083
	EXCMHLTH	0.0005	0.0015	-0.0024	0.0034	0.0006	0.0009	-0.0011	0.0023
	POORMHLTH	0.0063	0.0069	-0.0075	0.0196	0.0035	0.0029	-0.0022	0.0092
	EMPLOYEED	0.0015	0.0016	-0.0015	0.0046	0.0025	0.0010	0.0007	0.0044
	NORTHEAST	0.0025	0.0018	-0.0011	0.0061	0.0035	0.0011	0.0013	0.0057
	MIDWEST	0.0013	0.0017	-0.0020	0.0046	0.0034	0.0010	0.0014	0.0053
	WEST	0.0004	0.0017	-0.0029	0.0039	0.0017	0.0010	-0.0003	0.0037
	MALE	-0.0026	0.0013	-0.0052	-0.0002	-0.0030	0.0008	-0.0045	-0.0015
	BLACK	-0.0054	0.0022	-0.0098	-0.0011	-0.0062	0.0013	-0.0088	-0.0037
	MARRIED	-0.0010	0.0014	-0.0036	0.0017	0.0008	0.0008	-0.0008	0.0025
	COLLEGE	0.0035	0.0013	0.0009	0.0062	0.0035	0.0008	0.0019	0.0051
	MEDICAID	-0.0065	0.0042	-0.0147	0.0014	-0.0091	0.0025	-0.0141	-0.0042
	ANYLIM	0.0063	0.0014	0.0035	0.0091	0.0038	0.0008	0.0021	0.0054
	ω_{00}	0.0007	0.0000	0.0007	0.0008	0.0007	0.0000	0.0007	0.0008
	ω_{11}	0.0008	0.0000	0.0008	0.0009	0.0006	0.0000	0.0006	0.0007
	ω_{02}	-0.0041	0.0033	-0.0123	0.0005	-0.0172	0.0009	-0.0189	-0.0152
	ω_{12}	-0.0026	0.0017	-0.0056	0.0009	-0.0021	0.0012	-0.0043	0.0006

References

An, W. (2010), "Bayesian propensity score estimators: Incorporating uncertainties in propensity scores into causal inference," *Sociological Methodology*, 40, 151–189.

Austin, P. C. (2011), "Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies," *Pharmaceutical Statistics*, 10, 150–161.

Brand, J. E. and Halaby, C. N. (2006), "Regression and matching estimates of the effects of elite college attendance on educational and career achievement," *Social Science Research*, 35, 749–770.

Branson, Z., Rischard, M., Bornn, L., and Miratrix, L. W. (2019), "A nonparametric Bayesian methodology for regression discontinuity designs," *Journal of Statistical Planning and Inference*, 202, 14–30.

Caliendo, M. and Kopeinig, S. (2008), "Some practical guidance for the implementation of propensity score matching," *Journal of Economic Surveys*, 22, 31–72.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014a), "Robust data-driven inference in the regression-discontinuity design," *The Stata Journal*, 14, 909–946.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014b), "Robust nonparametric confidence intervals for regression-discontinuity designs," *Econometrica*, 82, 2295–2326.

Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2017), "rdrobust: Software for regression-discontinuity designs," *The Stata Journal*, 17, 372–404.

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

Chan, D., Kohn, R., Nott, D., and Kirby, C. (2006), "Locally adaptive semiparametric estimation of the mean and variance functions in regression models," *Journal of Computational and Graphical Statistics*, 15, 915–936.

Chan, J. C. and Jeliazkov, I. (2009a), "Efficient simulation and integrated likelihood estimation in state space models," *International Journal of Mathematical Modelling and Numerical Optimisation*, 1, 101–120.

Chan, J. C.-C. and Jeliazkov, I. (2009b), "MCMC estimation of restricted covariance matrices," *Journal of Computational and Graphical Statistics*, 18, 457–480.

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

Chaudhuri, K. and Howley, P. (2022), "The impact of COVID-19 vaccination for mental well-being," *European Economic Review*, 150, 104293.

Chib, S. (1995), "Marginal likelihood from the Gibbs output," *Journal of the American Statistical Association*, 90, 1313–1321.

Chib, S. (2007), "Analysis of treatment response data without the joint distribution of potential outcomes," *Journal of Econometrics*, 140, 401–412.

Chib, S. and Hamilton, B. H. (2000), "Bayesian analysis of cross-section and clustered data treatment models," *Journal of Econometrics*, 97, 25–50.

Chib, S. and Hamilton, B. H. (2002), "Semiparametric Bayes analysis of longitudinal data treatment models," *Journal of Econometrics*, 110, 67–89.

Chib, S. and Jacobi, L. (2016), "Bayesian fuzzy regression discontinuity analysis and returns to compulsory schooling," *Journal of Applied Econometrics*, 31, 1026–1047.

Chib, S. and Jeliazkov, I. (2001), "Marginal likelihood from the Metropolis–Hastings output," *Journal of the American Statistical Association*, 96, 270–281.

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

Chib, S., Greenberg, E., and Jeliazkov, I. (2009), "Estimation of semiparametric models in the presence of endogeneity and sample selection," *Journal of Computational and Graphical Statistics*, 18, 321–348.

Chib, S., Greenberg, E., and Simoni, A. (2023), "Nonparametric bayes analysis of the sharp and fuzzy regression discontinuity designs," *Econometric Theory*, 39, 481–533.

Deb, P., Munkin, M. K., and Trivedi, P. K. (2006), "Bayesian analysis of the two-part model with endogeneity: application to health care expenditure," *Journal of Applied Econometrics*, 21, 1081–1099.

Dehejia, R. H. and Wahba, S. (1999), "Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs," *Journal of the American Statistical Association*, 94, 1053–1062.

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

Dong, Y. (2019), "Regression Discontinuity Designs With Sample Selection," *Journal of Business & Economic Statistics*, 37, 171–186.

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

Dong, Y., Lee, Y.-Y., and Gou, M. (2023), "Regression Discontinuity Designs With a Continuous Treatment," *Journal of the American Statistical Association*, 118, 208–221.

Duan, Z., Feng, Y., Xu, S., Gao, D., Ji, Y., Sun, X., Chen, R., and Wang, Y. (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*, p. fdad060.

Fan, J. (1992), "Design-adaptive Nonparametric Regression," *Journal of the American Statistical Association*, 87, 998–1004.

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

Fletcher, J. M. and Tokmouline, M. (2018), "The effects of academic probation on college success: Regression discontinuity evidence from four texas universities," *Available at SSRN 3097330*.

Gamerman, D. (1997), "Sampling from the posterior distribution in generalized linear mixed models," *Statistics and Computing*, 7, 57–68.

Gelfand, A. E., Kottas, A., and MacEachern, S. N. (2005), "Bayesian Nonparametric Spatial Modeling With Dirichlet Process Mixing," *Journal of the American Statistical Association*, 100, 1021–1035.

Geneletti, S., Ricciardi, F., O'Keeffe, A. G., and Baio, G. (2019), "Bayesian modelling for binary outcomes in the regression discontinuity design," *Journal of the Royal Statistical Society Series A: Statistics in Society*, 182, 983–1002.

Gu, Y., Fiebig, D. G., Cripps, E., and Kohn, R. (2009), "Bayesian estimation of a random effects heteroscedastic probit model," *The Econometrics Journal*, 12, 324–339.

Hahn, J., Todd, P., and Van der Klaauw, W. (2001), "Identification and estimation of treatment effects with a regression-discontinuity design," *Econometrica*, 69, 201–209.

Heckman, J. J., Lopes, H. F., and Piatek, R. (2014), "Treatment effects: A Bayesian perspective," *Econometric Reviews*, 33, 36–67.

Imai, K. and van Dyk, D. A. (2004), "Causal Inference With General Treatment Regimes," *Journal of the American Statistical Association*, 99, 854–866.

Jacobi, L., Wagner, H., and Frühwirth-Schnatter, S. (2016), "Bayesian treatment effects models with variable selection for panel outcomes with an application to earnings effects of maternity leave," *Journal of Econometrics*, 193, 234–250.

Jeliazkov, I. (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, I. and Lee, E. H. (2010), "MCMC perspectives on simulated likelihood estimation," in *Maximum Simulated Likelihood Methods and Applications*, Emerald Group Publishing Limited.

Kass, R. E. and Raftery, A. E. (1995), "Bayes factors," *Journal of the American Statistical Association*, 90, 773–795.

Kline, B. and Tobias, J. L. (2008), "The wages of BMI: Bayesian analysis of a skewed treatment–response model with nonparametric endogeneity," *Journal of Applied Econometrics*, 23, 767–793.

Koop, G. and Tobias, J. L. (2004), "Learning about heterogeneity in returns to schooling," *Journal of Applied Econometrics*, 19, 827–849.

Li, M. and Tobias, J. L. (2008), "Bayesian analysis of treatment effects in an ordered potential outcomes model," in *Modelling and Evaluating Treatment Effects in Econometrics*, vol. 21, pp. 57–91, Emerald Group Publishing Limited.

Li, M. and Tobias, J. L. (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, 345–361.

Li, M. and Tobias, J. L. (2014), "Bayesian Analysis of Treatment Effect Models," *Bayesian Inference in the Social Sciences*, pp. 63–90.

Li, M., Poirier, D. J., and Tobias, J. L. (2004), "Do dropouts suffer from dropping out? Estimation and prediction of outcome gains in generalized selection models," *Journal of Applied Econometrics*, 19, 203–225.

McCrary, J. (2008), "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, 142, 698–714.

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

Munkin, M. K. (2003), "The MCMC and SML estimation of a self-selection model with two outcomes," *Computational Statistics and Data Analysis*, 42, 403–424, Computational Ecomometrics.

Munkin, M. K. (2011), "The Endogenous Sequential Probit model: An application to the demand for hospital utilization," *Economics Letters*, 112, 182–185.

Munkin, M. K. and Trivedi, P. K. (1999), "Simulated maximum likelihood estimation of multivariate mixed-Poisson regression models, with application," *The Econometrics Journal*, 2, 29–48.

Munkin, M. K. and Trivedi, P. K. (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, 197–220.

Nott, D. J. and Leonte, D. (2004), "Sampling schemes for Bayesian variable selection in generalized linear models," *Journal of Computational and Graphical Statistics*, 13, 362–382.

Rosenbaum, P. R. and Rubin, D. B. (1983), "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70, 41–55.

Rosenbaum, P. R. and Rubin, D. B. (1985), "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score," *The American Statistician*, 39, 33–38.

Rosenbaum, P. R. and Rubin, D. B. (2022), "Propensity scores in the design of observational studies for causal effects," *Biometrika*, 110, 1–13.

Roy, A. D. (1951), "Some thoughts on the distribution of earnings," *Oxford Economic Papers*, 3, 135–146.

Rubin, D. B. (1974), "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of Educational Psychology*, 66, 688.

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

Rubin, D. B. (1978), "Bayesian inference for causal effects: The role of randomization," *The Annals of Statistics*, pp. 34–58.

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

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

Silverman, B. W. (1985), "Some aspects of the spline smoothing approach to non-parametric regression curve fitting," *Journal of the Royal Statistical Society: Series B (Methodological)*, 47, 1–21.

Thistlethwaite, D. L. and Campbell, D. T. (1960), "Regression-discontinuity analysis: An alternative to the ex post facto experiment." *Journal of Educational Psychology*, 51, 309.

Thoemmes, F. (2012), "Propensity score matching in SPSS," *arXiv preprint arXiv:1201.6385*.

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, M. (2014), "Bayesian Analysis of Sample Selection Models," *Bayesian Inference in the Social Sciences*, pp. 91–113.

Vossmeyer, A. (2014a), "Determining the proper specification for endogeneous covariates in discrete data settings," *Advances in Econometrics*, 34, 223–247.

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

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

Wahba, G. (1978), "Improper priors, spline smoothing and the problem of guarding against model errors in regression," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 40, 364–372.

Williams, C. K. I. (1998), *Prediction with Gaussian Processes: From Linear Regression to Linear Prediction and Beyond", bookTitle="Learning in Graphical Models*, pp. 599–621, Springer Netherlands.

Wright, N. A. (2020), "Perform better, or else: Academic probation, public praise, and students decision-making," *Labour Economics*, 62, 101773.

Zhang, Z., Kim, H. J., Lonjon, G., Zhu, Y., et al. (2019), "Balance diagnostics after propensity score matching," *Annals of Translational Medicine*, 7.

Zhao, S., van Dyk, D. A., and Imai, K. (2020), "Propensity score-based methods for causal inference in observational studies with non-binary treatments," *Statistical Methods in Medical Research*, 29, 709–727.

Zhao, Z. (2008), "Sensitivity of propensity score methods to the specifications," *Economics Letters*, 98, 309–319.